REVISTA DE INGENIERÍA

UNIVERSIDAD DE NARIÑO

Vol. 04 No. 01 ISSN 2838-7732 (En línea) Noviembre 2026



Universidad de Nariño
FUNDADA EN 1904

Universidad de Nariño
ACREDITADA DE ALTA CALIDAD
RESOLUCIÓN MEN 000022 - ENERO 11 DE 2023

Facultad de INGENIERÍA
Universidad de Nariño

UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
CENTRO DE INVESTIGACIÓN E INTERACCIÓN SOCIAL (CIIS-FACING)

Dr. Alexander Barón Salazar
Decano de la Facultad de Ingeniería

COMITÉ EDITORIAL

Dr. Manuel Ernesto Bolaños González
Director del Departamento de Sistemas
Dr. Wagner Germán Suero Pérez
Director del Departamento de Electrónica
Dr. Guillermo Muñoz Ricaurte
Director del Departamento de Civil
Dr. Jesús Insuasti
Docente T.C. categoría Titular, Investigador
Dr. Oscar Revelo Sánchez
Director del Doctorado en Ingeniería

EQUIPO EDITORIAL

MSc. Jairo Antonio Guerrero García
Editor y coordinador del CIIS-FACING
Ing. María Camila Triviño Quiñonez
Monitor Estudiantil / Asistencia Editorial

INFORMACIÓN DE LA PUBLICACIÓN

Revista de Ingeniería
ISSN: 2838-7732 (En línea)
Periodicidad: Semestral
Contacto: revistadeingenieria@udenar.edu.co
Dirección: Ciudadela Universitaria - Calle 18 No. 50-02 - Bloque 6 - Piso 3. San Juan de Pasto, Nariño, Colombia.

Aviso Legal y Políticas de Publicación:

La **Revista de Ingeniería** es una publicación seriada oficial de la Facultad de Ingeniería de la Universidad de Nariño, avalada mediante el Acuerdo No. 038 de 2026 de su Consejo Editorial. Su objetivo es la difusión del conocimiento científico, técnico y de innovación de la región. Los artículos aquí publicados han sido sometidos a un proceso de evaluación por pares académicos bajo la modalidad de doble ciego, garantizando el rigor y la calidad científica de los contenidos.

Los conceptos, juicios y opiniones expresados en los artículos son de responsabilidad exclusiva de los autores y no reflejan necesariamente el pensamiento de la Universidad de Nariño, de la Facultad de Ingeniería ni del Comité Editorial.

Se autoriza la reproducción total o parcial de los textos aquí publicados, siempre y cuando se cite de manera completa la fuente original, reconociendo los derechos patrimoniales y morales de los autores y de la institución.

Universidad de Nariño
FUNDADA EN 1904

EDITORIAL

La concepción de la ingeniería encuentra su verdadero sentido cuando el rigor investigativo se traduce en respuestas tangibles para la sociedad. En 1906, bajo la dirección del ingeniero Fortunato Pereira Gamba, nació la Revista de Ingeniería de la Universidad de Nariño con una vocación muy clara: propagar el conocimiento y fomentar el desarrollo de las industrias en nuestra región. Hoy, más de un siglo después, desde el Centro de Investigación e Interacción Social (CIIS-FACING) reafirmamos ese mandato fundacional.

Esta publicación resurge y se consolida como el canal oficial para visibilizar la producción intelectual de nuestros investigadores, docentes, estudiantes y egresados. Es motivo de especial orgullo compartir que hemos sido ratificados para nuestra producción en línea y avalados por el Consejo Editorial de la Universidad de Nariño mediante el Acuerdo No. 038 de 2026. Este respaldo marca el inicio de una etapa de publicación semestral, cuyo lanzamiento celebramos este noviembre, rindiendo homenaje al aniversario de fundación de nuestra institución y de la Facultad de Ingeniería.

Nuestro propósito no se limita a documentar los hallazgos en las distintas ramas de la ingeniería; buscamos establecer un puente sólido y permanente entre el quehacer académico y las necesidades apremiantes del sector productivo local y regional. Fieles a la excelencia académica que caracteriza a nuestra facultad, asumimos la responsabilidad de mantener a la comunidad actualizada frente a los nuevos paradigmas tecnológicos. Sabemos que la divulgación científica es un esfuerzo colectivo. Por ello, cada artículo publicado en estas páginas representa el compromiso de someter nuestras ideas al rigor del método científico y garantizar que la Universidad de Nariño siga siendo un actor determinante en el progreso de nuestro territorio.

Invitamos a investigadores, profesionales y estudiantes a hacer de este espacio su principal plataforma de debate y difusión. La historia de nuestra facultad nos exige estar a la altura de los tiempos; con la presente edición, demostramos que seguimos construyendo el futuro desde la ciencia, el análisis y la innovación.

Jairo Antonio Guerrero García, MSc.
Editor Revista de Ingeniería
Universidad de Nariño

Universidad de Nariño
FUNDADA EN 1904

ÍNDICE

Editorial——[03]

ARTÍCULOS DE INVESTIGACIÓN:

DockerHard: Evaluación de seguridad y hardening de contenedores para proyectos de videojuegos desplegados en Linux
J. A. Collazos, R. Cadena——[05]

miniLOGO: A Multilingual, Multi-Turtle Turtle-Graphics Environment for Arts-Integrated Computational Thinking in High School
J. Insuasti y E. Insuasty——[13]

How Spell-Checkers Work — A Non-Technical Experience in Spanish
J. Insuasti E. Insuasty——[23]

Comparative Analysis of Glucose Control Strategies in T1D under Stress Scenarios
D. A. Ojeda Guzmán, M. A. Muriel Tobar, G. D. Obando Bravo y A .F. Arciniegas Mejía——[32]

A Predictive Model for Energy Consumption in P2P Networks Based on Knowledge Discovery and Artificial Intelligence
K. L. Yandar, O. Revelo Sánchez y M. Bolaños González——[45]

Control predictivo adaptativo para regulacion del estres mediante estimulos musicales basados en señales HRV
A. Pantoja, A. A. Jimenez Ocaña y J. Cabrera——[61]

Estimacion de la generacion fotovoltaica mediante modelos lineales basados en datos satelitales: estudio comparativo en Pasto, Colombia
D. F. Villaquiran Samboni y J. Revelo Fuelagán——[76]

Simulación de un brazo robotico bioinspirado controlado mediante tecnicas de Deep Learning
S. G. Eraso-Montenegro, P. J. Jiménez-Torres y W. O. Achicanoy-Martínez——[88]

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

DockerHard: Evaluación de seguridad y hardening de contenedores para proyectos de videojuegos desplegados en Linux

DockerHard: Security Assessment and Container Hardening for Video Game Projects Deployed on Linux

Jorge Andrés Collazos Narváez [1]*, Rodrigo Cadena Martínez [2]

[1] Programa de Ingeniería de Sistemas, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia. jorge.collazos@udenar.edu.co

[2] Doctorado en Informática, Universidad Americana de Europa UNADE, Cancún, Quintana Roo, México. rodrigo.cadena@unade.edu.co

* Autor de correspondencia: jorge.collazos@udenar.edu.co

Citación: J. A. Collazos Narváez y R. Cadena Martínez, "DockerHard: Evaluación de seguridad y hardening de contenedores para proyectos de videojuegos desplegados en Linux," *Revista de Ingeniería*, vol. 4, no. 1, pp. 05-12, Nov. 2026.

Abstract—The adoption of containerization technologies has driven the use of Docker images for application deployment in Linux environments; however, insecure configurations, inherited vulnerabilities, and third-party dependencies continue to pose significant risks to container security. This article presents an evaluation of Docker image hardening using DockerHard, a command-line tool developed to automate the security analysis of Dockerfiles and Docker images. The solution integrates static analysis based on rules derived from the OWASP Docker Security Cheat Sheet, the CIS Docker Benchmark, and NIST SP 800-190, as well as dynamic vulnerability analysis using Trivy and risk prioritization using FIRST-EPSS. Experimental validation was conducted on five video games developed for Linux using Python/Pygame and Unity. The results evidenced significant reductions in insecure configurations, a decrease in vulnerabilities, and quantifiable improvements in the security posture measured by the Hardening Score and Hardening Improvement Index metrics, demonstrating the effectiveness of the applied hardening activities.

Keywords—Container Security, Docker, Hardening, Linux Video Games, Vulnerability Analysis

Resumen—La adopción de tecnologías de contenedorización ha impulsado el uso de imágenes Docker para el despliegue de aplicaciones en entornos Linux; sin embargo, configuraciones inseguras, vulnerabilidades heredadas y dependencias de terceros continúan representando riesgos relevantes para la seguridad de los contenedores. Este articulo presenta una evaluación del hardening de imágenes Docker mediante DockerHard, una herramienta de línea de comandos desarrollada para automatizar el análisis de seguridad de Dockerfiles e imágenes Docker. La solución integra análisis estático basado en reglas derivadas de OWASP Docker Security Cheat Sheet, CIS Docker Benchmark y NIST SP 800-190, así como análisis dinámico de vulnerabilidades mediante Trivy y priorización de riesgos utilizando FIRST-EPSS. La validación experimental se realizó sobre cinco videojuegos desarrollados para Linux utilizando Python/Pygame y Unity. Los resultados evidenciaron reducciones significativas de configuraciones inseguras, disminución de vulnerabilidades y mejoras cuantificables en la postura de seguridad mediante las métricas Hardening Score e Índice de Mejora de Hardening, demostrando la efectividad de las actividades de hardening aplicadas.

Palabras claves—Análisis de vulnerabilidades, Docker, Hardening, Seguridad de Contenedores, Video Juegos Linux.

I. INTRODUCCIÓN

El uso de tecnologías de contenedorización continúa creciendo de manera sostenida en organizaciones de diferentes sectores. Sin embargo, este crecimiento también ha incrementado la dependencia de repositorios públicos y componentes reutilizables dentro de la cadena de suministro de software. Estudios recientes evidencian que las imágenes Docker incorporan una gran cantidad de dependencias de terceros y componentes externos cuyo estado de seguridad no siempre es conocido o gestionado adecuadamente. En una investigación desarrollada sobre imágenes ampliamente utilizadas en Docker Hub, NetRise identificó que las imágenes analizadas contenían en promedio 389 componentes de software y más de 600 vulnerabilidades conocidas, evidenciando la complejidad del ecosistema de dependencias que conforma una imagen moderna de contenedor [1]. Asimismo, se encontró que una proporción significativa de estas vulnerabilidades permanecía sin corregir durante años, incrementando el riesgo asociado a la reutilización de imágenes dentro de procesos de desarrollo y despliegue.

De acuerdo con el informe State of Container and Kubernetes Security Report de Red Hat, la adopción masiva de tecnologías basadas en contenedores ha incrementado la superficie de ataque de las organizaciones, generando nuevos vectores de explotación asociados a imágenes vulnerables, configuraciones incorrectas y errores de implementación [2]. De manera similar, el Cloud-Native Security and Usage Report de Sysdig indica que los entornos basados en contenedores se han convertido en objetivos recurrentes para actores maliciosos debido a la elevada presencia de vulnerabilidades conocidas en imágenes desplegadas en ambientes productivos [3].

Diversos estudios han demostrado que una proporción considerable de las imágenes Docker disponibles públicamente contienen vulnerabilidades de seguridad. Banyan et al. identificaron que una gran cantidad de imágenes analizadas incorporaban componentes vulnerables heredados de imágenes base y dependencias de terceros [4]. Asimismo, investigaciones desarrolladas por Shu, Tian y Williams evidenciaron que numerosas imágenes Docker distribuidas a través de repositorios públicos presentan vulnerabilidades conocidas que permanecen sin corregir durante largos periodos de tiempo, incrementando el riesgo de explotación por parte de atacantes [5].

Con el propósito de mitigar estos riesgos, organismos internacionales y comunidades especializadas han desarrollado diferentes estándares y guías de buenas prácticas orientadas al fortalecimiento de la seguridad en contenedores. Entre las más relevantes se encuentran el documento NIST SP 800-190: Application Container Security Guide [6], el CIS Docker Benchmark [7] y OWASP Docker Security Cheat Sheet [8], los cuales establecen recomendaciones para reducir privilegios, minimizar superficies de ataque, gestionar vulnerabilidades y fortalecer los mecanismos de protección de imágenes y contenedores.

A pesar de la existencia de estos lineamientos, la aplicación práctica de actividades de hardening continúa representando un desafío para organizaciones y equipos de desarrollo. En muchos casos, las actividades de endurecimiento se realizan de forma manual, heterogénea y sin mecanismos objetivos que permitan medir el impacto real de las acciones implementadas sobre la reducción del riesgo. Adicionalmente, las herramientas disponibles suelen concentrarse en tareas específicas como el análisis estático de Dockerfiles o el escaneo de vulnerabilidades sobre imágenes, sin proporcionar una visión integrada que permita evaluar sistemáticamente la efectividad de las actividades de hardening realizadas.

Esta problemática adquiere especial relevancia en proyectos de videojuegos ejecutados sobre entornos Linux. El crecimiento de las plataformas de distribución digital, los procesos de integración y despliegue continuo, así como la utilización de servidores de juego basados en Linux, han impulsado el uso de tecnologías de contenedorización dentro de diferentes etapas del ciclo de vida de los videojuegos.

En este contexto, resulta pertinente desarrollar un método formal de hardening que permita integrar buenas prácticas derivadas de estándares reconocidos internacionalmente con mecanismos automatizados de análisis y evaluación de seguridad. La construcción de la herramienta DockerHard responde precisamente a esta necesidad, proporcionando una plataforma capaz de realizar análisis estático de Dockerfiles, análisis dinámico de imágenes Docker, integración con motores de detección devulnerabilidades como Trivy, enriquecimiento de resultados mediante información de FIRST-EPSS y generación de métricas cuantitativas orientadas a medir el impacto de las actividades de hardening.

II. METODOLOGÍA

La investigación se desarrolló mediante un diseño metodológico de validación experimental orientado a evaluar la efectividad de la herramienta propuesta sobre imágenes Docker utilizadas en entornos de ejecución de videojuegos Linux.

La validación experimental consistió en ejecutar la herramienta sobre distintos Dockerfiles e imágenes Docker con diferentes niveles de seguridad, permitiendo comparar métricas obtenidas antes y después de aplicar mecanismos de hardening. Este proceso permitió medir el impacto de la propuesta en términos de reducción de

vulnerabilidades, disminución de configuraciones inseguras y fortalecimiento general de la postura de seguridad de las imágenes evaluadas.

Para el desarrollo de la solución se adoptó un enfoque iterativo-incremental, mediante el cual la herramienta fue construida progresivamente a través de módulos funcionales sucesivos. Cada incremento incorporó nuevas capacidades de análisis y endurecimiento, permitiendo validar continuamente el funcionamiento de la solución y realizar ajustes técnicos antes de avanzar a la siguiente etapa.

El uso de un enfoque iterativo-incremental resultó coherente con investigaciones recientes relacionadas con automatización de hardening y seguridad continua en entornos contenerizados, las cuales destacan la necesidad de integrar procesos automatizados de análisis, validación y remediación dentro de los ciclos de construcción y despliegue de imágenes Docker.

A diferencia de herramientas tradicionales enfocadas exclusivamente en vulnerabilidades de paquetes, DockerHard integro análisis estático, análisis dinámico, evaluación de riesgo y generación de métricas de mejora, permitiendo obtener una visión integral de la postura de seguridad de los contenedores.

Fase 1. Diseño de la arquitectura base y CLI

La primera fase correspondió a la definición de la arquitectura base de la solución y la construcción del núcleo funcional del CLI. El objetivo principal de esta etapa consistió en establecer la estructura modular del sistema y construir el punto de entrada encargado de orquestar las operaciones de análisis de seguridad.

La herramienta fue implementada en Python e implemento módulos especializados como analyzer.py, scanner.py, recommender.py, report.py y utils.py, permitiendo garantizar separación de responsabilidades, mantenibilidad y escalabilidad de la solución.

Asimismo, se implementó un CLI inicial utilizando la librería click, permitiendo ejecutar comandos personalizados y recibir parámetros asociados al análisis de Dockerfiles e imágenes Docker. El sistema fue empaquetado mediante setup.py, habilitando la ejecución de comandos.

La implementación mediante una interfaz de línea de comandos facilita la automatización de tareas de seguridad y su integración dentro de procesos DevSecOps y pipelines de integración continua. En este sentido, la automatización de actividades técnicas mediante herramientas ejecutables por línea de comandos constituye una práctica fundamental para garantizar consistencia, repetibilidad y trazabilidad durante el ciclo de vida del software [9].

Esta fase constituyó el núcleo estructural de la herramienta y definió el flujo interno de ejecución de los procesos de análisis.

Fase 2. Análisis estático de Dockerfiles

La segunda fase se centró en el desarrollo del módulo encargado del análisis estático (motor de análisis estático) de configuraciones declarativas presentes en Dockerfiles.

El motor de análisis estático es responsable de inspeccionar el contenido de los Dockerfiles sin necesidad de construir la imagen. Este módulo evalúa configuraciones inseguras, malas prácticas y desviaciones respecto a estándares de seguridad reconocidos, permitiendo identificar riesgos desde las etapas tempranas del ciclo de construcción del contenedor. Su propósito es detectar vulnerabilidades potenciales antes de que estas sean incorporadas a la imagen final, favoreciendo la adopción de principios de seguridad por diseño. El análisis estático constituye una de las estrategias más efectivas para identificar problemas de seguridad durante las fases inciales del desarrollo de software [10]. Complementariamente, las herramientas de análisis estático permiten detectar errores y defectos de manera temprana, reduciendo el costo asociado a su corrección en etapas posteriores del ciclo de desarrollo [11].

Las validaciones implementadas tomaron como referencia estándares y lineamientos internacionales relacionados con seguridad en contenedores, tales como OWASP Docker Security Cheat Sheet, Docker CIS Benchmark y NIST SP 800- 190.

Entre los elementos evaluados se encontraron:

Uso de imágenes base sin versionamiento explícito.

–Ejecución bajo privilegios root.

–Ausencia de mecanismos HEALTHCHECK.

–Uso incorrecto de instrucciones ADD.

–Instalación de paquetes innecesarios.

–Ausencia de limpieza de caché.

–Configuraciones que incrementaran la superficie de ataque.

Los hallazgos obtenidos fueron exportados al formato .XLSX, permitiendo su posterior correlación y análisis automatizado.

Fase 3. Escaneo de vulnerabilidades en imágenes Docker

La tercera fase amplió el análisis hacia la evaluación dinámica de vulnerabilidades presentes en las capas internas de las imágenes Docker construidas. Para ello se construyó el motor de análisis dinámico que integró la herramienta especializada Trivy, encargada de consultar bases de datos internacionales de vulnerabilidades CVE.

El objetivo de esta fase consistió en identificar componentes vulnerables, bibliotecas inseguras y dependencias obsoletas presentes en la imagen final. Adicionalmente, el sistema permitió clasificar vulnerabilidades según niveles de severidad y correlacionar los hallazgos con los componentes detectados.

La importancia de este enfoque ha sido reconocida por NIST, indicando que una parte significativa de las vulnerabilidades presentes en entornos basados en contenedores proviene de bibliotecas, paquetes del sistema operativo y componentes heredados incorporados durante la construcción de las imágenes. Por esta razón, la guía NIST SP 800-190 recomienda la evaluación continua de imágenes Docker como parte de las actividades de seguridad orientadas a identificar vulnerabilidades que no pueden detectarse únicamente mediante la inspección del Dockerfile [6].

De forma complementaria, Mills, White y Legg demostraron que el riesgo asociado a las imágenes Docker evoluciona con el tiempo debido a la aparición de nuevas vulnerabilidades en los componentes que las conforman, incluso cuando la imagen no ha sido modificada [12]. Este hallazgo refuerza la necesidad de realizar evaluaciones dinámicas periódicas sobre imágenes construidas, permitiendo identificar riesgos emergentes y mantener una postura de seguridad actualizada.

Fase 4. Motor de recomendaciones y puntuación de hardening

La cuarta fase incorporo un componente analítico avanzado orientado a interpretar los resultados obtenidos durante las fases previas y generar recomendaciones automatizadas de hardening. Este módulo implemento un conjunto de reglas formalizadas en archivos. YAML que contendrán identificadores, niveles de severidad, descripciones de riesgo, recomendaciones técnicas e impacto estimado.

Se incorporó el Hardening Score como una métrica cuantitativa utilizada para representar el nivel de seguridad de un Dockerfile o una imagen Docker a partir de los hallazgos identificados durante el proceso de análisis. El cálculo se basa en un modelo de penalización ponderada, donde cada vulnerabilidad detectada contribuye al riesgo total de acuerdo con su severidad. Las vulnerabilidades clasificadas como Critical (C) reciben una penalización mayor que las clasificadas como High (H), Medium (M), Low (L) o Unknown (U), reflejando su impacto potencial sobre la seguridad del contenedor. La Tabla 1 muestra el resumen de la penalización de los hallazgos detectados por Dockerhard.

Además, DockerHard incorpora el valor EPSS (Exploit Prediction Scoring System) dentro del cálculo del riesgo ponderado, permitiendo ajustar la penalización de cada vulnerabilidad según su probabilidad estimada de explotación. De esta manera, el Hardening Score refleja simultáneamente la severidad del hallazgo y la probabilidad de que dicho hallazgo sea explotado en escenarios reales.

TABLA I
SEVERIDAD DE HALLAZGOS DETECTADOS POR DOCKERHARD

Severidad	Penalización
Critical	15
High	8
Medium	4
Low	0,3
Unknown	2

Además, DockerHard incorpora el valor EPSS (Exploit Prediction Scoring System) dentro del cálculo del riesgo ponderado, permitiendo ajustar la penalización de cada vulnerabilidad según su probabilidad estimada de explotación. De esta manera, el Hardening Score refleja simultáneamente la severidad del hallazgo y la probabilidad de que dicho hallazgo sea explotado en escenarios reales.

El valor EPSS se calcula mediante la integración en tiempo real con la API de FIRST. El sistema EPSS es un modelo de aprendizaje automático basado en datos que estima la probabilidad de que un CVE publicado sea explotado durante los siguientes 30 días.

Adicionalmente, se incorporó la métrica Índice de Mejora de Hardening (IMH), la cual correspondió a una métrica propuesta durante la investigación para cuantificar la efectividad de las medidas de hardening aplicadas sobre Dockerfiles e imágenes Docker. El indicador expresa el porcentaje de reducción de hallazgos de seguridad obtenidos entre una línea base inicial y un estado posterior a la aplicación de controles de mitigación. Valores cercanos al 100% indican una reducción significativa de vulnerabilidades y una mejora sustancial en el nivel de seguridad del contenedor analizado.

Es importante resaltar que el Hardening Score mide el estado de seguridad del artefacto analizado en un momento determinado, mientras que el Índice de Mejora de Hardening (IMH) mide la efectividad de las acciones de mitigación aplicadas respecto a una línea base inicial. Por esta razón, ambas métricas se utilizan de manera complementaria para evaluar el resultado de los procesos de hardening realizados durante la investigación.

Para realizar el cálculo, DockerHard emplea en el módulo score.py en orden las ecuaciones: Ecuación 1, Ecuación 2 y Ecuación 3, y obtiene al final el hardening score y lo reporta al usuario.

$$Penalización\ Total = Peso_{Severidad} \times (1 + EPPS) \quad (1)$$

$$Riesgo\ Ponderado = \sum Penalización \quad (2)$$

$$Hardening\ Score = \frac{100}{1 + \left(\frac{Riesgo\ Ponderado}{100}\right)} \quad (3)$$

Fase 5. Generación de reportes y validación experimental

La fase final integro los resultados obtenidos por los distintos módulos y consolida la información en reportes exportables en formato .XLSX. Los reportes incluirán vulnerabilidades detectadas, configuraciones inseguras identificadas, score de hardening y recomendaciones priorizadas de seguridad.

Posteriormente, la herramienta fue validada experimentalmente mediante pruebas ejecutadas sobre imágenes Docker asociadas a entornos Linux para videojuegos, incluyendo builds headless de motores como Unity y frameworks basados en Python como Pygame.

III. RESULTADOS Y DISCUSIÓN

La validación experimental del método propuesto se realizó mediante la evaluación de cinco proyectos de videojuegos desarrollados para entornos Linux. Tres de los casos de estudio fueron implementados utilizando Python/Pygame (Aeroblasters, Flappy Bird y Memory Puzzle), mientras que dos proyectos fueron desarrollados utilizando Unity (2D Platformer y FPS Microgame). Para cada caso se analizaron dos artefactos: el Dockerfile y la imagen Docker resultante, obteniendo un total de diez artefactos evaluados. El proceso experimental consistió en ejecutar DockerHard sobre los Dockerfiles e imágenes iniciales, aplicar actividades de hardening recomendadas por la herramienta en los reportes individuales y que están derivadas de OWASP Docker Security Cheat Sheet, CIS Docker Benchmark y NIST SP 800-190, posteriormente repetir el proceso de evaluación para medir cuantitativamente las mejoras obtenidas.

Funcionamiento general de la herramienta Dockerhard

La herramienta se ejecuta mediante el comando dockerhard, el cual admite diferentes parámetros según el tipo de análisis requerido. Entre los más relevantes se encuentran --file, utilizado para el análisis estático de Dockerfiles, y --image, empleado para la evaluación dinámica de imágenes Docker. Estos comandos constituyen la base operativa de la herramienta y fueron utilizados de manera sistemática durante el desarrollo de todos los casos experimentales presentados en esta investigación.

La Tabla II presenta la descripción de los principales comandos disponibles y su propósito dentro del proceso de evaluación de seguridad implementado por DockerHard.

TABLA II
COMANDOS DE EJECUCIÓN DOCKERHARD CLI

Comandos de ejecución	Descripción
dockerhard scan --file <nombre_dockerfile>	Se emplea el comando **dockerhard scan** seguido el atributo **--file** y el nombre del archivo docker, para que la herramienta realice el análisis estático de archivos dockerfile ejemplo: **dockerhard scan --file midockerfile**
dockerhard scan --image <nombre_imagen:tag>	Se emplea el comando **dockerhard scan** seguido el atributo **--image** y el nombre de la imagen creada en docker(se puede emplear previamente docker images para visualizar los nombres y tags de las imágenes creadas), para que la herramienta realice el análisis de imágenes docker, ejemplo: **dockerhard scan --image mijuego:latest**
dockerhard scan --image <nombre_imagen:tag> --limit 20 --output reports/image-scan.xlsx	Dockerhard por medio del atributo **–image** realiza el escaneo dinámico de imágenes; **--limit**<# de resultados> permite obtener solo el número de hallazgos indicado en la terminal; **--output** permite exportar los resultados a un archivo .xlsx con el nombre de preferencia.
dockerhard compare --before <archivo.xlsx> --after <atchivo2.xlsx>	Dockerhard permitirá obtener el IMH(Índice de Mejora de Hardening), por medio de los atributos **--before**<nombre del archivo xlsx de análisis previo> y **--after**<nombre del archivo xlsx del análisis posterior>

Resultados del análisis estático

El análisis estático permitió identificar configuraciones inseguras recurrentes en los Dockerfiles evaluados, entre ellas la ejecución con privilegios elevados, ausencia de usuarios no privilegiados, instalación de paquetes innecesarios, utilización de imágenes base con una mayor superficie de ataque y ausencia de mecanismos de monitoreo mediante HEALTHCHECK. La Tabla III y la Figura, presentan la distribución de severidades de los

hallazgos detectados por DockerHard durante la evaluación inicial de los Dockerfiles inseguros y los resultados obtenidos después de aplicar las actividades de hardening y ejecutar nuevamente el proceso de análisis.

TABLA III
SEVERIDAD DE HALLAZGOS DETECTADOS POR DOCKERHARD

Tipo	Estado	C	H	L	M	U	Total
DockerFile	Antes	13	34	23	25	0	95
DockerFile	Después	0	0	0	0	0	2

Los resultados evidencian una reducción sustancial de los hallazgos identificados por DockerHard, pasando de 95 observaciones de seguridad a únicamente 2 hallazgos residuales, lo que representa una disminución global del 97,89 %. Particularmente relevante resulta la eliminación completa de los hallazgos clasificados como Critical, High y Medium, categorías que agrupan los riesgos con mayor potencial de impacto sobre la confidencialidad, integridad y disponibilidad de los contenedores. Este comportamiento demuestra que las actividades de hardening aplicadas fueron efectivas para corregir configuraciones inseguras asociadas principalmente a la gestión de privilegios, manejo de secretos, selección de imágenes base y adopción de buenas prácticas durante la construcción de contenedores.

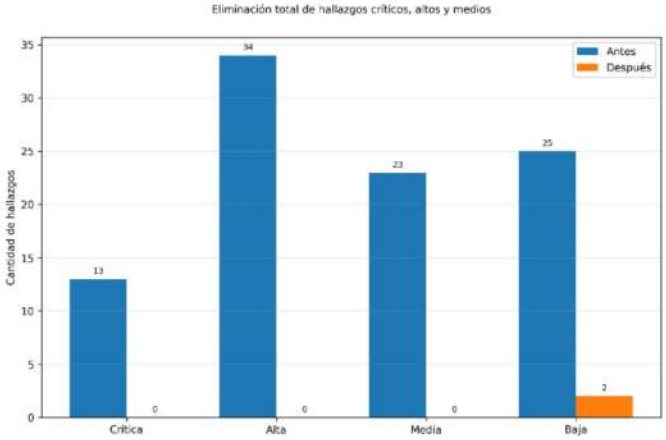


Fig. 1. Distribución de severidades antes y después del hardening

La Tabla IV, muestra los resultados que DockerHard identificó, inicialmente 95 hallazgos distribuidos entre los cinco Dockerfiles analizados, siendo FPS Microgame y Memory Puzzle los proyectos con mayor cantidad de configuraciones inseguras, con 25 y 24 hallazgos respectivamente. Asimismo, cuatro de los cinco proyectos presentaron un Hardening Score inicial clasificado como Riesgo Crítico, evidenciando una postura de seguridad deficiente y una alta exposición a configuraciones inseguras durante la fase de construcción.

TABLA IV
SEVERIDAD DE HALLAZGOS DETECTADOS POR DOCKERHARD

Proyecto	HI	HF	IMH	Score inicial	Score final
Aeroblasters	10	0	100%	67.16%	100%
Flappy Bird	18	0	100%	49.95%	100%
MemoryPuzzle	24	0	100%	38.94%	100%
2D Platformer	18	1	94.44%	55.49%	97.7%
FPSMicrogame	25	1	96%	38.4%	97.7%

Después de aplicar las actividades de hardening, los resultados evidencian una mejora sustancial en todos los casos de estudio. Aeroblasters, Flappy Bird y Memory Puzzle alcanzaron un Índice de Mejora de Hardening (IMH) del 100 %, eliminando completamente los hallazgos identificados durante la evaluación inicial. Por su parte, 2D Platformer y FPS Microgame alcanzaron valores de IMH de 94,44 % y 96,00 %, respectivamente, manteniendo únicamente un hallazgo residual de severidad baja en cada caso.

Desde la perspectiva del Hardening Score, los cinco proyectos alcanzaron valores finales superiores a 99 puntos, clasificándose dentro del nivel de Endurecimiento óptimo. Este comportamiento evidencia que las actividades de hardening implementadas permitieron corregir efectivamente configuraciones inseguras relacionadas con privilegios excesivos, gestión de dependencias, selección de imágenes base y aplicación de buenas prácticas recomendadas para la construcción de contenedores.

Un aspecto particularmente relevante es la transición observada entre los niveles de riesgo iniciales y finales. Mientras que el 80 % de los proyectos evaluados se encontraban inicialmente en la categoría Riesgo Crítico, después de aplicar las medidas de endurecimiento el 100 % de los Dockerfiles alcanzó el endurecimiento óptimo. Este resultado demuestra la efectividad de DockerHard no solo para identificar desviaciones respecto a estándares como OWASP Docker Security Cheat Sheet, CIS Docker Benchmark y NIST SP 800-190, sino también para cuantificar objetivamente la mejora obtenida mediante las métricas Hardening Score e Índice de Mejora de Hardening.

Resultados del análisis dinámico

La Tabla V presenta los resultados consolidados obtenidos durante el análisis dinámico de las imágenes Docker asociadas a los cinco videojuegos evaluados. A diferencia del análisis estático, cuyo objetivo fue identificar configuraciones inseguras presentes en los Dockerfiles, el análisis dinámico permitió evaluar la superficie de ataque real de las imágenes construidas, identificando vulnerabilidades heredadas desde bibliotecas del sistema operativo, paquetes instalados, dependencias de terceros y componentes incorporados durante el proceso de construcción.

TABLA V
SCORE E IMHOBTENIDOS

Proyecto	HI	HF	IMH	Score inicial	Score final
Aeroblasters	1899	105	100%	67.16%	100%
Flappy Bird	382	108	100%	49.95%	100%
MemoryPuzzle	1089	108	100%	38.94%	100%
2D Platformer	96	77	94.44%	55.49%	97.7%
FPSMicrogame	589	285	96%	38.4%	97.7%

Los resultados evidencian diferencias significativas entre los proyectos evaluados, tanto en la cantidad inicial de vulnerabilidades como en la efectividad de las actividades de hardening aplicadas, según lo presenta la Figura 2.

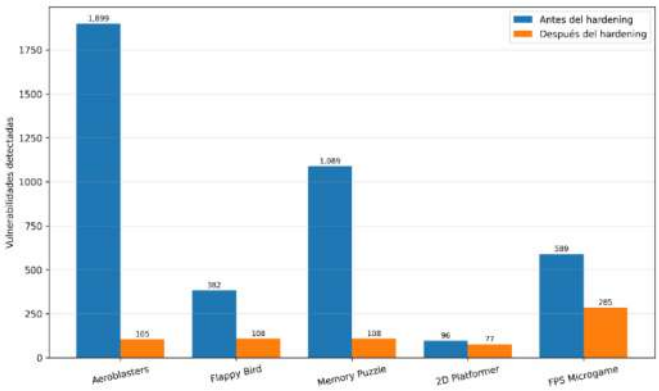


Fig. 2. Vulnerabilidades antes y después del hardening
[1].

En términos absolutos, Aeroblasters presentó la mayor cantidad de vulnerabilidades iniciales con 1899 hallazgos, seguido por Memory Puzzle con 1089 vulnerabilidades. Este comportamiento sugiere una fuerte dependencia de componentes heredados desde imágenes base y bibliotecas externas, lo que incrementa considerablemente la superficie de ataque inicial de las imágenes Docker.

Después de aplicar las actividades de hardening, Aeroblasters redujo sus vulnerabilidades a 105 hallazgos, alcanzando un Índice de Mejora de Hardening (IMH) de 94,47 %, el valor más alto observado en los experimentos. De forma similar, Memory Puzzle alcanzó un IMH de 90,08 %, reduciendo el número de vulnerabilidades de 1089 a 108. Estos resultados indican que las medidas implementadas lograron eliminar una proporción significativa de componentes vulnerables y optimizar la composición final de las imágenes Docker.

Flappy Bird presentó un comportamiento intermedio, alcanzando una reducción del 71,73 % de las vulnerabilidades identificadas inicialmente. Aunque el número absoluto de vulnerabilidades iniciales fue considerablemente menor que en Aeroblasters y Memory Puzzle, el resultado evidencia que la aplicación de actividades de hardening también produce beneficios significativos en imágenes con una superficie de ataque inicial moderada.

Por otra parte, los resultados obtenidos en los proyectos desarrollados con Unity presentan características particulares. En el caso de FPS Microgame, el número de vulnerabilidades se redujo de 589 a 285, alcanzando un IMH de 51,61 %. Aunque la mejora obtenida es relevante, el volumen residual de vulnerabilidades continúa siendo considerablemente superior al observado en los proyectos desarrollados con Python/Pygame. Este comportamiento puede explicarse por la complejidad inherente de las dependencias requeridas por los entornos de ejecución asociados a Unity, las cuales incrementan la cantidad de bibliotecas presentes en la imagen y limitan las posibilidades de eliminación de componentes sin afectar la funcionalidad de la aplicación.

El caso de 2D Platformer resulta especialmente relevante desde una perspectiva de investigación. Aunque presentó el menor número de vulnerabilidades iniciales (96 hallazgos), también registró el menor Índice de Mejora de Hardening (19,79 %). Este resultado evidencia que una menor cantidad de vulnerabilidades iniciales no necesariamente implica una mayor capacidad de mitigación. En este caso, la mayoría de los hallazgos identificados correspondieron a vulnerabilidades asociadas a componentes esenciales de la imagen base y dependencias requeridas para la ejecución del videojuego, limitando la efectividad de las actividades de hardening tradicionales.

El comportamiento observado en el Hardening Score complementa los resultados obtenidos mediante el IMH. Todos los proyectos experimentaron incrementos en su puntuación de seguridad después del proceso de hardening, lo que confirma una mejora objetiva de la postura de seguridad de las imágenes analizadas. Sin embargo, los valores finales obtenidos permanecen significativamente por debajo de los alcanzados durante el análisis estático de los Dockerfiles. Mientras los Dockerfiles hardenizados alcanzaron puntuaciones cercanas al 100%, las imágenes Docker finales oscilaron entre 13,1% y 38,99 %.

Este resultado constituye uno de los hallazgos más importantes de la investigación, ya que evidencia que la corrección de configuraciones inseguras durante la fase de construcción no garantiza la eliminación de las vulnerabilidades presentes en los componentes heredados de las imágenes base. En consecuencia, la seguridad de una imagen Docker depende no solamente de la calidad del Dockerfile utilizado para construirla, sino también de la seguridad de los paquetes, bibliotecas y componentes externos incorporados durante el proceso de construcción.

Desde una perspectiva práctica, los resultados obteni-

dos demuestran que DockerHard permite identificar y cuantificar de forma objetiva tanto las mejoras derivadas de las actividades de hardening como el riesgo residual que permanece presente después de su aplicación. Asimismo, la combinación de análisis estático, análisis dinámico, priorización mediante FIRST-EPSS y métricas cuantitativas como Hardening Score e IMH proporciona una visión integral de la postura de seguridad de los contenedores, permitiendo orientar de forma más precisa los esfuerzos de mitigación y fortalecimiento de imágenes Docker utilizadas en entornos Linux.

IV. CONCLUSIONES

Los resultados experimentales permiten concluir que DockerHard es una herramienta viable para apoyar procesos de análisis de seguridad y hardening de imágenes Docker en videojuegos ejecutados sobre Linux. Su principal aporte radica en integrar análisis estático de Dockerfiles, análisis dinámico de imágenes, priorización mediante FIRST-EPSS y métricas cuantitativas como el Hardening Score y el Índice de Mejora de Hardening (IMH).

El análisis estático evidenció la mayor efectividad del proceso de hardening. Los hallazgos en Dockerfiles se redujeron de 95 a 2, lo que representa una mejora del 97,89 %. Además, se eliminaron completamente los hallazgos de severidad Critical, High y Medium, demostrando que las configuraciones inseguras asociadas a privilegios excesivos, secretos embebidos, paquetes innecesarios y ausencia de buenas prácticas pueden corregirse de forma significativa desde la fase de construcción.

El análisis dinámico mostró también una reducción relevante de vulnerabilidades en las imágenes Docker, pasando de 4055 a 683 hallazgos, equivalente a una mejora global del 83,16 %.

No obstante, los resultados evidencian que el hardening aplicado sobre Dockerfiles no elimina por completo el riesgo presente en las imágenes, debido a vulnerabilidades heredadas desde imágenes base, bibliotecas del sistema operativo y dependencias de terceros.

Los casos experimentales mostraron diferencias importantes según la tecnología empleada. Los proyectos desarrollados con Python/Pygame presentaron los mayores índices de mejora, destacándose Aeroblasters con un IMH de 94,47 % y Memory Puzzle con 90,08 %. En contraste, los proyectos Unity conservaron mayor riesgo residual, especialmente 2D Platformer, con un IMH de 19,79 %, lo cual evidencia que la complejidad del entorno de ejecución influye directamente en la efectividad del hardening.

La integración de información de vulnerabilidades, severidades, EPSS y recomendaciones basadas en estándares como OWASP Docker Security Cheat Sheet, NIST SP 800- 190 y CIS Docker Benchmark permitió construir un proceso de evaluación alineado con buenas prácticas reconocidas internacionalmente para la protección de contenedores.

V. REFERENCIAS

[1].NetRise, Software Supply Chain Visibility and Risk Report, NetRise, 2024.

[2].Red Hat, State of Kubernetes and Container Security Report, Red Hat, 2024.

[3].Cloud Native Computing Foundation (CNCF), CNCF Annual Survey Report, CNCF, 2023.

[4].M. Banyan et al., “An Empirical Study of Vulnerabilities in Docker Images,” IEEE Access, vol. 8, 2020.

[5].R. Shu, X. Tian, and D. Williams, “A Study of Security Vulnerabilities on Docker Hub,” in Proc. ACM Workshop on Container Technologies and Container Clouds (CODASPY), 2017.

[6].M. Souppaya, J. Morello, and K. Scarfone, Application Container Security Guide, NIST Special Publication 800-190. Gaithersburg, MD, USA: National Institute of Standards and Technology, Sep. 2017. doi: 10.6028/NIST.SP.800-190.

[7].Center for Internet Security, CIS Docker Benchmark, CIS, 2024.

[8].OWASP Foundation, OWASP Docker Top 10. [Online]. Available: https://owasp.org/www-project-docker-top-10/

[9].J. Humble and D. Farley, Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation. Boston, MA, USA: Addison-Wesley, 2011.

[10].B. Chess and G. McGraw, “Static Analysis for Security,” IEEE Security & Privacy, vol. 2, no. 6, pp. 76–79, Nov.–Dec. 2004.

[11].N. Ayewah, W. Pugh, J. D. Morgenthaler, J. Penix, and Y. Zhou, “Using Static Analysis to Find Bugs,” IEEE Software, vol. 25, no. 5, pp. 22–29, Sep.–Oct. 2008, doi: 10.1109/MS.2008.130.

[12].A. Mills, J. R. White, and P. Legg, “Longitudinal Risk-based Security Assessment of Docker Software Container Images,” Computers & Security, vol. 132, 2023.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

miniLOGO: A Multilingual, Multi-Turtle TurtleGraphics Environment for Arts-Integrated Computational Thinking in High School

miniLOGO: Un entorno TurtleGraphics multilingüe y multitortuga para el pensamiento computacional integrado a las artes en educación secundaria

Jesus Insuasti [1]*, Edwin Insuasty [2]

[1] Galeras.NET Research Group, Universidad de Nariño, Pasto, Nariño, Colombia. insuasti@udenar.edu.co

[2] Galeras.NET Research Group, Universidad de Nariño, Pasto, Nariño, Colombia. edwin@udenar.edu.co

* Autor de correspondencia: insuasti@udenar.edu.co

Citación:. J. Insuasti y E. Insuasty, "miniLOGO: A Multilingual, Multi-Turtle TurtleGraphics Environment for Arts-Integrated Computational Thinking in High School," Revista de Ingeniería, vol. 4, no. 1, pp. 13-22, Nov. 2026.

Abstract—This paper presents miniLOGO, a lightweight turtlegraphics environment designed for arts-integrated computational thinking (CT) activities in high school, and reports a descriptive pilot based on 11 group guide submissions (33 worksheet pages) and contextual visual records. The study does not test learning gains or causal effects. At the groupartifact level, all 11 submissions contained movementplanning, geometric-procedure, and repetition indicators; 8/11 contained visible correction or revision marks; and 10/11 contained a positive experience mark, while one reflection was not reliably interpretable. These observations are compatible with sequencing, iteration, debugging, and geometry-to-code translation, but they do not establish improvement in CT attributable to miniLOGO. The revised system description separates implemented capabilities from those examined in the pilot: Spanish, English, and Portuguese command vocabularies and support for up to five turtles are implemented features, whereas the pilot evidence primarily concerns Spanish commands, single-pattern geometric procedures, repetition, and immediate visual feedback. The methodology specifies the unit of analysis, inclusion rules, coding decisions, reliability requirements, and ethical safeguards. The findings are therefore interpreted as exploratory evidence about classroom use and artifact characteristics, motivating a larger study with validated CT measures and independently coded data.

Keywords—arts integration, computational thinking, high school,multilingual programming education, turtle graphics

Resumen—Este artículo presenta miniLOGO, un entorno ligero de gráficos de tortuga diseñado para actividades de pensamiento computacional (PC) integradas a las artes en la escuela secundaria, y reporta un piloto descriptivo basado en 11 entregas de guías grupales (33 páginas de hojas de trabajo) y registros visuales contextuales. El estudio no evalúa ganancias de aprendizaje ni efectos causales. A nivel de los artefactos grupales, las 11 entregas contuvieron indicadores de planificación de movimiento, procedimientos geométricos y repetición; 8 de 11 mostraron marcas visibles de corrección o revisión; y 10 de 11 incluyeron una marca de experiencia positiva, mientras que una reflexión no pudo interpretarse de manera confiable. Estas observaciones son compatibles con la secuenciación, la iteración, la depuración y la traducción de geometría a código, pero no establecen una mejora en el PC atribuible a miniLOGO. La descripción revisada del sistema separa las capacidades implementadas de las examinadas en el piloto: los vocabularios de comandos en español, inglés y portugués, junto con el soporte para hasta cinco tortugas, son características implementadas; mientras que la evidencia del piloto se centra principalmente en comandos en español, procedimientos geométricos de un solo patrón, repetición y retroalimentación visual inmediata. La metodología especifica la unidad de análisis, las reglas de inclusión, las decisiones de codificación, los requisitos de confiabilidad y las salvaguardas éticas. Por lo tanto, los hallazgos se interpretan como evidencia exploratoria sobre el uso en el aula y las características de los artefactos, lo que motiva un estudio más amplio con medidas de PC validadas y datos codificados de forma independiente.

Palabras claves—Análisis de vulnerabilidades, Docker, Hardening, Seguridad de Contenedores, Video Juegos Linux.

I. INTRODUCTION

Computational thinking has become a central goal of K-12 computing education because it names a family of practices that support problem formulation and solution design beyond writing syntactically correct code: decomposition, pattern recognition, abstraction, algorithm design, and debugging [1]- [4]. Recent syntheses confirm the rapid expansion of CT research while also emphasizing continuing variation in definitions, instructional designs, and assessment practices [16], [22]. For classroom studies, this makes a clear distinction essential between evidence that a CT-related practice is visible in student work and evidence that an intervention caused measurable improvement in CT.

The LOGO tradition offers a durable response to this tension. Papert's [5] turtle graphics framed programming as a way to think with objects, movement, and mathematics. Rather than treating code only as a formal language, turtle graphics lets learners reason through embodied directions: move forward, turn, repeat, draw, and revise. The screen becomes a feedback surface where mathematical ideas and aesthetic choices can be tested. Modern creative coding environments such as Scratch extend this tradition by emphasizing low-floor, high-ceiling participation, remixing, and visible results [6].

miniLOGO builds on this lineage with a focused high-school purpose: to provide a small turtle-graphics command vocabulary that can be used as a design medium for geometric and artistic tasks. The environment implements commands in Spanish, English, and Portuguese and supports up to five turtles in one composition. These design features are motivated by the broader educational-language principle that novice environments should favor simplicity, readability, and immediate feedback [20]. The present pilot, however, did not compare languages and did not systematically test multi-turtle work; the student artifacts available for analysis predominantly use Spanish commands and emphasize single-pattern geometric constructions.

This study updates the relevant literature, describes the miniLOGO feature set and the subset actually exercised in the pilot, specifies the classroom artifact-analysis procedure, and reports descriptive findings from student guide submissions, screenshots, and contextual classroom records. The research questions concern observable uses and artifact characteristics rather than pre/post change or treatment effects.

Across the supplied materials, the recurring descriptive pattern is a progression from movement planning to repeated geometric procedures and brief reflection. The classroom records show paper planning used alongside command entry and geometric screen output. These materials permit identification of artifact-level indicators compatible with sequencing, repetition, correction, and geometric construction. They do not by themselves establish that students improved their CT, nor that any observed performance is attributable specifically to miniLOGO.

II. LITERATURE REVIEW

Computational thinking in K-12 education

CT is commonly discussed as an approach to formulating problems and expressing solutions through computational steps, but the literature cautions against treating every programming success as evidence of broad CT development [3], [7]. Recent systematic reviews of K-12 computer science education report generally positive associations between programming experiences and CT-related outcomes while also documenting substantial heterogeneity in interventions, duration, and measures [16]. A 2026 meta-review similarly notes that the field has grown quickly but remains methodologically diverse, reinforcing the need for precise construct definitions and restrained inference [22]. In this paper, CT is operationalized only through observable artifact indicators tied to sequencing, iteration, decomposition, and revision; no claim of measured CT gain is made.

Turtle graphics, constructionism, and visual feedback

Papert's constructionist approach [5] emphasized learning by making public, shareable artifacts. Turtle graphics is a paradigmatic constructionist tool because it links code to visible motion and drawing. Students can treat each line of code as a design decision, observe its consequence, and revise accordingly. This feedback loop is valuable for beginners because it externalizes errors. A wrong angle, missing repetition, or misplaced movement produces a visible mismatch between the intended and generated image. That mismatch can motivate debugging without requiring the teacher to diagnose every error verbally.

Arts integration and creative coding

Arts integration can frame programming as a medium for expression rather than only as a technical exercise. Earlier creative-media and STEAM work emphasizes personally meaningful artifacts, aesthetic choice, and iterative refinement [6], [9], [10]. More recent cross-curricular research combining CT with arts and design has documented creativity, tinkering, and debugging as practices that can become visible in project work [17]. Research on creative-coding assessment likewise shows the value of explicit rubrics and rater agreement when open-ended artistic artifacts are used as data [23]. For miniLOGO, geometry and generative art provide the task

context in which procedures, repeated patterns, and visual revision can be observed.

Multilingual and beginner-friendly programming environments

Programming environments designed for novices often reduce unnecessary syntactic burden, emphasize readability, and provide rapid feedback [11], [12], [20]. miniLOGO adds a multilingual design choice: its command vocabulary is implemented in Spanish, English, and Portuguese. Current research on multilingual computer science education indicates that language is an important access and participation dimension, and recent bilingual-programming work suggests that native-language support warrants direct empirical study [21], [24]. The present pilot does not provide such a comparison. Its artifacts primarily contain Spanish commands such as avanzar, derecha, izquierda, repetir, and fin; therefore, multilingualism is treated here as an implemented system capability, not as an evaluated learning effect.

Assessing computational thinking through artifacts

Assessing CT remains challenging because CT can be represented in products, processes, explanations, and test performance, and different instruments capture different dimensions [13], [14]. Recent assessment syntheses continue to report wide variation in tests, questionnaires, observations, interviews, and product analysis [19]. Artifact-based assessment is useful for creative coding because programs, drawings, annotations, and revisions can make specific practices visible, but product evidence should not be interpreted as proof of change unless the design includes an appropriate baseline or comparison. Accordingly, this study codes only observable indicators in group artifacts and reports counts descriptively.

Current state of the art and study gap

Recent literature points to four issues that directly shape the present study. First, interdisciplinary CT work is increasingly common, including mathematics and arts/design contexts [17], [18]. Second, programming-language design for education still prioritizes simplicity and readability but requires explicit reporting of the features learners actually use [20]. Third, multilingual computing is an emerging equity-oriented research area rather than a settled evidence base [21], [24]. Fourth, assessment scholarship increasingly expects transparent coding rules and reliability checks for project-based evidence [19], [23]. The gap addressed by this pilot is therefore narrow: it documents how a small multilingual turtle-graphics environment was used in one arts-integrated high-school activity and identifies artifact-level patterns that can inform a more rigorous subsequent study.

III. PROBLEM STATEMENT AND RESEARCH QUESTIONS

The instructional problem addressed by miniLOGO is the gap between the recognized importance of CT and the difficulty of making CT accessible, creative, and assessable for high-school novices. Many introductory programming tools either demand too much syntax too early or separate programming from meaningful classroom domains. Visual art provides a promising context, but teachers need lightweight tools and classroom- ready learning sequences that connect code, geometry, and artistic decision-making.

This pilot paper is guided by the following research questions:

RQ1. How do high-school students use miniLOGO commands to express movement, repetition, and geometric pattern in arts-integrated activities?

RQ2. What CT practices are visible in students' solved guides, generated artifacts, and classroom photographs?

RQ3. What initial evidence of perceived usability and engagement appears in student reflections from the classroom pilot?

IV. SYSTEM AND PEDAGOGICAL DESIGN

miniLOGO is a lightweight, text-command turtle-graphics environment intended for introductory geometric and artistic programming. The available interface captures show a command-entry/editor area alongside a drawing canvas, so each executed movement-turn sequence produces an immediately visible geometric result. The language is deliberately compact, with 25 commands, and the interface also exposes controls for visual properties such as color, line thickness, background, and turtle visibility. This design follows the educational-language rationale that a novice tool should minimize incidental complexity while keeping program behavior readable and inspectable [20].

Command and graphics functions

The pilot materials document movement and orientation commands, including avanzar (forward), derecha (right), and izquierda (left), together with repetir ... fin as the principal visible control structure for geometric repetition. These commands are sufficient to express paths, regular polygons, stars, rosette-like figures, and repeated line compositions. The interface also implements drawing-style controls for color, line thickness, background, and turtle visibility. Because the archived pilot

artifacts do not systematically vary those style controls, they are described as available functions rather than analyzed outcomes.

Multilingual command model

The same small command vocabulary is implemented in Spanish, English, and Portuguese so that introductory activities can be written in the classroom language. In the analyzed pilot corpus, the commands visible in student work are overwhelmingly Spanish. Consequently, the study does not test whether multilingual command support changes comprehension, participation, confidence, or error rates. Any such effect remains a question for a comparative multilingual study [21], [24].

Multi-turtle capability

miniLOGO supports up to five turtles within a composition, allowing separate drawing agents to be positioned and used as digital brushes. This capability can support layered or coordinated artwork, but the current guide tasks and coded artifacts focus mainly on single-pattern geometric reasoning. Multi-turtle coordination was therefore not an evaluated variable in this pilot and no learning claim is made for that feature.

Implemented features versus pilot-evaluated features

The distinction is important for interpreting the paper. Implemented features include the 25-command vocabulary, Spanish/English/Portuguese command forms, up to five turtles, visual styling controls, control-structure formatting, and immediate turtle-graphics output. Features evidenced in the pilot are narrower: movement and turn commands, repetition, geometric construction, Spanish-language command use, visible revision marks on paper, and screen-based geometric feedback. Multilingual comparison, multi-turtle coordination, and systematic use of color/background/thickness were not evaluated and are not presented as findings.

The instructional sequence represented in the guide materials begins with turtle movement and orientation, advances to polygon and star-like constructions using repetition, and ends with more decorative geometric patterns and a brief experience reflection. The pedagogical mapping is prospective rather than causal: movement-turn sequences are compatible with algorithmic sequencing; repeated blocks with iteration and abstraction; and visible corrections with testing and revision. A representative miniLOGO output is shown in Fig. 1

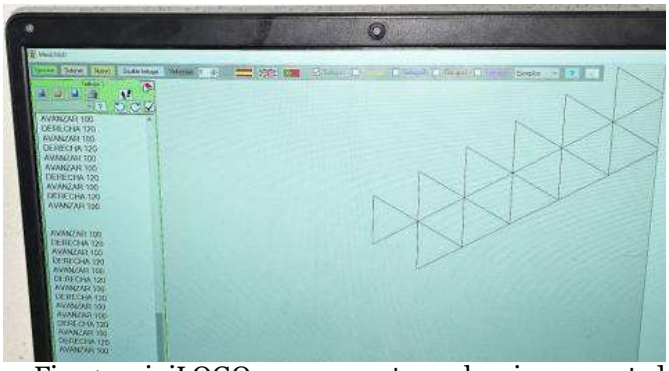


Fig. 1. miniLOGO screen capture showing repeated commands producing a geometric pattern.

V. METHODOLOGY

This study uses an exploratory descriptive pilot design. Its purpose is to document the classroom use of miniLOGO and characterize observable features in the available artifacts, not to estimate learning gains, compare treatments, or attribute outcomes to the software. The unit of analysis for the coded results is the group guide submission. Student artifacts are summarized with frequencies only. Photographs and screen captures are used as contextual records and are not scored as additional student observations. The miniLOGO abstract and system description are used only to describe intended functionality and are excluded from the artifact counts. Because there is no pretest, posttest, control group, or validated CT outcome measure, all findings are framed as descriptive and non-causal.

TABLE I

DATA CORPUS AND ROLE IN THE DESCRIPTIVE PILOT ANALYSIS

Data source	Quantity	Role in analysis
Student guide submissions	11 group documents; 33 worksheet pages	Primary coded corpus. All available group submissions were reviewed at the group level.
Classroom / screen visual records	7 supplied images	Contextual description only; not added to artifact-level counts. Published figures use non-identifying excerpts/screens.
miniLOGO system description	1 supplied abstract / description	System-feature description only; excluded from student evidence and coding counts.

Participants and classroom context

The pilot involved 24 students enrolled in nineth grade at Escuela Normal de Pasto. Participants were approximately 13 years old and worked in pairs producing 11 group guide submissions. The classroom language was Spanish, and the activity was facilitated by the teacher in informatics and the researchers in this paper. Individual names are not reported, and the analysis uses group-level evidence only. These details must be verified from the original classroom records before submission.

Intervention duration and procedure

The intervention consisted of one session of approximately 45 minutes, conducted over the course of informatics. The procedure represented in the guide materials had five stages: (1) introduction to turtle position, movement, and turning; (2) paper-based planning of short paths; (3) translation of paths into miniLOGO commands; (4) use of repetition to construct polygons, stars, and decorative geometric patterns; and (5) a brief smile-scale and open-comment reflection. Students moved between printed guides and the miniLOGO interface, comparing planned commands with the resulting geometry. No pre/post CT test was administered.

Artifact inclusion and figure selection

The coded corpus consisted of all 11 group guide submissions available for the pilot analysis; no available guide was excluded, and all 33 worksheet pages within those submissions were reviewed. Each group document contributed one binary code per indicator, regardless of how many qualifying marks occurred on its pages. Visual records were used only to document the paper-screen context. Figure excerpts were selected after coding as illustrative examples of categories already present in the rubric; their selection did not create additional cases or change the reported counts. Any page or image containing a direct identifier was to be excluded or cropped before publication.

Artifact coding protocol

Each guide was coded as one group-level case using five indicators defined a priori from the task structure. Codes were binary (1 = indicator present; 0 = indicator absent), except that an illegible or missing reflection response was coded NA rather than negative. A code of 1 required visible evidence in the artifact itself; inferred intentions were not coded. Multiple examples of the same indicator within one guide still counted as one case. The indicators were: (a) movement planning, (b) polygon/star or related geometric procedure, (c) repetition/iteration, (d) debugging/correction marks, and (e) positive experience mark. Table II provides the decision rules. The resulting frequencies are descriptive artifact counts, not psychometric scores and not estimates of individual student mastery.

Inter-rater reliability and disagreement resolution

For the revised analysis, we independently code all 11 group submissions using the finalized rubric before seeing each other's decisions. A short calibration pass should be used only to clarify the codebook before the formal independent round. Report percent agreement and Cohen's kappa for each binary indicator, for example: movement planning, geometric procedure, repetition, debugging, and positive experience. After reliability statistics are calculated, disagreements may be discussed and resolved by consensus for the final descriptive counts. These values must be based on a real independent recoding and must not be invented.

Ethical considerations

All instruments and procedures used in this research were reviewed and approved by the Ethics Committee of the University of Nariño, ensuring strict compliance with current research ethics regulations. The protection, anonymity, and well-being of the participants were always guaranteed, especially for underage participants. Analysis reports only group-level data; names and other direct identifiers are omitted; and published worksheet excerpts are cropped or anonymized. In this revised manuscript, the classroom photograph that showed part of a student's body has been removed, leaving only non-identifying worksheet excerpts and screen captures. If institutional policy determined that formal review was not required, the manuscript should state the basis for that determination rather than implying approval.

TABLE II
DESCRIPTIVE CODING RUBRIC FOR STUDENT GUIDE ANALYSIS

Indicator	Operational decision rule in guide artifacts	CT-related interpretation
Movement planning	Code 1 if a guide contains at least one explicit route or command/path representation combining forward movement with one or more turns; otherwise 0.	Compatible with algorithmic sequencing and spatial reasoning.

Indicator	Operational decision rule in guide artifacts	CT-related interpretation
Polygon/star procedures	Code 1 if commands or annotations specify a polygon, star, rosette, circle-like construction, or comparable geometric procedure; otherwise 0.	Compatible with decomposition and geometry-to-code translation.
Repetition iteration	Code 1 if a repeat/ loop instruction or explicitly delimited repeated block encodes a recurring movement-turn pattern; manual rewriting alone is not sufficient.	Compatible with iteration, pattern generalization, and abstraction.
Debugging correction marks	Code 1 if a visible revision changes an angle, step count, command, or proposed alternative. Uninterpretable erasure alone is not counted.	Compatible with testing, debugging, and iterative refinement.
Positive experience mark	Code 1 for a positive-side smile-scale selection and/or clearly positive written comment; neutral/ negative = 0; missing/illegible = NA.	Descriptive self-report of perceived experience; not a validated engagement score.

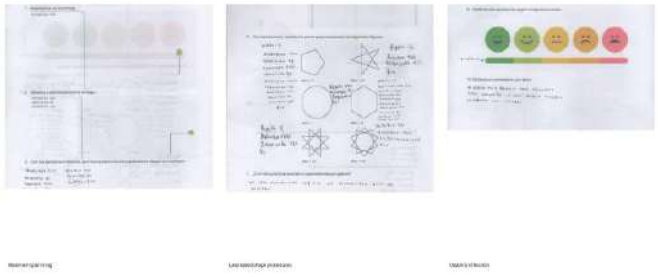


Fig. 2. Anonymized illustrative excerpts from the coded corpus showing movement planning, loop-based geometric procedures, and reflection. Excerpts are examples of rubric categories and were not counted as additional cases.

VI. FINDINGS

The artifact review produced five descriptive frequencies. All 11 group submissions contained the movement-planning indicator, all 11 contained at least one geometric-procedure indicator, and all 11 contained repetition/ iteration evidence under the rubric definition. Eight of 11 submissions contained visible correction or revision marks. Ten of 11 submissions contained a positive experience mark; the remaining reflection was not reliably interpretable and was treated as missing rather than negative. These frequencies describe what is visible in the artifacts. They are not standardized scores, do not represent independent individual observations, and do not establish improvement attributable to miniLOGO. Table III summarizes the observations, and Fig. 3 reports the group-artifact counts.

TABLE III
SUMMARY OF DESCRIPTIVE FINDINGS FROM THE STUDENT GUIDE ANALYSIS

Finding	Evidence observed	Descriptive interpretation
Movement sequencing indicator	11/11 group / submissions included a completed movement or path task.	The artifacts contain explicit mappings between visible paths and command sequences.
Iteration geometry	11/11 submissions included repeat/ loop- based or explicitly repeated geometric procedures.	Loop notation co-occurs with regular geometric construction in the guide tasks.
Visible revision marks	8/11 guides contained changed angles, overwritten commands, corrected counts, or alternatives.	The marks are compatible with testing and correction, but the reason for each change cannot be inferred from the artifact alone.
Immediate visual output	Screen records show geometric output displayed after command entry.	The records document the feedback mechanism; they do not isolate its effect on learning.
Positive experience response	10/11 guides contained a positive mark/comment; 1 interpretable response was not reliably interpretable.	The brief self-report was generally favorable among responses but is not a validated engagement measure.

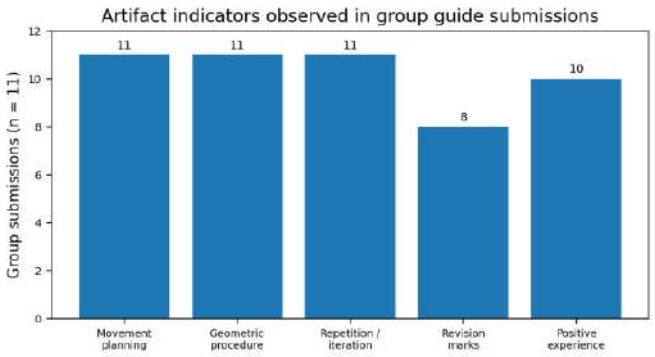


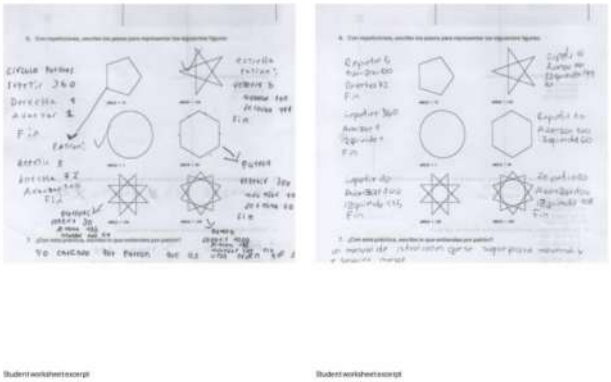
Fig. 3. Aggregate artifact-level indicators in 11 group guide submissions. Counts are descriptive observations, not standardized CT scores or estimates of learning gain.

Student solutions and CT practices

The solved guides most clearly document the relationship between geometry and command repetition. Students wrote short procedures around printed polygons and stars, often pairing a repetition count with movement distance and turn angle. Under the coding protocol, this constitutes visible evidence of a repeated procedural rule. It is compatible with CT interpretations of iteration and abstraction, but the artifact alone cannot show whether the learner newly acquired that idea during the intervention.

The guide tasks also structure larger visual goals into smaller units: forward movement, turn, repeated block, finished figure, and screen composition. Some student entries follow this structure explicitly. These features are compatible with decomposition and geometry-to-code translation. Because no baseline task or verbal explanation was collected, the study does not estimate change in students' decomposition skill.

Visible revision marks occurred in 8 of the 11 group guides. Examples include changed angle values, rewritten commands corrected repetition counts, and annotated alternatives. Such marks are consistent with testing and revision, but they do not reveal whether a change resulted from screen feedback, peer discussion, teacher prompting, or simple rewriting. Figure 4 therefore illustrates the form of the evidence rather than proving a debugging process or learning gain.



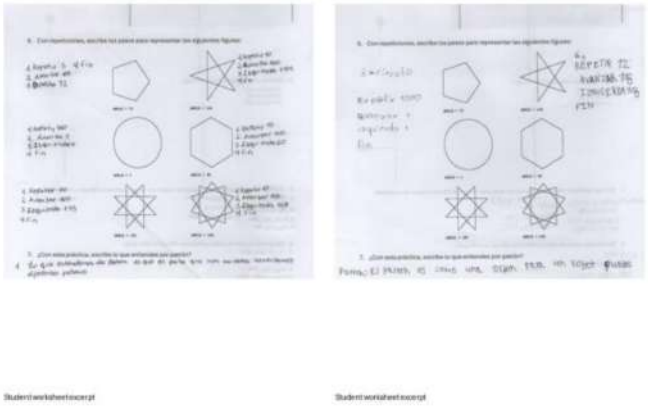


Fig. 4. Anonymized examples of geometric procedures and visible revisions in group guide submissions. The examples illustrate coded categories; causal or developmental interpretation is not inferred from the images alone.

Classroom workflow

The visual records document a paper-to-screen classroom arrangement: printed guides contain planned commands and geometric annotations, while miniLOGO screen captures display the resulting line patterns. These records are used only to characterize the learning environment and representations available to students. They are not time-ordered process logs and do not establish how often particular behaviors occurred or whether the visual feedback caused conceptual change. Figure 5 shows a non-identifying miniLOGO screen record from the pilot.

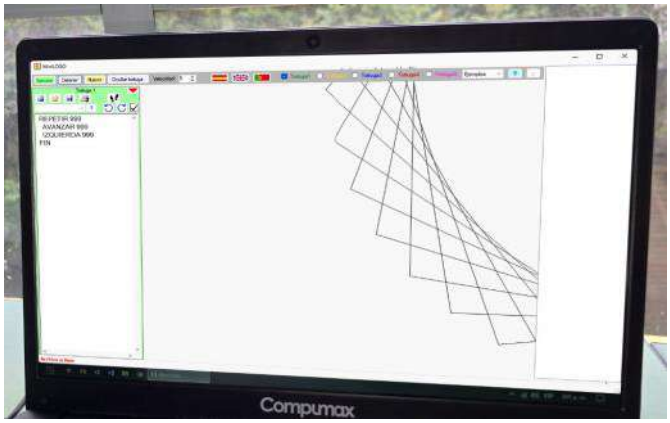


Fig. 5. Non-identifying miniLOGO screen record from the pilot showing geometric output after command entry.

Usability and engagement reflections

The reflection pages provide a brief descriptive self-report of the experience. Ten of the 11 group submissions contain a positive-side smile-scale selection and/or a clearly positive written comment. One remaining reflection is not reliably interpretable and is treated as missing

rather than negative. Comments include brief descriptions such as the activity being fun or the explanation being good. Because the item was short, group-based, and not validated as an engagement or usability instrument, these responses should be reported only as favorable classroom reactions, not as evidence of improved motivation or usability in a psychometric sense.



Fig. 6. Anonymized reflection excerpts illustrating positive experience markings and brief comments. These are descriptive classroom responses, not validated engagement scores.

VII. DISCUSSION

Within the limits of this descriptive pilot, miniLOGO provides a compact visual context in which movement, turning, repetition, and geometric output can be represented with a small command vocabulary. The artifacts show that these forms were used in the classroom, particularly through Spanish commands such as avanzar, derecha, izquierda, repetir, and fin. This is consistent with the design goal of reducing unnecessary syntactic overhead, but the pilot does not compare miniLOGO with another environment and therefore cannot establish that the tool is easier to learn or that it improves CT.

The multilingual and multi-turtle capabilities should be separated from the pilot findings. The language implements Spanish, English, and Portuguese command forms, yet the coded corpus provides evidence almost exclusively for Spanish use. Similarly, the system can support up to five turtles, but the guide artifacts focus mainly on single-pattern geometry rather than coordinated multi-turtle compositions. These features are technically implemented and pedagogically promising, but their educational value remains unevaluated in the present study.

The arts-integrated approach is also relevant to assessment design. Project artifacts can expose relationships among plan, code, geometry, visible revision, and reflection, but recent assessment research emphasizes that open-ended products require explicit rubrics and reliability procedures [19], [23]. In the present study, the rubric is intentionally narrow and binary. A revised star angle or looped polygon is treated as an observable indicator compatible with a CT practice, not as proof of individual mastery or improvement.

Angle-related annotations and revisions are especially visible in the star and rosette tasks. These marks identify angle selection as a useful target for future observation and interview, but the current artifacts cannot determine whether students held a misconception, why they changed a value, or which classroom interaction prompted the change. A stronger study could pair screen/program logs with think-alouds or short interviews to distinguish geometric reasoning from trial-and-error editing.

What the pilot does and does not establish

The pilot establishes that the available group artifacts contain observable examples of sequencing, repeated geometric procedures, correction marks, and generally positive brief reflections. It also documents that miniLOGO's immediate turtle-graphics output was present in the classroom setting. It does not establish pre-to-post improvement in CT, the superiority of miniLOGO over another tool, a causal effect of visual feedback, a benefit of multilingual commands, or an effect of multi-turtle programming. Those claims require additional measures and a stronger design.

VIII. LIMITATIONS

The study has several limitations. First, it is a small single- context pilot with 11 group submissions and no pretest, posttest, or comparison condition. Second, the unit of analysis is the group artifact, so individual learning and individual contributions cannot be separated. Third, handwritten pages vary in legibility, and the reflection item is not a validated usability or engagement instrument. Fourth, the available student artifacts mainly document Spanish command use and do not systematically exercise the multi-turtle or multilingual features. Fifth, the visual records are contextual rather than continuous process data. Finally, the credibility of the artifact counts depends on completion and reporting of the independent recoding and inter-rater agreement described in the methodology. These constraints limit the findings to descriptive feasibility and artifact patterns in this pilot context.

IX. IMPLICATIONS FOR PRACTICE AND FUTURE RESEARCH

For practice, miniLOGO can be positioned as a short exploratory module linking geometric drawing, programming notation, and visual-art tasks. A plausible sequence is to introduce movement, ask students to predict output, build regular figures with repetition,

compare turn-angle choices, generate an artwork, and reflect on the process. The current study documents this sequence as a classroom design possibility; it does not demonstrate that the sequence is more effective than alternatives.

For research, the next step is to retain the explicit codebook and independent double-coding while adding stronger outcome and process measures. Future studies should collect original miniLOGO code files, generated images, screen/program logs, brief student explanations, and a validated CT assessment aligned with sequence, iteration, abstraction, and debugging. A pre/post or comparison design would be required to examine change attributable to the intervention. Where artifact coding is used, rater agreement should be reported before consensus resolution.

The multilingual dimension deserves a separate research design. Because miniLOGO implements Spanish, English, and Portuguese command forms, future work can compare otherwise equivalent tasks across language conditions or examine translanguaging choices in authentic multilingual classrooms. Recent studies show that language support in computing is an active research area [21], [24], but the present pilot provides no comparative evidence on language effects. The same principle applies to multi-turtle programming: coordinated turtle use should be evaluated with tasks designed specifically to elicit that feature.

X. CONCLUSION

This exploratory pilot documents a feasible arts-integrated use of miniLOGO in which group guide submissions contain visible movement sequences, repeated geometric procedures, correction marks, and generally favorable brief experience responses. The environment combines a small turtle-graphics command set with immediate visual output and implements multilingual and multi-turtle capabilities. The pilot evidence, however, is narrower than the system feature set and should remain descriptive: it does not demonstrate improvement in computational thinking or effects attributable to miniLOGO. The main contribution of the revised study is therefore a clearly bounded system description, an explicit artifact-coding framework, and a set of classroom observations that can inform a larger, ethically documented, independently coded, and outcome-validated investigation.

APPENDIX A
ARTIFACT CODING SUMMARY

Table IV summarizes the group-artifact counts reported in the pilot. The entries are binary presence/absence observations under the rubric in Table II; they are not test scores.

TABLE IV
ARTIFACT CODING SUMMARY

Indicator	Count observed	Notes
Completed movement task	11-11	All group submissions contained movement/path-planning evidence.
Polygon/star procedure	11-11	All submissions contained at least one procedural geometric construction.
Repetition iteration	11-11	Repeat/loop or explicitly delimited repeated patterns were visible under the rubric rule.
Debugging correction marks	8-11	Visible revisions changed angles, counts, commands, or alternatives.
Positive experience mark	10 positive; 1 NA	Ten submissions contained a positive mark/comment; one reflection was not reliably interpretable.

APPENDIX B
RECOMMENDED FULL-PAPER NEXT COLLECTION

1) Export original miniLOGO code files, generated images, and timestamped screen/program logs for each team.

2) Collect a validated or validity-supported pre/post CT measure aligned with sequence, iteration, abstraction, debugging, and geometry; add a comparison condition if causal inference is a goal.

3) Use a validated usability or engagement questionnaire in addition to any brief classroom reflection item.

4) Conduct short student interviews or think-alouds asking learners to explain one geometric procedure and one revision decision.

5) Preserve independent double-coding with the finalized rubric and report percent agreement and Cohen's kappa before consensus resolution.

REFERENCES

[1].D. Barr, J. Harrison, and L. Conery, “Computational thinking: A digital age skill for everyone,” Learning & Leading with Technology, vol. 38, no. 6, pp. 20-23, 2011.

[2].V. Barr and C. Stephenson, “Bringing computational thinking to K-12: What is involved and what is the role of the

computer science education community?," ACM Inroads, vol. 2, no. 1, pp. 48-54, 2011, doi: 10.1145/1929887.1929905.

[3].S. Grover and R. Pea, "Computational thinking in K-12: A review of the state of the field," Educational Researcher, vol. 42, no. 1, pp. 38-43, 2013, doi: 10.3102/0013189X12463051.

[4].J. M. Wing, "Computational thinking," Communications of the ACM, vol. 49, no. 3, pp. 33-35, 2006, doi: 10.1145/1118178.1118215.

[5].S. Papert, Mindstorms: Children, Computers, and Powerful Ideas. New York, NY, USA: Basic Books, 1980.

[6].M. Resnick et al., "Scratch: Programming for all," Communications of the ACM, vol. 52, no. 11, pp. 60-67, 2009, doi: 10.1145/1592761.1592779.

[7].V. J. Shute, C. Sun, and J. Asbell-Clarke, "Demystifying computational thinking," Educational Research Review, vol. 22, pp. 142-158, 2017, doi: 10.1016/j.edurev.2017.09.003.

[8].D. Weintrop, E. Beheshti, M. Horn, K. Orton, K. Jona, L. Trouille, and U. Wilensky, "Defining computational thinking for mathematics and science classrooms," Journal of Science Education and Technology, vol. 25, no. 1, pp. 127-147, 2016, doi: 10.1007/s10956-015-9581-5.

[9].K. A. Peppler and Y. B. Kafai, "From SuperGoo to Scratch: Exploring creative digital media production in informal learning," Learning, Media and Technology, vol. 32, no. 2, pp. 149-166, 2007, doi: 10.1080/17439880701343337.

[10].E. Perignat and J. Katz-Buonincontro, "STEAM in practice and research: An integrative literature review," Thinking Skillsand Creativity, vol. 31, pp. 31-43, 2019, doi: 10.1016/j.tsc.2018.10.002.

[11].M. U. Bers, Coding as a Playground: Programming and Computational Thinking in the Early Childhood Classroom. New York, NY, USA: Routledge, 2018.

[12].S. Y. Lye and J. H. L. Koh, "Review on teaching and learning of computational thinking through programming: What is next for K-12?," Computers in Human Behavior, vol. 41, pp. 51-61, 2014, doi: 10.1016/j.chb.2014.09.012.

[13].K. Brennan and M. Resnick, "New frameworks for studying and assessing the development of computational thinking," in Proc. 2012 Annual Meeting of the American Educational Research Association, 2012. [Online]. Available: http://web.media.mit.edu/~kbrennan/files/Brennan_Resnick_AERA2012_ CT.pdf

[14].L. Werner, J. Denner, S. Campe, and D. C. Kawamoto, "The Fairy Performance Assessment: Measuring computational thinking in middle school," in Proc. 43rd ACM Technical Symposium on Computer Science Education, pp. 215-220, 2012, doi: 10.1145/2157136.2157200.

[15].A. Yadav, C. Stephenson, and H. Hong, "Computational thinking for teacher education," Communications of the ACM, vol. 60, no. 4, pp. 55-62, 2017, doi: 10.1145/2994591.

[16].S. J. Lee, G. M. Francom, and J. Nuatomue, "Computer science education and K-12 students' computational thinking: A systematic review," International Journal of Educational Research, vol. 114, Art. no. 102008, 2022, doi: 10.1016/j.ijer.2022.102008.

[17].L. Mannila, T. Leinonen, M. Bauters, and M. Veermans, "Student and teacher co-agency when combining CT with arts and design in a cross- curricular project," Computers and Education Open, vol. 4, Art. no. 100132, 2023, doi: 10.1016/j.caeo.2023.100132.

[18].H. Ye, B. Liang, O.-L. Ng, and C. S. Chai, "Integration of computational thinking in K-12 mathematics education: A systematic review on CT-based mathematics instruction and student learning," International Journal of STEM Education, vol. 10, Art. no. 3, 2023, doi: 10.1186/s40594-023-00396-w.

[19].S. Forsström and M. Bond, "Computational Thinking Assessment in Primary and Secondary Education: A Meta-synthesis of tools, Methods and Pedagogical Approaches," International Journal of Computer Science Education in Schools, vol. 6, no. 4, 2024, doi: 10.21585/ijcses.v6i4.208.

[20].M. Kölling, "Principles of Educational Programming Language Design," Informatics in Education, vol. 23, no. 4, pp. 823-836, 2024, doi: 10.15388/infedu.2024.29.

[21].E. Tshukudu, E. Dodoo, F. Hermans, and M. Mudongo, "Bilingual Programming: A Study of Student Attitudes and Experiences in the African Context," in Proc. 24th Koli Calling International Conference on Computing Education Research, 2024, pp. 1-11, Art. no. 42, doi: 10.1145/3699538.3699561.

[22].K. Astor, J. Rönnlund, C. Fawcett, and G. Gredebäck, "Computational thinking: A meta-review of systematic reviews and meta-analyses," Educational Research Review, vol. 52, Art. no. 100794, 2026, doi: 10.1016/j.edurev.2026.100794.

[23].J. Blake-West, G. Alrawashdeh, and M. Bers, "Validating a Creative Coding Rubric through expressive activities for elementary grades," Journal of Research on Technology in Education, vol. 57, no. 6, pp. 1350-1369, 2025, doi: 10.1080/15391523.2024.2398502.

[24].S. R. Jacob and M. Warschauer, "Elementary Computing for All: A Computational Thinking Curriculum for Multilingual Students," CATESOL Journal, vol. 35, no. 1, 2024, doi: 10.5070/B5.34830.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

How Spell-Checkers Work — A Non-Technical Experience in Spanish

Cómo funcionan los correctores ortográficos: Una experiencia no técnica en español

Jesus Insuasti [1]*, Edwin Insuasty [2]

[1] Galeras.NET Research Group, Universidad de Nariño, Pasto, Nariño, Colombia. insuasti@udenar.edu.co

[2] Galeras.NET Research Group, Universidad de Nariño, Pasto, Nariño, Colombia. edwin@udenar.edu.co

* Autor de correspondencia: insuasti@udenar.edu.co

Citación: J. Insuasti y E. Insuasty, "How Spell-Checkers Work — A Non-Technical Experience in Spanish," Revista de Ingeniería, vol. 4, no. 1, pp. 23-31, Nov. 2026.

Abstract—Spell checking has evolved from dictionary lookup and edit- distance methods to context-aware systems that also use statistical and neural language models. This article explains, for a non-specialist audience, how spelling correction can be understood through Levenshtein distance and illustrates the process with a pedagogical Spanish-language prototype. The revised experimentation section documents the prototype, the dictionary, the test procedure, and the observed candidate counts for a reproducible illustrative case. The discussion also distinguishes exhaustive candidate generation from the contextual ranking used by modern writing assistants.

Keywords—algorithm; correction; distance; spelling; words.

Resumen—La corrección ortográfica ha evolucionado desde la búsqueda en diccionarios y métodos de distancia de edición hasta sistemas conscientes del contexto que también utilizan modelos de lenguaje estadísticos y neuronales. Este artículo explica, para un público no especializado, cómo se puede entender la corrección ortográfica a través de la distancia de Levenshtein e ilustra el proceso con un prototipo pedagógico en lengua española. La sección de experimentación revisada documenta el prototipo, el diccionario, el procedimiento de prueba y los recuentos de candidatos observados para un caso ilustrativo reproducible. La discusión también distingue la generación exhaustiva de candidatos de la clasificación contextual utilizada por los asistentes de escritura modernos.

Palabras claves—Algoritmo, corrección, distancia, ortografía, palabras.

INTRODUCTION

For most people, the computer is a handy "black box" in everyday life. We consider it a black box because we do not need to know its components or understand how it works internally to facilitate numerous tasks, such as browsing the Internet, viewing and creating images or videos, listening to and recording sounds, and reading or writing texts, among others. One of the most practical functions that computers perform when we use word processors, such as Microsoft Word, Apache OpenOffice Writer, or Corel WordPerfect, is spell checking. Although it is not perfect, this feature is an excellent ally in writing texts.

People do not wonder how a word processor performs spell checking; they need it to work. However, there is an interesting mathematical background behind this task, related to number manipulation, as well as a computational background based on data handling [1], [2]. If you have ever wondered, "How does a spell checker work?" this article will answer that question.

Spell checking and text prediction, built into cell phones and tablets, are closely related technologies that share underlying mechanisms but differ in approach.

Both technologies rely on algorithms that analyze the text the user is typing. However, while spell checking focuses on identifying and correcting typing errors, text prediction aims to anticipate the words or phrases that the user might type next, facilitating faster and more efficient typing. Both features use linguistic resources and statistical models, but their purposes differ slightly.

Spell checking compares each typed word against a built-in dictionary and suggests corrections for words that do not match. A transparent baseline uses an edit-distance model, such as Levenshtein distance, to generate candidate corrections, while modern systems may also use contextual statistical or neural models to rank or generate suggestions [2], [3]. Text prediction, on the other hand, anticipates likely continuations based on context and usage patterns.

On mobile phones and tablets, both spell checking and text prediction have evolved through natural language processing, neural networks, and large pretrained language models. Recent grammatical-error-correction research describes a progression from rule-based and statistical approaches to neural machine translation and transformer-based systems [3]. Multilingual instruction-tuned models have also demonstrated strong results for grammatical error correction and other text-editing tasks, including Spanish [4]. These developments complement, rather than invalidate, edit distance: distance remains useful for interpretable candidate generation, while contextual models improve ranking and correction in complete sentences.

This article stems from the development of a learning object designed to teach students in the Bachelor's degree program in Computer Science at the University of Nariño about Similarity Algorithms. The objective of this learning object is to facilitate the understanding of the mathematical and computational foundations that support this theory and its practical applications. From this basis, we intend to expand the usefulness of the developed learning object by explaining, in a simple, accessible way for any reader, how machines perform spelling correction without resorting to complex mathematical formulas or intricate computational algorithms. However, it is necessary to mention scientific concepts such as mathematics and computing; their technical details will not be delved into, as the objective is to provide an understandable reflection on the spelling correction process accessible to all audiences.

BACKGROUND

Historical Evolution of the word Processor

It is often thought that the emergence of computers in the middle of the last century brought with it tools such as word processors and other utilities we use today. However, many of these functions existed in previous technological forms, long before the invention of the computer. A clear example is the use of mechanical tools to write and edit texts. Price and Urban classify word processors into three broad groups: mechanical, electronic, and computer-based [5].

Mechanical devices have their origins in the 1714 patent obtained by the English engineer Henry Mill, who invented a precursor to the typewriter. Although there is no detailed record of his invention, it was an important precursor. Later, in 1829, the American William Austin Burt developed a mechanical device called the "Typograph," which allowed documents to be written by selecting characters using a lever, leaving an ink mark on the paper [6]. Christopher Latham Sholes later designed a commercially successful typewriter whose keyboard evolved into the QWERTY layout [7]. These advances gave rise to the craft of typing, although at that time it was not possible to edit documents once they were written.

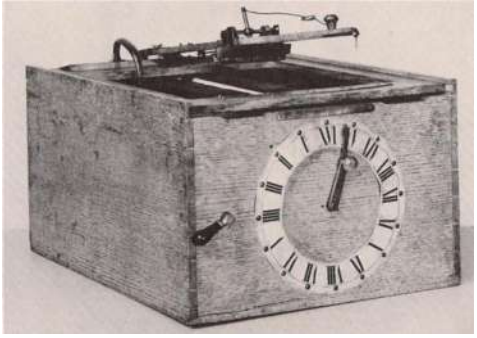


Fig. 1. William Austin Burt's typographer, the precursor to all typewriters. Source: [6].

Electronic devices enabled significant advances in text editing, making it easier to correct errors using special ribbons that could erase characters. Furthermore, thanks to the invention of the ball type (see Fig. 2), it became possible to easily change fonts on typewriters, thereby solving the problem of hammer jams on mechanical machines. This innovation not only improved efficiency but also allowed the use of additional balls with memorable characters, such as those used in mathematics, expanding the writing possibilities.



Fig. 2. IBM Correcting Selectric type ball (Photo: Ernesto Pelzin).

The first true word-processor automation was developed by IBM in 1964 with a new typewriter model that included magnetic storage for entered text, allowing users to retrieve, reuse, and correct previously written material [6].

With the invention of the floppy disk, which is now obsolete, the evolution of computer-based word processors accelerated, first in machines dedicated exclusively to that task and then in personal computers. In the 1970s, large organizations relied on mainframe and minicomputer systems, technologies not available to the public. With the emergence of microcomputers, these machines began to reach homes and small offices [8].

In 1978, WordStar was introduced, one of the first commercial word processors that was later surpassed in popularity by WordPerfect during the 1980s. The rise of Windows cemented Microsoft Word's position as the world's most widely used word processor.

Free, open-source word processors have had a significant impact on computing and technological development. These programs allow users to edit and create documents without the costs associated with proprietary software licenses. Furthermore, being open source, they allow developers to collaborate, improve, and adapt the software to the specific needs of their communities and projects.

One of the best-known word processors in this field is LibreOffice Writer, part of the LibreOffice suite, which emerged as a fork of the OpenOffice project. LibreOffice supports Microsoft Word document formats and is developed by a global open-source community. In education and other resource-constrained settings, free software can reduce licensing barriers and give users greater control over the technology they use [9].

Another notable example is AbiWord, a lightweight, cross-platform word processor that enables users with limited resources to access essential document-creation and editing features. While it is not as popular as LibreOffice, its efficiency in resource-constrained environments makes it a valuable option.

The impact of these open-source word processors is also reflected in government and educational institutions. Some organizations have adopted free software to reduce dependence on proprietary vendors and to foster technological autonomy. The collaborative, transparent open-source development model is discussed by Raymond [10].

Evelyn Berezin, an American engineer, is widely credited with developing an early computerized word-processing system. In 1969, she founded Redactron Corporation, where she designed the Data Secretary, a machine that automated text editing and processing in offices and helped bridge the transition from mechanical to digital word processing [11].

Berezin's contribution was not only technical but also social, because office automation changed repetitive document- management tasks and the work practices surrounding them. Her career also underscores the contribution of women to the development of computing technology [11].

Origins of Spell Checking

In 1947, the American mathematician and engineer Richard W. Hamming proposed a method for error correction in computer-to-computer transmissions and a way to measure the distance between equal-length strings [12]. Later, Vladimir Levenshtein generalized the idea to strings of different lengths through insertion, deletion, and substitution operations [13]. These methods became foundational for approximate string matching and spelling-correction software.

The first spell checkers appeared in the 1960s, during the mainframe era. These systems were primarily used in large organizations and universities that could afford powerful mainframe computers. Spell checkers in this environment were rudimentary programs that worked quite differently from the ones we use today.

A mainframe is a high-powered computer designed to process large volumes of data and run multiple applications simultaneously. Early spell-checking systems compared each typed word with entries in a stored dictionary. If a word did not match an entry, it could be flagged for review; early systems generally offered far less contextual assistance than modern tools [14].

The speed and efficiency of early spelling checkers were limited by the technology of the time. Storing large dictionaries in memory and processing long texts were significant challenges. Search and hashing techniques helped make dictionary lookup more efficient [15], but contextual ranking comparable to contemporary writing assistants was not yet feasible.

FUNDAMENTALS OF SPELLING CORRECTION AND TEXT PREDICTION

The metric or distance between words

Measuring and counting are activities that humans have performed since ancient times. Throughout different cultures, these tasks have been carried out in various ways to meet the needs of social life. For example, the Eskimos developed a number system based on the number 20, while the Vikings used a combined system based on 10 and 12. Similarly, measurement methods have been diverse, leading mathematicians to develop generalizations such as measure theory and the theory of metric spaces, and to formalize concepts such as distance.

This formalization of the concept of distance eventually led some mathematicians to wonder if it was possible to measure the distance between words in a language. This reasoning laid the groundwork for tools such as spell checkers and text predictors, which we use today on computers and mobile phones.

Throughout our lives, we learn to measure things like time, weight, volume, and, particularly, distance. Without realizing it, we employ various methods to measure distances. The most common method is the straight-line distance, measured in units such as meters, which is known as the Euclidean Metric or Distance. However, we also use another metric known as the 'Taxi Driver's Distance or Metric'. If a taxi can move along all streets in any direction, Fig. 3 illustrates how a taxi driver measures the distance between points A and B: it will always be four blocks, regardless of the chosen route, since the streets follow an orthogonal pattern.

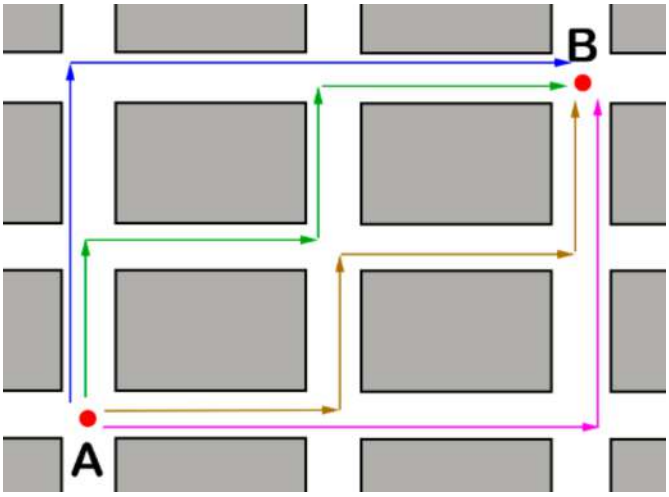


Fig. 3. Taxi Driver Metric or Distance. Source: The Authors.

The Russian mathematician Vladimir Levenshtein, mentioned above, developed a metric known as Levenshtein distance, which measures how many single-character edits separate two strings. A smaller distance indicates greater similarity in form. The calculation is based on three operations: insertion, deletion, or substitution of characters [16].

Next, we explain these operations using simple examples, without resorting to complex formulas or algorithms. Consider the Spanish words ‘espada’ and ‘espalda’. As Fig. 4 shows, converting ‘espada’ into ‘espalda’ requires inserting the letter ‘l’ between ‘a’ and ‘d’. Since only one operation is required, their Levenshtein distance is 1.

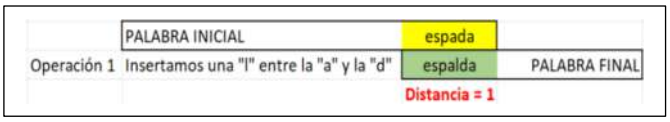


Fig. 4. Levenshtein distance between ‘espada’ and ‘espalda’. Source: The Authors.

This distance value of 1 indicates that the two words are very similar. Now, consider the case of the words ‘zapato’ and ‘sapote’. As shown in Fig. 5, two substitution operations are required to convert ‘zapato’ to ‘sapote’. This suggests that the words are less similar compared to the previous example, as they require more changes to transform one into the other.

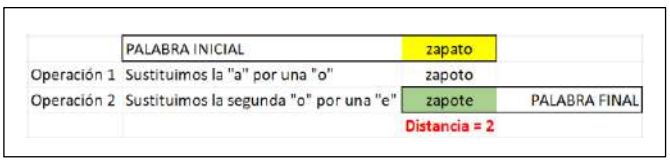


Fig. 5. Levenshtein distance between ‘zapato’ and ‘sapote’. Source: The Authors.

The following example measures the distance between the words ‘convicción’ and ‘contención’. At first glance, they might seem less similar than the previous word

pairs, as demonstrated in Fig. 6, where three simple operations are required to transform one into the other. This result reflects a lower similarity between the two words.

	PALABRA INICIAL	convicción	
Operación 1	Sustituimos la "v" por una "t"	conticción	
Operación 2	Sustituimos la primera "i" por una "e"	contección	
Operación 3	Sustituimos la primera "c" por una "n"	contención	PALABRA FINAL
		Distancia = 3	

Fig. 6. Levenshtein distance between 'convicción' and 'contención'. Source: The Authors.

The previous examples show word pairs in which the first word has the same number of characters or fewer than the second, and in which insertion and substitution operations are applied to transform the first into the second. Now consider the case of 'Bolivia' and 'boca', where the first word is longer than the second. As shown in Fig. 7, deletion operations are necessary. Fig. 8 shows that reversing the order of the two strings produces the same distance.

	PALABRA INICIAL	bolivia	
Operación 1	Sustituimos la "l" por una "c"	bocivia	
Operación 2	Eliminamos la primera "i"	bocvia	
Operación 3	Eliminamos la "v"	bocia	
Operación 4	Eliminamos la "i"	boca	PALABRA FINA
		Distancia = 4	

Fig. 7. Levenshtein distance between 'bolivia' and 'boca'. Source: The Authors.

	PALABRA INICIAL	boca	
Operación 1	Sustituimos la "c" por una "l"	bola	
Operación 2	Insertamos una "i" entre la "l" y la "a"	bolia	
Operación 3	Insertamos una "v" entre la "i" y la "a"	boliva	
Operación 4	Insertamos una "i" entre la "v" y la "a"	bolivia	PALABRA FINAL
		Distancia = 4	

Fig. 8. Levenshtein distance between 'boca' and 'bolivia'. Source: The Authors.

Now consider the words 'glorieta' and 'camisa', which at first glance appear very different in form and meaning, making it unlikely that a writer would confuse them when typing. Fig. 9 shows that the distance between them is 6 units, reflecting the multiple operations required to transform one word into the other.

	PALABRA INICIAL	glorieta	
Operación 1	Sustituimos la "g" por una "c"	clorieta	
Operación 2	Sustituimos la "l" por una "a"	caorieta	
Operación 3	Sustituimos la "o" por una "m"	camrieta	
Operación 4	Eliminamos la "r"	camieta	
Operación 5	Sustituimos la "e" por una "s"	camista	
Operación 6	Eliminamos la "t"	camisa	PALABRA FINAL
		Distancia = 6	

Fig. 9. Levenshtein distance between 'glorieta' and 'camisa'. Source: The Authors.

The process of measuring the distance between words is straightforward, as illustrated in the previous examples. However, demonstrating mathematically that it satisfies all the properties of a distance function was a significant achievement for Levenshtein. Furthermore, expressing this concept in a computational algorithm that a machine can execute adds considerable complexity, making it difficult for those unfamiliar with mathematics or computer science to understand. Over time, other methods emerged with the same purpose, each with its own characteristics and applications.

How Spell Check Works

Similarity functions can be character-based or token-based. For spelling correction, character-based edit distance is particularly useful because it directly models insertions, deletions, and substitutions [17]. The examples above illustrate that small edit distances often correspond to morphologically similar strings. However, a fixed threshold should be treated as a candidate-generation heuristic rather than a universal rule: the usefulness of a candidate also depends on word length, frequency, language, and context.

For the pedagogical prototype used in this article, candidates within one to three edits were inspected, with special attention to distances of one and two. The prototype uses a Spanish dictionary of 646,612 entries [18]. Microsoft Word is used only as a descriptive comparator because its internal dictionary and ranking algorithm are proprietary and may vary by version.

When the user types a word, the spell checker first determines whether the token appears in its dictionary. If it is not found, Microsoft Word typically marks it with a red underline, as illustrated by 'tuho' in Fig. 10. By contrast, valid dictionary entries such as 'tubo', 'tuvo', or 'tuco' are not flagged even when the sentence itself may be semantically odd.

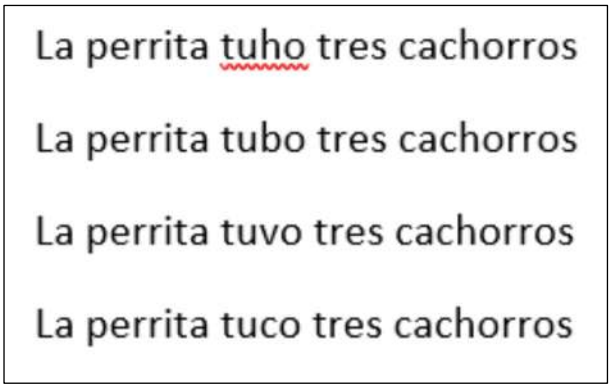


Fig. 10. Microsoft Word Spell Checker. Source: The Authors.

As you can see, a human spell checker would flag errors in three of the four sentences, except for the one that says "La perrita tuvo tres cachorros." This demonstrates that the computer spell checker is not perfect, but it is a good aid in the task of constructing documents written on a computer.

In a simple edit-distance prototype, an unknown word such as 'tuho' can be compared with dictionary entries and the nearest strings can be returned as candidate replacements. Production spell checkers generally perform additional ranking rather than displaying every near neighbor; contextual language models can further improve which suggestion appears first [3].

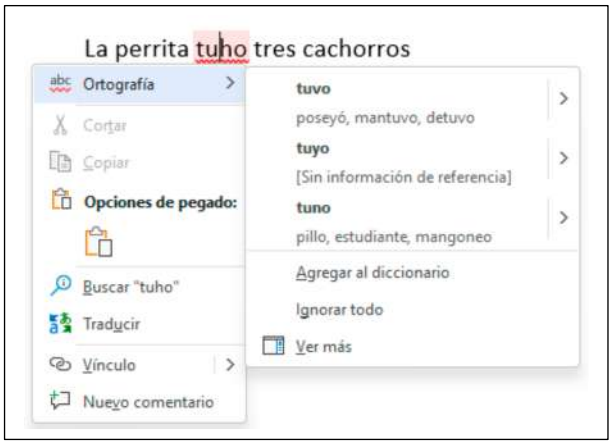


Fig. 11. List of replacement words for 'tuho'. Source: The Authors.

EXPERIMENTATION WITH A SPELL CHECKER PROTOTYPE

When the deliberately misspelled Spanish token 'conticción' is typed in Microsoft Word, it is underlined in red because it is not recognized as a dictionary entry. Fig. 12 shows the context menu displayed for this token and the correction suggestions offered by the word processor.

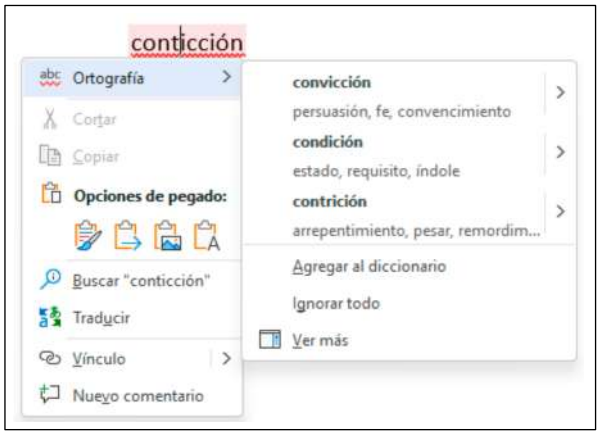


Fig. 12. List of words suggested by Microsoft Word as a 'conticción' replacement. Source: The Authors.

METHODOLOGICAL DESIGN AND REPLICATION PROTOCOL

The experimental instrument is a pedagogical spell-checker prototype developed for the Computer Science program at the University of Nariño to demonstrate similarity algorithms. The prototype accepts a single input token, compares it against a Spanish dictionary containing 646,612 entries [18], computes Levenshtein distance for each dictionary entry, and groups the resulting candidates by exact distance. In the experiment reported here, only candidates at distances 1, 2, and 3 were retained for inspection. This is an illustrative case study rather than a performance benchmark: the reported quantitative sample is one deliberately misspelled token (n = 1), 'conticción'.

Replication procedure: (1) load the same 646,612-entry Spanish dictionary [18]; (2) enter the token 'conticción'; (3) calculate the Levenshtein distance between the input and every dictionary entry; (4) count candidates whose distance is exactly 1, 2, or 3; and (5) compare the resulting candidates descriptively with the suggestions displayed by Microsoft Word for the same input. The Microsoft Word version/build is not recorded in the original experiment, so the commercial-tool side of the comparison cannot be reproduced exactly and is therefore not used to support claims of superior performance.

Using the procedure above, the prototype returned 1 candidate at distance 1, 12 candidates at distance 2, and 56 candidates at distance 3 for 'conticción', as shown in Fig. 13. These counts are reported systematically below and can be reproduced with the same dictionary and distance procedure.

TABLE I
REPRODUCIBLE PARAMETERS AND OBSERVED RESULTS FOR THE DOCUMENTED TEST CASE.

Replication parameter	Reported value
Input token	conticción (deliberately misspelled)
Spanish dictionary	646,612 entries [18]
Distance thresholds	Exact Levenshtein distances 1, 2, and 3
Distance 1 candidates	1
Distance 2 candidates	12
Distance 3 candidates	56
Comparator	Microsoft Word suggestions for the same token (descriptive comparison)

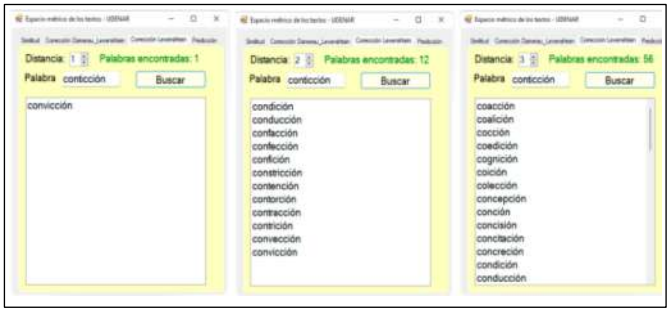


Fig. 13. Spell checker prototype using the word 'conticción'. Source: The Authors.

The original classroom trials included additional words, but they were not collected under a predefined sampling plan and are therefore not treated as a formal dataset in this article. Consequently, no general claim is made that the prototype is more efficient or more accurate than Microsoft Word. A future comparative evaluation should predefine a representative set of misspellings with gold-standard corrections and report measures such as top-1/top-k correction accuracy, candidate recall, and processing time.

FINDINGS

Spell checkers have transformed everyday writing by identifying typographical and orthographic problems and by suggesting possible corrections. The technology has progressed from dictionary lookup and character-level similarity to systems that incorporate statistical and neural context [3].

One of the most important considerations about spell checkers is their role in democratizing writing. Before these tools, spelling errors were more common, especially in formal texts. Today, anyone, regardless of their educational level, can access spell checkers that improve the quality of their writing, making written communication more accurate and accessible.

However, despite their efficiency, spell checkers have their limitations. They are still imperfect at handling context and the multiple meanings of words. For example, they may fail to identify incorrect homophone use or misinterpret expressions that are correct in a particular context. This highlights the need for users not to rely exclusively on these tools, but to use them as support.

AI-based writing assistants increasingly analyze complete context rather than isolated words. Google has described on-device neural grammar correction in Gboard, designed under memory and latency constraints [19], while recent text- editing research uses instruction-tuned language models for grammatical error correction, simplification, and paraphrasing [4], [20]. These approaches illustrate why commercial systems can rank a short list of contextually useful corrections even when a simple edit-distance prototype generates many formsimilar candidates.

The prototype spell checker performs its intended educational function by exposing candidate generation based on edit distance. For 'conticción', Microsoft Word displays a short list of suggestions, while the prototype returns a larger candidate set at distance 2. Because all suggestions displayed by Microsoft Word in the screenshot are also contained in the prototype's broader set, the comparison illustrates coverage rather than superiority: more candidates do not necessarily mean better correction.

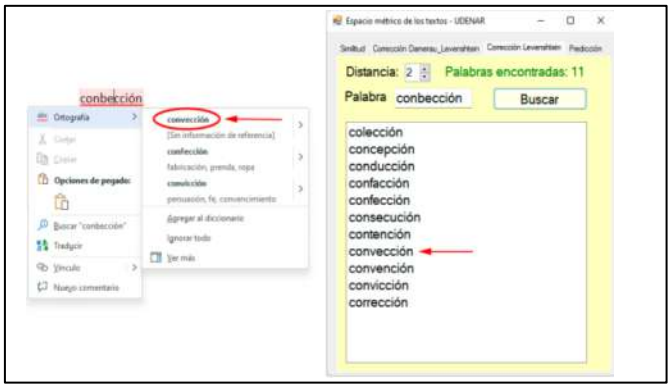


Fig. 14. Comparison between the word spell checker and the prototype. Source: The Authors.

The smaller number of suggestions displayed by Microsoft Word should not be interpreted as evidence that its dictionary contains fewer entries. Fig. 15 shows that a word such as 'convección' can exist in Microsoft Word's dictionary even when it is not displayed among the first suggestions for 'conticción'. This indicates that the commercial tool applies selection or ranking criteria beyond simple dictionary membership.

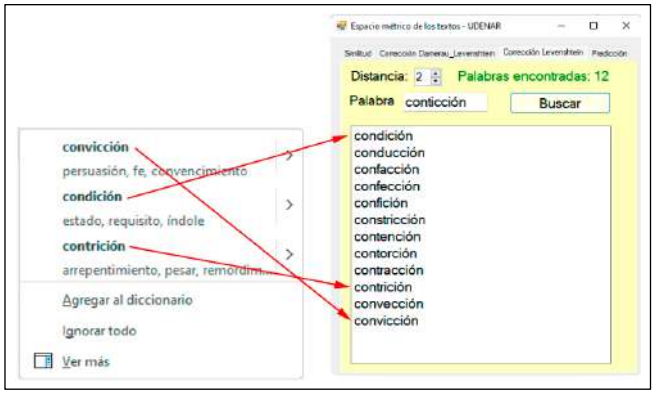


Fig. 15. Comparison between the word spell checker and the prototype. Source: The Authors.

Microsoft Word displays only a few suggestions even when additional dictionary entries are available. This behavior is consistent with the idea that mature writing tools rank candidate corrections using signals beyond raw edit distance, potentially including frequency and contextual information. The present experiment does not reveal Microsoft's proprietary algorithm; it only

demonstrates that exhaustive candidate generation and user-facing ranking are different stages.

Spell checkers in free, open-source word processors have improved significantly in recent years. However, in terms of efficiency and accuracy, they may differ from proprietary software checkers such as Microsoft Word.

LibreOffice Writer offers a spell checker based on the Hunspell engine, an open-source spell checker also used in other applications. Hunspell is efficient for morphologically rich languages and supports large, customizable dictionaries. Its performance depends strongly on the installed linguistic resources and configuration [21]. Compared with modern context-aware writing assistants, a dictionary-centered checker may provide less support for subtle grammatical or semantic errors.

Hunspell-based checking remains valuable because it is transparent, lightweight, and adaptable. However, dictionary and morphology rules alone do not provide the same sentence-level contextual modeling as neural grammatical-error- correction systems [3].

Spell checkers in free word processors, such as LibreOffice and AbiWord, are efficient and functional for most users who need a basic tool. However, for those who require more in-depth grammar analysis or more advanced suggestions, these checkers may fall short compared to commercial solutions.

CONCLUSION

Spell checking has evolved from simple dictionary lookups to context-aware writing assistants, and the article's prototype sits in that arc as a transparent didactic system that makes one foundational mechanism visible. By anchoring candidate generation in Levenshtein distance over a large Spanish lexicon of 646,612 entries [18], the prototype shows how near-neighbor candidates can be generated within predefined edit radii.

The revised experiment should be interpreted narrowly. For the single reported test token 'conticción', the prototype produced 1 candidate at distance 1, 12 at distance 2, and 56 at distance 3. The comparison with Microsoft Word shows that a commercial checker may present a much shorter list even when more dictionary neighbors exist. The exercise therefore demonstrates the distinction between candidate coverage and suggestion ranking, not comparative efficiency or accuracy.

A broader contribution of the work is its educational value. The prototype was conceived as part of a learning object to teach similarity algorithms in an accessible way, bridging formal definitions and hands-on experimentation without overwhelming readers with mathematical detail. Students and practitioners can trace how a misspelled input triggers a dictionary lookup, how distances are computed, and how candidate sets emerge, thereby demystifying a capability that most writers take for granted inside modern word processors. This transparency advances the pedagogical goal stated in the manuscript: to explain how machines perform spelling correction in terms that a general audience can follow, while still grounding the explanation in core computational ideas.

At the same time, the findings reaffirm well-known limits of purely character-based approaches. Edit distance can surface strings that are similar in form, but it does not by itself resolve context, semantics, or grammatical agreement. Contemporary grammatical-error-correction systems increasingly use neural and pretrained language models to address these limitations [3], [4]. A natural next step for the prototype is to preserve transparent distance-based candidate generation while adding frequency- or context-aware re-ranking.

In summary, the prototype serves as a working educational demonstration of edit-distance-based spelling correction. Its value lies in making the candidate-generation process visible and reproducible for a simple documented case. Future work should evaluate a predefined sample of Spanish misspellings with gold-standard corrections, record software versions and execution conditions, and compare systems using top-k accuracy, candidate recall, and processing time. Such an evaluation would retain the prototype's instructional clarity while supporting stronger empirical claims.

REFERENCES

[1].R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," Journal of the ACM, vol. 21, no. 1, pp. 168–173, 1974, doi: 10.1145/321796.321811.

[2].D. Jurafsky and J. H. Martin, Speech and Language Processing, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2009.

[3].C. Bryant, Z. Yuan, M. R. Qorib, H. Cao, H. T. Ng, and T. Briscoe, "Grammatical error correction: A survey of the state of the art," Computational Linguistics, vol. 49, no. 3, pp. 643–701, 2023, doi: 10.1162/coli_a_00478.

[4].V. Raheja, D. Alikaniotis, V. Kulkarni, B. Alhafni, and D. Kumar, "mEdIT: Multilingual text editing via instruction tuning," in Proc. 2024 Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Mexico City, Mexico, 2024, pp. 979–1001, doi: 10.18653/v1/2024.naacl-long.56.

[5].J. Price and L. Urban, The Definitive Word-Processing Book. New York, NY, USA: Viking Penguin, 1984.

[6].A. Barredo, "Historia de la tecnología: el primer procesador de texto," Hipertextual, Jan. 2016. [Online]. Available: https://hipertextual.com/2016/01/historia-la-tecnologia-primer-procesador-texto. [Accessed: Feb. 12, 2022].

[7].Wikipedia contributors, "Christopher Sholes," Wikipedia,

2022. [Online]. Available: https://es.wikipedia.org/wiki/Christopher_Sholes. [Accessed: Feb. 12, 2022].

[8].T. Bergin, "The origins of word processing software for personal computers: 1976–1985," IEEE Annals of the History of Computing, vol. 28, no. 4, pp. 32–47, 2006, doi: 10.1109/MAHC.2006.76.

[9].R. Stallman, Free Software, Free Society: Selected Essays of Richard M. Stallman. Boston, MA, USA: GNU Press, 2002.

[10].E. S. Raymond, The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. Sebastopol, CA, USA: O'Reilly Media, 1999.

[11].BBC News Mundo, "Evelyn Berezin, la pionera de la computación que creó el primer procesador de texto," Dec. 2018. [Online]. Available: https://www.bbc.com/mundo/noticias-46567517. [Accessed: Feb. 13, 2022].

[12].C. Polcino, "Group algebras and coding theory: A short survey," Revista Integración, Temas de Matemáticas, vol. 37, no. 1, pp. 153–166, 2019.

[13].V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," Soviet Physics Doklady, vol. 10, no. 8, pp. 707–710, 1966.

[14].M. Y. Vardi, Early Computer Systems and Text Processing. New York, NY, USA: Computing Press, 1975.

[15].D. E. Knuth, The Art of Computer Programming, Vol. 3: Sorting and Searching. Reading, MA, USA: Addison-Wesley, 1981.

[16].Wikipedia contributors, "Distancia de Levenshtein," Wikipedia, 2021. [Online]. Available: https://es.wikipedia.org/wiki/Distancia_de_Levenshtein. [Accessed: Feb. 13, 2022].

[17].I. Amón, F. Moreno, and J. Echeverri, "Algoritmo fonético para detección de cadenas de texto duplicadas en el idioma español," Revista Ingenierías Universidad de Medellín, vol. 11, no. 20, pp. 127–138, 2012.

[18].JorgeDuenasLerin, "Diccionario español TXT," GitHub repository, 2022. [Online]. Available: https://github.com/JorgeDuenasLerin/diccionario- espanol-txt. [Accessed: Feb. 14, 2022].

[19].T. Mak and S. Tong, "Grammar correction as you type, on Pixel 6," Google Research, Oct. 27, 2021. [Online]. Available: https://research.google/blog/grammar-correction-as-you-type-on-pixel-6/.

[20].V. Raheja, D. Kumar, R. Koo, and D. Kang, "CoEdIT: Text editing by task-specific instruction tuning," in Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 2023, pp. 5274–5291, doi: 10.18653/v1/2023.findings-emnlp.350.

[21].O. Tymkiv, LibreOffice Development and Enhancement. Warsaw, Poland: FLOSS Publications, 2019.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

Comparative Analysis of Glucose Control Strategies under Stress Scenarios

Análisis Comparativo de Estrategias de Control de la Glucosa bajo Escenarios de Estrés

David Andrés Ojeda Gúzman[1*], Manuela María Muriel Tobar [2], Germán Darío Obando Bravo[3], Andrés Fernando Arciniegas Mejía[4]

[123]Department of Electronic Engineering, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia. daojedag@udenar.edu.co, murieltmanu@udenar.edu.co, gdobando@udenar.edu.co

[4]SEG - Strategic Engineering Group, ADASI, Edge Group, Abu Dhabi, United Arab Emirates. arciandres@gmail.com

*Autor de correspondencia: daojedag@udenar.edu.co

Citación: D. A. Ojeda Guzmán, M. M. Muriel Tobar, G. D. Obando Bravo y A. F. Arciniegas Mejía, "Comparative Analysis of Glucose Control Strategies under Stress Scenarios," Revista de Ingeniería, vol. 4, no. 1, pp. 32–44, Nov. 2026.

***Abstract*—** Type 1 Diabetes (T1D) requires lifelong exogenous insulin therapy to maintain glycemic control. This study compares four control strategies: ON/OFF, proportional–integral–derivative (PID) with feedforward, Gain Scheduling (GS), and Model Predictive Control (MPC), using the modified Bergman and Hovorka physiological models under six metabolic stress scenarios, including high carbohydrate intake and severe glycemic excursions. Results show that MPC achieves the best overall clinical trade-off and lowest insulin consumption with the Bergman model. However, its robustness decreases under initial hyperglycemia in the Hovorka model, where aggressive corrections induce severe hypoglycemia, while GS exhibits performance degradation due to nonlinear dynamics and physiological delays. These findings highlight the potential of predictive control for artificial pancreas systems and the importance of physiologically detailed models, while evaluating computational cost and insulin requirements for embedded implementation.

***Keywords*—**closed-loop control, glucose-insulin system, glycemic control, model predictive control, physiological modeling, type 1 diabetes.

***Resumen*—** La diabetes mellitus tipo 1 (DM1) requiere terapia de insulina exógena para mantener el control glucémico. Este estudio compara cuatro estrategias de control: ON/OFF, PID con acción anticipativa *(feedforward*), programación de ganancias (GS) y control predictivo basado en modelos (MPC), utilizando los modelos fisiológicos de Bergman y Hovorka bajo seis escenarios de estrés metabólico. Los resultados muestran que MPC logra el mejor compromiso clínico y el menor consumo de insulina con el modelo de Bergman. Sin embargo, en el modelo de Hovorka, su desempeño se ve afectado por la hiperglucemia inicial, que puede provocar hipoglucemia severa, mientras que GS presenta limitaciones ante dinámicas no lineales y retardos fisiológicos. Estos hallazgos destacan el potencial del control predictivo en sistemas de páncreas artificial y la importancia de evaluar el desempeño clínico, el consumo de insulina y el costo computacional para su implementación en sistemas embebidos.

***Palabras clave*—** control en lazo cerrado, sistema glucosa-insulina, control glucémico, control predictivo basado en modelos, modelado fisiológico, diabetes mellitus tipo 1.

I. INTRODUCTION

In 2024, it is estimated that approximately 14% of the global population suffers from some form of diabetes, representing a sustained increase compared to the 7% reported in 1990 [1]. While this prevalence encompasses various forms of the disease, the clinical management of type 1 diabetes (T1D) remains challenging due to the persistent risk of acute and chronic complications associated with inadequate glycemic control. Despite advances in multiple daily injections and insulin pumps, achieving stable metabolic control remains difficult, underscoring the need for artificial pancreas (AP) technologies [2], [3].

From a control engineering perspective, the glucose-insulin system is characterized by high nonlinearity, significant phys-iological transport delays, and substantial interpatient variabil-ity [4]. Furthermore, the system is subject to external distur-bances from meal intake, inherent sensor noise in continuous glucose monitors (CGM), and parametric uncertainty due to changing metabolic conditions [5], [6].

The state of the art in glycemic regulation includes sev-eral strategies with varying levels of complexity. Rule-based ON/OFF control is one of the simplest approaches, in which insulin delivery is determined according to predefined glucose ranges [7]. PID controllers are valued for their simplicity; however, their reactive nature may limit the management of postprandial disturbances and physiological delays [7], [8]. Gain Scheduling (GS) controllers aim to address nonlinearity by interpolating between local linear controllers, yet their performance may degrade when operating conditions deviate substantially from those represented by the local models [4]. Conversely, MPC has been widely investigated as an advanced strategy due to its ability to anticipate future glucose trends and explicitly incorporate physiological constraints [9]. Nevertheless, MPC performance depends on model accuracy and typically entails a higher computational burden [3].

A critical factor in evaluating these strategies is the choice of mathematical representation. The modified Bergman model is appropriate for preliminary designs due to its relatively low complexity and 5-state structure [10], [11]. In contrast, the Hovorka model provides a detailed 8-state multicompartmen-tal representation, offering higher physiological fidelity [9]. This disparity raises a fundamental question: do controllers that perform adequately in simplified models maintain their effectiveness when evaluated in more detailed physiological representations?

Current research often focuses on individual strategies or limited scenarios, and comparative evaluations integrating these factors remain limited. To bridge this gap, this study presents a systematic comparative analysis of four control strategies (ON/OFF, PID, GS, and MPC) across both Bergman and Hovorka models. The main scientific contributions in-clude: a) the comparison of four control strategies under six metabolic stress scenarios; b) the evaluation of the impact of physiological complexity; and c) a joint assessment of glycemic performance, computational requirements, and in-sulin consumption [6].

II. METHODOLOGY

This section presents the methodological framework used to evaluate and compare the performance of automated in-sulin delivery strategies. The methodology is organized into three sequential phases. First, mathematical representations of the glucose-insulin system are established by considering the reduced-order modified Bergman model and the detailed multicompartmental Hovorka model. Second, the design, tun-ing, and physical-constraints addressing of the four evaluated control strategies ON/OFF, PID, Gain Scheduling, and MPC are defined under comparable operating conditions. Finally, a simulation and evaluation framework is established, in-corporating six metabolic stress scenarios and clinical and control performance metrics to assess glycemic regulation, computational requirements, and insulin consumption.

A. Physiological Modeling

To design and evaluate closed-loop insulin delivery strate-gies, mathematical models that describe the dynamic behavior of the glucose-insulin system are essential [11]. These models must capture the patient's physiological responses to external disturbances, such as carbohydrate intake, and the absorption kinetics of exogenously administered insulin [11]. In this study, two models of different physiological complexity—the modified Bergman model and the Hovorka model—are employed to establish a systematic and rigorous evaluation platform for the control strategies [9], [10].

1) Modified 5-State Bergman Model: The modified Bergman minimal model is a commonly used mathematical framework for representing the fundamental glucose-insulin dynamics in T1D [10], [11]. In this study, an extended 5-state formulation is adopted to incorporate carbohydrate absorption and the delay associated with subcutaneous glucose mea-surements, providing a suitable representation for closed-loop artificial pancreas simulations [12]. The dynamic behavior of the system is described by the following ordinary differential equations (ODEs):

$$\dot{G}(t) = -\left(p_1 + X(t)\right) G(t) + p_1 G_b + D(t), \quad (1)$$

$$\dot{X}(t) = -p_2 X(t) + p_3 \left(I(t) - I_b\right), \quad (2)$$

$$\dot{I}(t) = -p_4 I(t) + \frac{1}{V_I} U(t), \quad (3)$$

$$\dot{D}(t) = -d_{\text{rate}} D(t), \quad (4)$$

$$\dot{G}_{sc}(t) = \frac{1}{5} r_u \left(G(t) - G_{sc}(t)\right). \quad (5)$$

where G(t) represents the plasma glucose concentration (mg/dL), X(t) represents the remote insulin action (min−1), I(t) denotes the plasma insulin concentration (mU/L), D(t) represents the rate of glucose appearance resulting from car-bohydrate absorption (mg/dL/min), and Gsc(t) denotes the subcutaneous glucose concentration (mg/dL). The control in-put U (t) corresponds to the exogenous insulin infusion rate (U/min), while Gb and Ib represent the basal glucose and insulin concentrations, respectively.

Each state represents a distinct component of the glucose-insulin dynamics:

- **Plasma glucose dynamics (1):** The glucose concen-tration is affected by insulin-independent and insulin-dependent glucose utilization, represented by −p1G(t) and −X(t)G(t), respectively. The term p1Gb represents the basal glucose production required to maintain the nominal equilibrium, while D(t) accounts for glucose appearance following carbohydrate absorption.
- **Remote insulin action (2):** The state X(t) represents the delayed effect of plasma insulin on glucose utilization. Its dynamics depend on the decay parameter p2 and the deviation of plasma insulin from its basal value Ib, scaled by the insulin sensitivity parameter p3.
- **Plasma insulin kinetics (3):** Plasma insulin follows first-order clearance dynamics represented by p4I(t) and is increased by the exogenous insulin infusion U (t) scaled by the insulin distribution volume VI .
- **Carbohydrate absorption (4):** The state D(t) repre-sents the rate of glucose appearance associated with gastrointestinal carbohydrate absorption. Its dynamics are represented by a first-order decay governed by drate.
- **Subcutaneous glucose dynamics (5):** The state Gsc(t) represents the delayed subcutaneous glucose concentra-tion used to characterize the measurement dynamics of the CGM. Its response to plasma glucose G(t) is represented by a first-order lag governed by ru.

The physiological parameters, their definitions, nominal values, and units are summarized in Table I.

TABLE I

PHYSIOLOGICAL PARAMETERS OF THE MODIFIED BERGMAN MODEL

Parameter	Nominal Value	Unit	Definition
p_1	0.03	min^{-1}	Insulin-independent glucose removal
p_2	0.02	min^{-1}	Dissipation rate of the insulin action
p_3	4.5×10^{-5}	$\text{L}/(\text{mU} \cdot \text{min}^2)$	Peripheral insulin sensitivity
p_4	0.08	min^{-1}	Plasma insulin removal (clearance)
V_I	12	L	Insulin distribution volume (plasma)
d_{rate}	5	min^{-1}	Intestinal glucose absorption
I_b	10	mU/L	Basal insulin during fasting
r_u	1	min^{-1}	Subcutaneous glucose lag constant

The reduced-order structure of this 5-state formulation pro-vides a computationally tractable representation for controller design and preliminary evaluation, while the Hovorka model is subsequently used to assess controller performance under a more physiologically detailed representation.

2) 8-State Hovorka Model: To evaluate the glycemic con-trol strategies using a physiologically detailed representation of T1D, the 8-state Hovorka multicompartmental model is imple-mented. Compared with reduced-order physiological models, this formulation provides a higher level of physiological detail by explicitly representing three interconnected subsystems: subcutaneous insulin absorption, intestinal glucose absorption, and the delayed metabolic actions of insulin on glucose distribution, utilization, and endogenous glucose [9].

The physiological behavior of the system is governed by the following set of ODEs:

$$\dot{Q}_1(t) = -\frac{F_{01}^C}{V_C} - x_1(t)Q_1(t) + k_{12}Q_2(t) - F_R + E_C(1 - x_3(t)) + U_G(t), \quad (6)$$

$$\dot{Q}_2(t) = x_1(t)Q_1(t) - (k_{12} + x_2(t))Q_2(t), \quad (7)$$

$$\dot{x}_1(t) = -k_{a1}x_1(t) + k_{b1}I(t), \quad (8)$$

$$\dot{x}_2(t) = -k_{a2}x_2(t) + k_{b2}I(t), \quad (9)$$

$$\dot{x}_3(t) = -k_{a3}x_3(t) + k_{b3}I(t), \quad (10)$$

$$\dot{S}_1(t) = U(t) - \frac{S_1(t)}{t_{\max}}, \quad (11)$$

$$\dot{S}_2(t) = \frac{S_1(t)}{t_{\max}} - U_I(t), \quad (12)$$

$$\dot{I}(t) = \frac{U_I(t)}{V_I} - k_e I(t). \quad (13)$$

where the state vector is defined as

$$x(t) = [Q_1(t), Q_2(t), x_1(t), x_2(t), x_3(t), S_1(t), S_2(t), I(t)]^T, \quad (14)$$

In this multicompartmental representation, Q1(t) (mmol/kg) represents the glucose mass in the accessible compartment, while Q(t) (mmol/kg) represents the glucose mass in the non-accessible compartment. The measurable plasma glucose concentration is obtained as G(t) = Q1(t)/VG. The states x1(t), x2(t), and x3(t) (min−1) represent delayed insulin actions associated with glucose transport and utilization and the suppression of endogenous glucose production, respectively. The states S1(t) and S2(t) (mU/kg) represent the subcutaneous insulin masses in the two absorption compartments, while I(t) (mU/L) denotes the plasma insulin concentration. The input U (t) (mU/kg/min) represents the exogenous insulin infusion rate delivered by the pump, normalized by patient body weight.

Physiologically, the state equations represent the following metabolic processes:

- **Glucose compartmental kinetics (6)–(7):** Glucose is distributed between the accessible and non-accessible compartments through the exchange rate k12. In the accessible compartment, glucose is affected by insulin-independent utilization, represented by FC /VG, insulin-dependent utilization represented by the interaction x1(t)Q1(t), renal clearance FR, endogenous glucose production EG(1 – x3(t)), and the glucose appearance rate UG(t) resulting from carbohydrate absorption. The non-accessible compartment is coupled to the accessible compartment through the corresponding exchange and insulin-dependent utilization terms.
- **Delayed insulin action (8)–(10):** The effects of plasma insulin on glucose metabolism are represented by three delayed action states. These states follow first-order acti-vation dynamics, in which plasma insulin I(t) stimulates the corresponding insulin actions through the parameters kb1, kb2, and kb3, while the active effects decay according to ka1, ka2, and ka3.
- **Subcutaneous insulin absorption and plasma kinetics (11)–(13):** Exogenous insulin infusion U (t) enters the first subcutaneous compartment S1(t) and subsequently transits to S (t). The insulin absorption rate into the systemic circulation is given by UI (t) = S2(t)/tmax. Plasma insulin I(t) is then increased by the absorbed insulin and eliminated according to the first-order clearance term keI(t).

The nominal values, physical definitions, and units of the Hovorka model parameters for the simulated patient are sumarized in Table II. The multi-state coupling and explicit representation of physiological delays make this model suit-able for evaluating controller performance under increased physiological complexity.

3) Meal Disturbance Modeling: Carbohydrate intake con-stitutes the primary meal disturbance affecting the glucose-insulin dynamics. Following the extended formulation [11], the rate of glucose appearance resulting from a meal is represented by a gamma-shaped function, which provides a physiologically plausible description of the progressive rise, postprandial peak and subsequent decay of plasma glucose after food intake [12]. The total glucose appearance rate is expressed as the sum of the individual meal contributions:

$$D(t) = \sum_i D_i(t), \quad D_i(t) = \frac{C_i}{T_i}\Gamma\left(\frac{t - t_i}{T_i}\right) \quad (15)$$

in (15), Ci is the amount of carbohydrates ingested in meal i, ti denotes the time instant of the intake, and Ti is the duration of the absorption process. In this study, three meal durations are considered: Ti = 90 min for breakfast, Ti = 120 min for lunch, and Ti = 90 min for dinner. This gamma-shaped formulation avoids non-physiological represen-tations such as ideal steps or impulses, capturing instead the temporal distribution of glucose absorption [12].

In the modified Bergman model, the meal input D(t) enters to the glucose state dynamics as an exogenous disturbance (1), while the corresponding compartment state follows a first-order decay governed by the absorption rate constant drate, as expressed in (4). The same carbohydrate intake profile is

TABLE II

PHYSIOLOGICAL PARAMETERS OF THE HOVORKA MODEL

Parameter	Nominal Value	Unit	Definition
V_I	0.12	L/kg	Insulin distribution volume (plasma)
V_G	0.16	L/kg	Glucose distribution volume (accessible)
E_G	0.0161	mmol/kg/min	Basal endogenous hepatic glucose production
F_{01}^C	0.0097	mmol/kg/min	Insulin-independent glucose utilization
F_R	0.003	mmol/kg/min	Renal glucose excretion rate
k_{12}	0.066	min^{-1}	Inter-compartmental glucose exchange
t_{max}	55	min	Mean subcutaneous insulin absorption time
k_e	0.138	min^{-1}	Plasma insulin elimination
k_{a1}	0.006	min^{-1}	Insulin action decay, glucose uptake
k_{a2}	0.06	min^{-1}	Insulin action decay, hepatic production
k_{a3}	0.03	min^{-1}	Insulin action decay, hepatic utilization
k_{b1}	0.0005	min^{-2}	Insulin action activation, glucose uptake
k_{b2}	0.0006	min^{-2}	Insulin action activation, hepatic production
k_{b3}	0.0003	min^{-2}	Insulin action activation, hepatic utilization

applied as the external input UG(t) in the Hovorka model, where it represents the rate of glucose appearance into the accessible compartment (6).

4) Open-Loop Baseline and Insulin Bolus Modeling: Be-fore the closed-loop strategies were evaluated, an *open-loop* (manual-therapy) simulation was carried out to establish a baseline of the uncontrolled behavior of both models. For the Bergman model, the exogenous insulin delivery is rep-resented by the effective infusion rate $U(t)$, which models the subcutaneous absorption of rapid-acting insulin boluses and their subsequent appearance in the bloodstream. Rather than being idealized impulses, the boluses reproduce the delay and transient nature of the subcutaneous action. Each bolus is described by an exponential-shaped function,

$$U_i(t) = \begin{cases} B_i \frac{\tau^{\alpha} e^{-\beta\tau}}{k}, & t \geq t_i, \\ 0, & t < t_i, \end{cases} \quad \tau = \frac{t - t_i}{T_i}, \quad (16)$$

where Bi is the bolus dose, ti its application time, Ti the characteristic absorption duration, α and β shape parameters, and k a normalization constant ensuring that the area under the curve is proportional to the administered dose. For all boluses, Ti = 180 min, α = 1.5, and β = 2, with a constant basal infusion Ub = 0.0001 U/min. The same exogenous insulin infusion UI (t) is applied in the Hovorka model.

Both models were simulated under an identical open-loop scenario comprising three carbohydrate meals of 40, 60, and 50 g at 7:00 AM, 1:00 PM, and 8:00 PM, respectively, together with three corresponding manual insulin boluses of 4, 4, and 5 U at the same times [12]. As shown in Fig. 1 and Fig. 2, the uncontrolled responses exhibit clearly deficient glycemic regu-lation: sustained hyperglycemic episodes throughout the simu-lation and high glycemic variability, reflecting the progressive depletion of the insulin action and the inability of the open-loop system to compensate for meal-induced disturbances. In the Hovorka model, marked hyperglycemia is accompanied by pronounced glucose drops after each insulin dose, evidencing the absence of feedback regulation. These baseline responses motivate the design of the closed-loop strategies described in the next section.

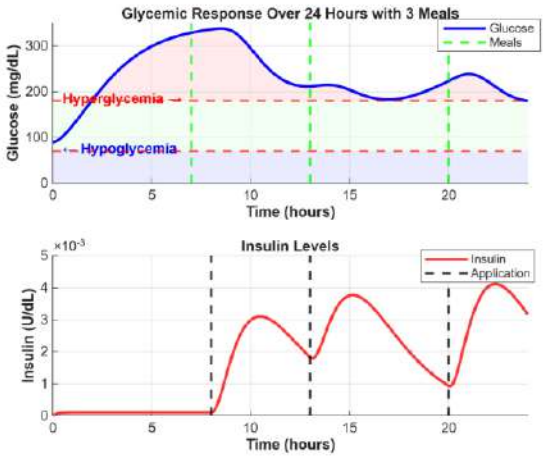


Fig. 1. Open-loop response of the modified Bergman model under three meals (40, 60, 50 g) and three manual insulin boluses (4, 4, 5 U). The uncontrolled system exhibits sustained hyperglycemia and high variability.

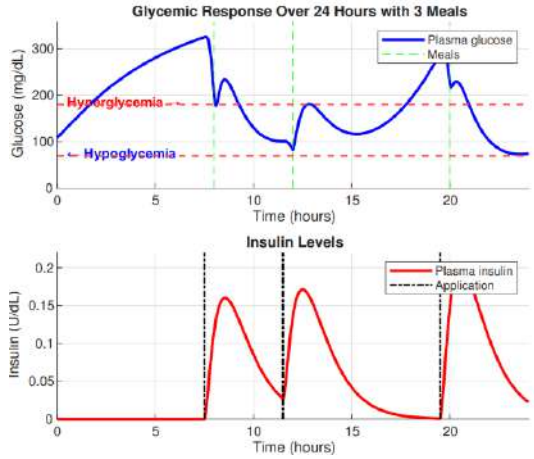


Fig. 2. Open-loop response of the Hovorka model under the same meal and bolus schedule. Hyperglycemic episodes and pronounced post-dose glucose drops confirm the absence of feedback regulation.

B. Glycemic Control Strategies

Four classical strategies of increasing complexity were implemented over the two physiological models, and their parameters were tuned empirically to ensure adequate closed-loop performance. Each controller uses the exogenous insulin infusion as the manipulated variable, whereas meal intake acts as an unmanipulable disturbance.

1) ON/OFF Control With Hysteresis: The simplest strategy investigated is a binary controller that toggles insulin delivery according to the blood glucose level, following the rule

$$u(t) = \begin{cases} u_{\text{on}}, & G(t) > 185 \text{ mg/dL}, \\ 0, & G(t) < 175 \text{ mg/dL}, \end{cases} \qquad \tau = \frac{t - t_i}{T_i}, \tag{16}$$

where uon = 0.019 U/min is a fixed infusion rate selected within a physiologically safe range, above the typical basal delivery and sufficiently high to elicit a corrective action without approaching the maximum output of commercial pumps [13]. The ±5 mg/dL hysteresis band prevents rapid actuator switching around the activation threshold, and a hard safety limit of 70 mg/dL forces the pump to stop whenever glucose falls below it. While computationally inexpensive and straightforward to implement, its binary nature precludes fine dose adjustment, leading to wide oscillations and slow responses to postprandial disturbances.

2) PID with Feedforward and Anti-Windup: A fixed-gain PID controller was designed by linearizing each model around an equilibrium point corresponding to 100 mg/dL using MATLAB's pidtune tool. A dead zone around the reference sup-presses unnecessary actuation for small glycemic deviations, and an anti-windup mechanism enforces the physical actuator limit of 0.032 U/min. The derivative term is implicitly filtered through the model dynamics, reducing sensitivity to measure-ment noise. A feedforward action delivers anticipatory insulin boluses approximately 30 min before each meal, computed from a predefined insulin-to-carbohydrate ratio, to compensate intestinal glucose absorption and improve the postprandial response. Table III summarizes the resulting gains; note that the Hovorka gains are negative because, with the error defined as measured glucose minus reference, the static input–output gain of this model is negative and negative gains therefore yield effective negative feedback.

TABLE III

PID PARAMETERS FOR THE BERGMAN AND HOVORKA MODELS

Parameter	Bergman	Hovorka
Proportional gain K_p	0.08	−0.0002
Integral gain K_I	0.002	−0.0000
Derivative gain K_d	0.02	−0.0161

3) Gain Scheduling LQI: Gain scheduling is implemented as a set of local linear-quadratic-integral (LQI) controllers, each tuned for a different operating point of the glucose–insulin system to cover a glucose range between 80 and 300 mg/dL. Five operating regions are defined for the Bergman model and twelve for the Hovorka model, the latter using a finer spacing because the coarser configuration degraded performance. Each local controller results from linearizing the model around a physiological equilibrium characterized by a specific glucose level, basal plasma insulin, and no meal disturbance. The active controller is selected through con-tinuous interpolation based on a scheduling variable defined as a weighted combination of the measured glucose and the reference, which smooths the transitions between regions and anticipates incoming changes. The weighting matrices Q and R are adjusted per region to balance tracking error against control effort, increasing the penalty on control action near hypoglycemia and allowing more aggressive actions during hyperglycemia. In the Bergman model only four of the five states are fed back, since D˙ (t) is treated as an exgenous disturbance. Tables IV and V list the resulting gain sets.

TABLE IV

LQI GAINS PER OPERATING POINT FOR THE BERGMAN MODEL

G (mg/dL)	K_1	K_2	K_3	K_4	K_i
80	−239.62	81948.92	8.75	−0.03	−31.62
100	−227.91	91944.30	9.29	−0.03	−31.62
140	−211.28	109376.67	10.17	−0.03	−31.62
200	−335.30	190804.36	13.77	−0.05	−63.24
300	−305.50	234789.63	15.30	−0.05	−63.24

TABLE V

LQI GAINS PER OPERATING POINT FOR THE HOVORKA MODEL

G	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_i
80	-5.4×10^{-4}	-4.5×10^{-4}	9.8×10^{-6}	2.5×10^{-5}	4.5×10^{-4}	0.44	0.31	0.02	-5.4×10^{-6}
100	-6.1×10^{-4}	-5.0×10^{-4}	1.1×10^{-5}	2.9×10^{-5}	5.3×10^{-4}	0.59	0.39	0.03	-5.4×10^{-6}
120	-6.2×10^{-4}	-5.2×10^{-4}	1.1×10^{-5}	3.1×10^{-5}	5.8×10^{-4}	0.71	0.46	0.03	-5.4×10^{-6}
140	-6.4×10^{-4}	-5.3×10^{-4}	1.1×10^{-5}	3.3×10^{-5}	6.3×10^{-4}	0.84	0.53	0.03	-5.4×10^{-6}
160	-6.5×10^{-4}	-5.4×10^{-4}	1.2×10^{-5}	3.5×10^{-5}	6.8×10^{-4}	0.96	0.59	0.03	-5.4×10^{-6}
180	-6.1×10^{-4}	-5.1×10^{-4}	1.2×10^{-5}	3.5×10^{-5}	6.8×10^{-4}	1.01	0.61	0.03	-5.4×10^{-6}
200	-6.2×10^{-4}	-5.3×10^{-4}	1.2×10^{-5}	3.7×10^{-5}	7.3×10^{-4}	1.13	0.66	0.03	-5.4×10^{-6}
220	-6.4×10^{-4}	-5.4×10^{-4}	1.3×10^{-5}	3.9×10^{-5}	7.7×10^{-4}	1.26	0.70	0.03	-5.4×10^{-6}
240	-6.5×10^{-4}	-5.6×10^{-4}	1.3×10^{-5}	4.0×10^{-5}	8.1×10^{-4}	1.38	0.74	0.03	-5.4×10^{-6}
260	-6.6×10^{-4}	-5.6×10^{-4}	1.3×10^{-5}	4.1×10^{-5}	8.4×10^{-4}	1.48	0.77	0.03	-5.4×10^{-6}
280	-6.8×10^{-4}	-5.8×10^{-4}	1.3×10^{-5}	4.3×10^{-5}	9.0×10^{-4}	1.64	0.82	0.03	-5.4×10^{-6}
300	-6.9×10^{-4}	-5.9×10^{-4}	1.4×10^{-5}	4.5×10^{-5}	9.4×10^{-4}	1.77	0.85	0.03	-5.4×10^{-6}

4) Nonlinear MPC: Because both physiological models are nonlinear, the predictive controller is formulated as a nonlinear MPC without relying on a fixed linearization point. At each sampling instant the strategy minimizes a discrete quadratic cost that penalizes the deviation of glucose from its reference and the abrupt changes in insulin infusion, subject to the actuator limits:

$$\min_{u} \sum_{k=1}^{N_p} W_Q e(k)^2 + \sum_{k=1}^{N_c} W_R \Delta u(k)^2, \qquad \text{s.t. } 0 \le u(k) \le u_{\max}, \quad u_{\max} = 0.032 \text{ U/min}, \tag{18}$$

where e(k) is the glycemic tracking error and Δu(k) the change in infusion rate. The prediction horizon is Np = 40 steps (i.e., 40 min at the sampling time Ts = 1 min), capturing the delayed dynamics of insulin action, while

the control hori-zon is kept shorter than Np to limit computational complexity. The predicted glucose trajectory of each model is obtained by numerically integrating its nonlinear state equations with the explicit Runge–Kutta (4,5) scheme implemented in MATLAB ode45, so the discrete cost in Eq. (18) is evaluated over each sampling interval of Ts = 1 min. Since the only manipulated variable is exogenous insulin, the weights WQ and WR are implemented as positive scalars that balance tracking perfor-mance against control effort. Table VI summarizes the tuning used for both models.

TABLE VI
NONLINEAR MPC PARAMETERS FOR THE BERGMAN AND HOVORKA MODELS

Parameter	Bergman	Hovorka
Sampling time T_s [min]	1	1
Prediction horizon N_p [steps]	40	40
Control horizon N_c [steps]	5	5
Glucose weight W_Q	400	20
Control-effort weight W_R	50	10
Prediction integrator	ode45	ode45

C. Simulation and Evaluation Framework

1) Glycemic Reference and Metabolic Stress Scenarios: Six simulation scenarios were designed to assess the response of the control strategies under different physiological and clinical conditions in patients with T1D: standard, fasting, high car-bohydrate (CHO) intake, snacks, severe hyperglycemia, and severe hypoglycemia. Each run covers a 24 h period driven by the smooth, continuous glycemic reference depicted in Fig. 3, which reproduces realistic physiological variations without abrupt transitions. The biomedical control literature

supports the use of smoothed reference trajectories when designing drug - infusion systems [14],while clinical indices should be kept within a safe range to avoid sudden glycemic swings [15].

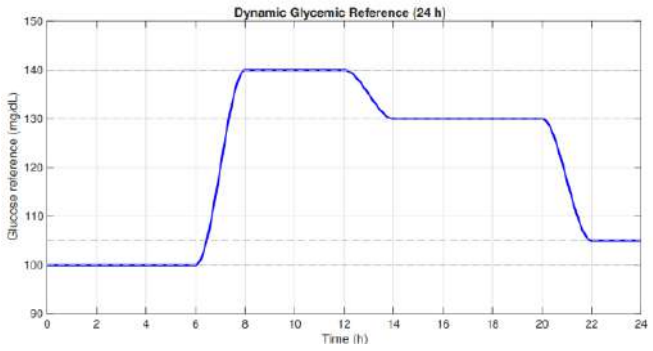


Fig. 3. Glycemic reference signal used in the simulations (24 h, 90–150 mg/dL).

The six scenarios are defined as follows:

- **Standard.** A typical day for a patient with T1D under normal physiological conditions, with three main meals at 8:00, 12:00, and 20:00 corresponding to 40, 60, and 50 g of CHO. Initial conditions are G0 = 100 mg/dL and I0 = 0 mU/L.
- **Fasting**. Prolonged fast with no CHO intake over the 24 h period, evaluating the ability to maintain glycemia within range using only basal exogenous insulin (G0 = 100 mg/dL, I0 = 0 mU/L).
- **High CHO intake.** Three main meals at 8:00, 12:00, and 20:00 with 70, 80, and 90 g of CHO, respec-tively, stressing postprandial compensation and insulin efficiency (G0 = 100 mg/dL, I0 = 0 mU/L).
- **Snacks.** Additional mid-morning, mid-afternoon, and post-dinner snacks of 20 g of CHO each, emulating frequent low-magnitude disturbances (G0 = 100 mg/dL, I0 = 0 mU/L).
- **Severe hyperglycemia.** The system starts from a severe hyperglycemic condition of G0 = 300 mg/dL with no CHO intake during the first hours, isolating the response that drives glucose back toward the normoglycemic range without overcorrection.
- **Severe hypoglycemia.** The system starts below the phys-iological range at G0 = 50 mg/dL with no CHO in-take during the first hours, assessing recovery without overcompensation or unstable oscillations—a clinically critical safety scenario [16].

2) Clinical, Control, Computational, and Cost Metrics: The strategies were assessed from a clinical, control, com-putational, and economic perspective. Clinical metrics [17] quantify safety, stability, and effectiveness in terms of pa-tient well-being. The percentage of time in the recommended physiological range of 70–180 mg/dL, with values ≥ 70% indicating stable regulation [18], is given by

$$\mathrm{TIR} = \frac{100}{T}\int_0^T I\{70 \le G(t) \le 180\}\,dt, \quad (19)$$

where $I\{\cdot\}$ is the indicator function. Time below and above range,

$$\mathrm{TBR} = \frac{100}{T}\int_0^T I\{G(t) < 70\}\,dt, \quad (20)$$

capture the risk of hypoglycemia, which can compro-

mise neu-rological and cardiovascular function [19], and of prolonged hyperglycemia, associated with oxidative stress and progres-sive organ damage [20], [21]. The coefficient of variation quantifies relative glycemic variability.

$$\mathrm{TAR} = \frac{100}{T}\int_0^T I\{G(t) > 180\}\,dt, \quad (21)$$

where σG and μG are the standard deviation and mean of the glucose signal; values above 36% indicate greater metabolic instability and a higher probability of hypoglycemic episodes [22]. Finally, average and total insulin quantify the efficiency and operational cost of the therapy.

Control metrics [23] assess the efficiency of the control action in tracking the reference Gref (t) with error e(t) = G(t) – Gref (t):

$$\begin{aligned}\mathrm{IAE} &= \int_0^T |e(t)|\,dt,\\ \mathrm{ISE} &= \int_0^T e^2(t)\,dt,\\ \mathrm{ITAE} &= \int_0^T t\,|e(t)|\,dt.\end{aligned} \quad (22)$$

Here ISE penalizes large deviations (identifying oscillatory responses), while ITAE favors controllers that minimize per-sistent steady-state error.

Computational cost measures the total execution time of each controller under identical conditions, reflecting its suit-ability for embedded insulin-pump hardware where resources and energy are limited. The simulations were run on the platform summarized in Table VII. Monetary cost of insulin, a critical economic factor for T1D therapy, was computed from the international unit (IU) price. Although the production cost of rapid-acting insulin is very low (roughly USD $0.0037–0.0173 per IU [24]), consumer acquisition cost ranges from USD $0.04–0.12 per IU [25], [26]; a mean price of USD $0.05 was used in this study.

TABLE VII

HARDWARE CONFIGURATION USED FOR THE SIMULATIONS

Component	Specification
Operating system	Microsoft Windows 11 Home Single Language
MATLAB version	MATLAB R2024a (version 24.1)
Processor (CPU)	AMD Ryzen 5 4600H with Radeon Graphics
CPU frequency	3.0 GHz (base), up to 4.0 GHz (boost)
Physical / logical cores	6/12
RAM	8 GB
Graphics (GPU)	NVIDIA GeForce GTX 1650 Ti
GPU total memory	4.29 GB
GPU available memory	3.45 GB
GPU clock speed	1485 MHz

III. RESULTS

A. Standard Scenario

The quantitative metrics reported below summarize the be-havior of the four conventional controllers under the standard scenario. As expected from their structural simplicity, the ON/OFF and PID schemes exhibited markedly inferior clinical and control performance on both physiological models, with high glycemic variability. On the Bergman model the ON/OFF controller achieved a coefficient of variation (CV) of 51.0% and a time in range (TIR) of 90.1%, while the PID controller reached a CV of 19.6% and a TIR of 96.3%. The difference was far more pronounced on the Hovorka model, where the ON/OFF controller registered a CV of 138.8% and a TIR of only 19.8%, with 72.1% of the time spent below range — a clinically hazardous behavior — whereas the PID controller attained a TIR of 61.2% and a CV of 19.3%. Because of this poor performance, additional metrics such as total insulin, monetary cost, and simulation time were not further evaluated for the ON/OFF and PID controllers, as they would not meaningful information to the comparison. The complete standard-scenario metrics are listed in Table VIII.

TABLE VIII

STANDARD-SCENARIO PERFORMANCE METRICS OF THE ON/OFF AND PID CONTROLLERS ON THE BERGMAN AND HOVORKA MODELS

Model	Ctrl	TIR (%)	Hyper (%)	Hypo (%)	CV (%)	IAE (mg/dL · min)	ISE (mg/dL^2 · min)	ITAE (mg/dL · min^2)
Bergman	On/Off	90.1	9.9	0.0	51.0	35.27	3481.3	29337.7
Bergman	PID	96.3	3.7	0.0	19.6	28.85	1433.7	25448.2
Hovorka	On/Off	19.8	8.07	72.1	138.8	87.4	92.4	55527.8
Hovorka	PID	61.2	38.8	0.0	19.3	46.5	3006.3	34207.8

The gain-scheduling LQI and nonlinear MPC controllers, in contrast, produced stable and physiologically coherent regula-tion. On the Bergman model the gain-scheduling controller provided adequate global stabilization, although its perfor-mance exhibited a greater dependence on the operating point: under hyperglycemic initial conditions it showed the highest total time out of range (≈ 6.6%, comprising 5.4% below and 1.2% above range) and a transitory overshoot in which severe hypoglycemia (glucose below 54 mg/dL, level 2 of the

international consensus [17]) accounted for roughly 4% of the time, the remaining below-range excursions being reflected in the 5.4% hypoglycemia entry of Table IX. The nonlinear MPC, on the other hand, delivered the most balanced distribution of hypo- and hyperglycemic episodes and the lowest total insulin, reflecting a more efficient dosing policy, consistent with its explicit predictive optimization.

B. Behavior Under Stress Scenarios

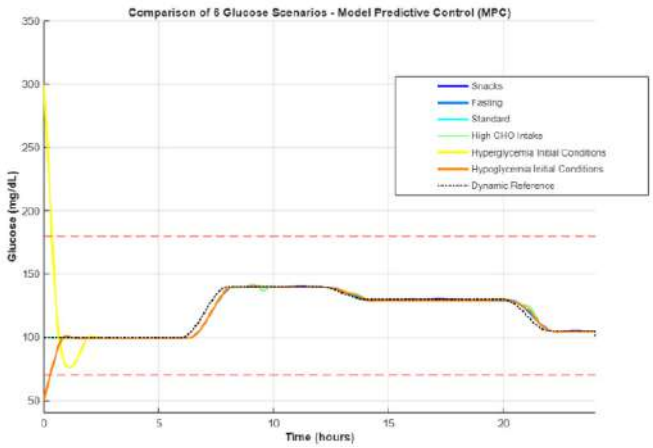


Fig. 4. Glycemic response of the nonlinear MPC controller across the six metabolic stress scenarios on the Bergman model. The nearly overlapping trajectories reflect anticipative control and reduced inter-scenario variability.

The six-scenario evaluation revealed substantial differences in robustness. On the Bergman model the MPC controller (Fig. 4) exhibited homogeneous trajectories across scenarios, with almost coincident curves even under high-carbohydrate intake and snacking, evidencing anticipative rejection of meal disturbances. The gain-scheduling controller (Fig. 5) also stabilized the system but with more pronounced transients under extreme initial conditions.

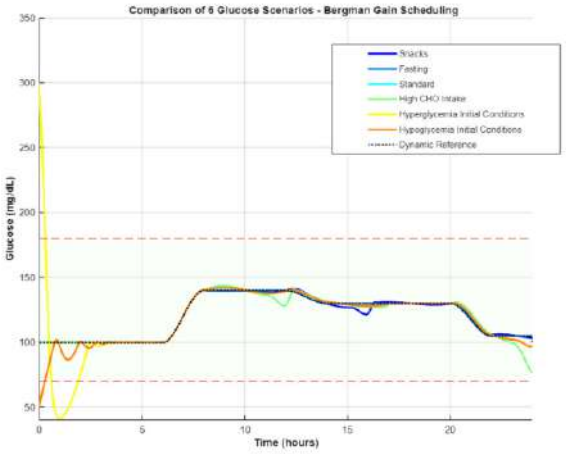


Fig. 5. Glycemic response of the gain-scheduling LQI controller across the six metabolic stress scenarios on the Bergman model. The controller stabilizes the system in all scenarios, although with more pronounced transients under extreme initial conditions.

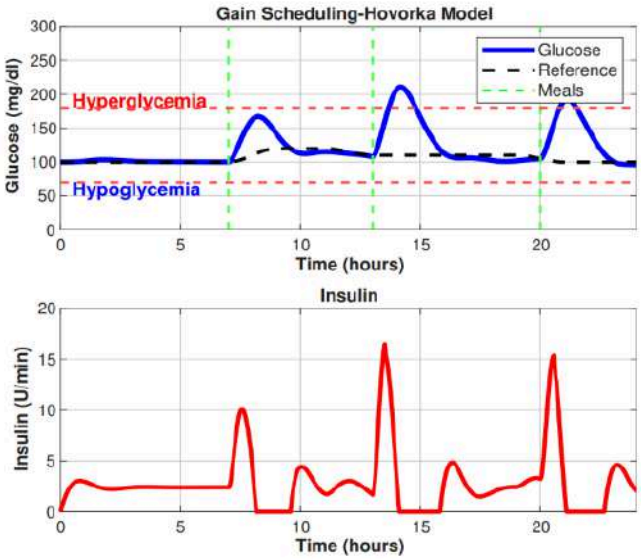


Fig. 6. Response of the gain-scheduling LQI controller on the Hovorka model in the standard scenario. The controller cannot reject postprandial hyperglycemic peaks.

The choice of physiological model proved decisive for the linear reactive strategies. Although the gain-scheduling con-troller behaved adequately on the Bergman model, it was insuf-ficient on the Hovorka model (Fig. 6), where it could not reject postprandial hyperglycemic peaks and only maintained basal conditions. This failure stems from the intrinsic limitation of gain-scheduled linear controllers when faced with significant pure delays (greater than 45 min) and the complex, coupled absorption dynamics of the Hovorka model, whose operating points are not adequately captured by local linearizations. On the Hovorka model (Fig. 7) the nonlinear MPC also lost robustness for the hyperglycemic initial-condition scenario, exhibiting an aggressive corrective action that induced severe hypoglycemia (≈ 22.3%, with a total time out of range near 25%, an IAE of 24.12, and an ISE exceeding 2000), reflecting the sensitivity of its predictive optimization to states far from the nominal and the critical role of constraint and weight tuning.

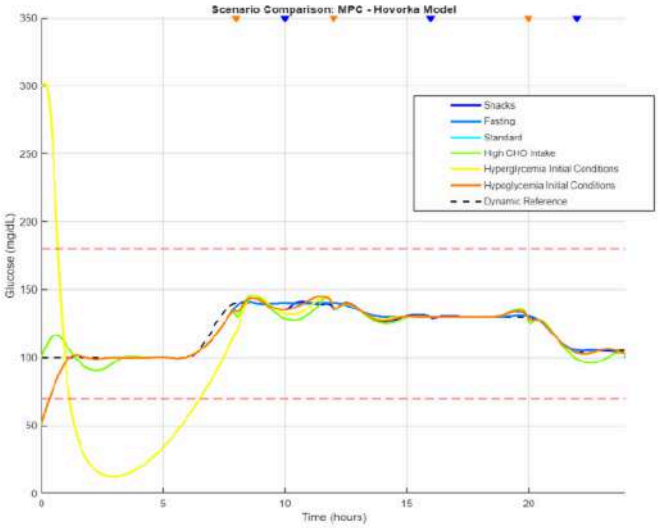


Fig. 7. Glycemic response of the nonlinear MPC controller across the six metabolic stress scenarios on the Hovorka model. Robust regulation is preserved in the routine scenarios, but the severe-hyperglycemia initial condition triggers an aggressive corrective action that drives glucose into hypoglycemia.

The clinical and control metrics across the six scenarios are summarized in Fig. 8 and Fig. 9, and consolidated numerically in Table IX (clinical) and Table X (control and dosing). On the Bergman model the gain-scheduling controller exhibited the largest total time out of range (≈ 6.6%) under the hyperglycemic initial condition, whereas the nonlinear MPC delivered the most balanced distribution of hypo- and hyper-glycemic episodes and the lowest total insulin, evidencing a more efficient dosing policy. On the Hovorka model, the MPC maintained time in range above 95% for all but the severe-hyperglycemia scenario, in which the aggressive corrective action drove glucose into hypoglycemia. The ISE remained low in routine scenarios but increased sharply under the extreme initial conditions for all strategies, reflecting a larger control effort and reduced energetic stability.

The total insulin delivered by each strategy is reported in Table XII. On the Bergman model, the nonlinear MPC delivered a slightly lower insulin total than the gain-scheduling controller in every routine scenario while maintaining time in range at 100%, confirming its dosing efficiency. Under the severe-hyperglycemia initial condition both controllers degraded, but the MPC preserved a larger time in range (97.09% vs. 93.3%) at the cost of a higher ISE. On the Hovorka model, the nonlinear MPC held time in range at 100% in the four routine scenarios, but the hyperglycemic initial condition induced severe hypoglycemia (22.30%) and a sharp rise of the error metrics, evidencing the sensitivity of receding-horizon control to states far from the nominal operating point.

TABLE IX

MASTER CLINICAL METRICS PER SCENARIO

Scenario	GS-LQI (Bergman)				MPC (Bergman)				MPC (Hovorka)			
	TIR (%)	Hypo (%)	Hyper (%)	CV (%)	TIR (%)	Hypo (%)	Hyper (%)	CV (%)	TIR (%)	Hypo (%)	Hyper (%)	CV (%)
Standard	100	0.0	0.0	13.70	100	0.0	0.0	13.65	100	0.0	0.0	13.78
Fasting	100	0.0	0.0	13.40	100	0.0	0.0	13.66	100	0.0	0.0	13.66
With snacks	100	0.0	0.0	13.50	100	0.0	0.0	13.48	100	0.0	0.0	13.20
High CHO	100	0.0	0.0	14.10	100	0.0	0.0	13.43	100	0.0	0.0	13.85
I.C.: Hyper	93.3	5.4	1.2	22.80	97.09	0.0	2.91	22.93	74.87	22.30	2.84	47.21
I.C.: Hypo	98.9	1.1	0.0	15.40	94.93	5.07	0.0	18.13	98.61	1.31	0.0	15.67

TABLE X

MASTER CONTROL METRICS PER SCENARIO

Scenario	GS-LQI (Bergman)			MPC (Bergman)			MPC (Hovorka)		
	IAE	ISE	ITAE	IAE	ISE	ITAE	IAE	ISE	ITAE
Standard	1.11	3.41	982.8	1.32	5.84	812.4	3.06	29.82	1133.8
Fasting	0.19	0.82	138.4	1.24	5.33	740.1	2.25	26.95	500.4
With snacks	1.13	3.58	1074.8	1.13	4.28	768.1	2.37	16.11	1028.7
High CHO	2.22	19.17	2176.6	1.39	6.13	1045.3	4.31	41.74	2321.8
I.C.: Hyper	6.14	427.04	1254.6	6.15	784.43	1284.7	24.12	2076.14	5363.6
I.C.: Hypo	3.01	30.31	1011.3	4.76	123.12	1490.1	2.44	36.00	1040.2

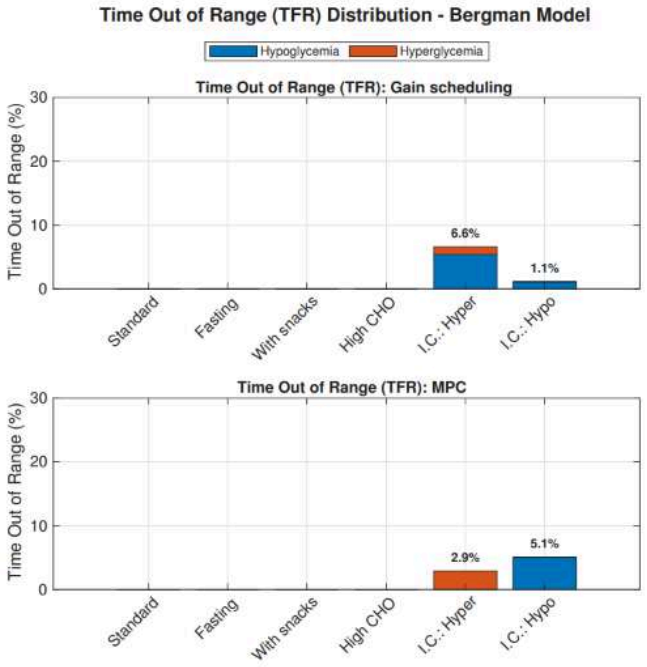


Fig. 8. Clinical ranges for the Bergman model across the six stress scenarios. Time out of range remains near zero in routine scenarios and increases only under extreme initial conditions.

C. Computational and Monetary Cost

Table XI reports the execution-time statistics for the con-trollers under identical conditions and hardware (Table VII). Because the gain-scheduling controller failed to maintain glycemia regulation on the Hovorka model (Fig. 6), its simulation-time and monetary-cost statistics on that model are not reported, following the same rationale applied to the ON/OFF and PID controllers: such figures would not correspond to a valid closed-loop configuration and would not contribute

meaningful information to the cost comparison.

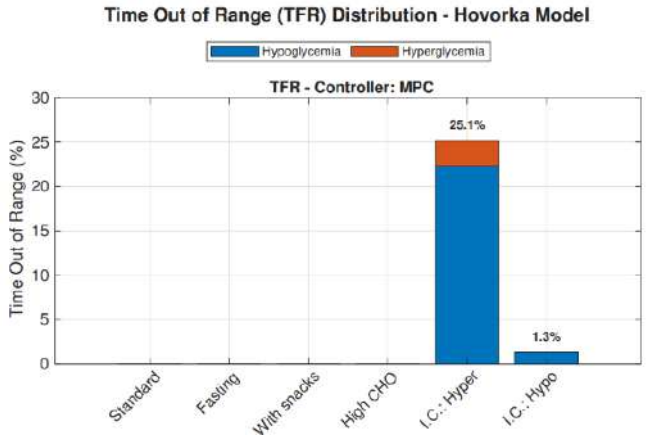


Fig. 9. Time out of range distribution for the nonlinear MPC controller on the Hovorka model across the six stress scenarios. Hypoglycemia dominates only under the severe-hyperglycemia initial condition, where the aggressive corrective action induces excessive insulin delivery.

On the Bergman model the gain-scheduling controller ran in approximately 8.6 s on average, whereas the nonlinear MPC required, on average, 210.52 s (median 206 s) because of its iterative on-line optimization. On the more complex Hovorka model the MPC execution time rose to a mean of 568 s, underscoring the computational burden associated with receding-horizon control on higher-order physiological models. The corresponding execution-time distributions on the Bergman model are depicted in Fig. 10, which evidences the substantially higher and more variable computational demand of the MPC relative to the gain-scheduling strategy. The MPC execution-time distribution on the Hovorka model is shown in Fig. 11, confirming the higher computational burden of receding-horizon control on the higher-order model.

TABLE XI
SIMULATION-TIME STATISTICS (SECONDS) OF THE CONVENTIONAL CONTROLLERS

Model	Controller	Mean	Median	Std	Min	Max	IQR
Bergman	G. Scheduling	8.59	8.47	0.70	7.90	11.05	0.77
Bergman	MPC	210.52	206.00	18.77	180.00	260.00	29.00
Hovorka	MPC	568.00	566.00	13.41	540.00	610.00	11.25

From an economic standpoint (Table XII), all closed-loop strategies markedly reduced the insulin cost relative to open-loop (manual) administration, which consumed 64 U at a cost of USD $1168 per year on the Bergman model. In the standard scenario the gain-scheduling controller used 25.2 U (USD $1.26/day, $460/year), whereas the nonlinear MPC used only 10.12 U (USD $0.51/day, $185/year) on the Bergman model and 26.54 U (USD $1.33/day, $484/year) on the Hovorka model, confirming that automatic regulation translates into a significant reduction of therapy cost compared with manual management of T1D.

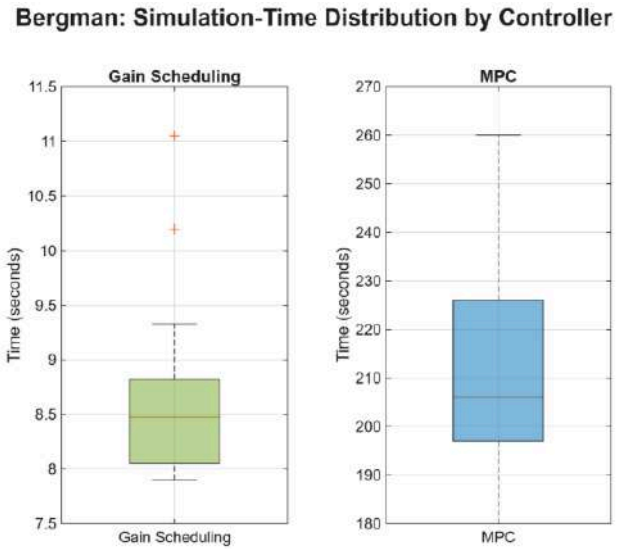


Fig. 10. Distribution of simulation time on the Bergman model. The gain-scheduling controller exhibits moderate variability around a median of ≈8.5 s, whereas the nonlinear MPC shows markedly higher and more variable execution times.

TABLE XII
ESTIMATED MONETARY COST OF INSULIN PER CONTROLLER IN THE STANDARD SCENARIO

Model	Controller	Insulin (U)	1 day (USD)	1 year (USD)
Bergman	Open-loop (manual)	64.00	$3.20	$1168
Bergman	Gain scheduling	25.20	$1.26	$460
Bergman	MPC	10.12	$0.51	$185
Hovorka	MPC	26.54	$1.33	$484

IV. CONCLUSIONS

This work compared four conventional closed-loop glucose-control strategies — ON/OFF with hysteresis, PID with feed-forward and anti-windup, gain-scheduling LQI, and nonlinear MPC — on the modified Bergman and Hovorka physiological models under a standard scenario and five metabolic stress scenarios, using clinical, control, computational, and economic metrics.

The results confirm that the choice of physiological model depends on the design objective and context. Owing to its re-duced structure and lower dynamical complexity, the Bergman model is well suited for the preliminary design, analysis, and validation of control strategies, whereas the more detailed Hovorka model provides greater physiological fidelity and is more appropriate for evaluating clinically complex scenarios.

The simple ON/OFF and PID controllers proved insufficient on both models, and especially hazardous on the Hovorka model, where the ON/OFF scheme remained below range for 72.1% of the simulation with a CV of 138.8%.

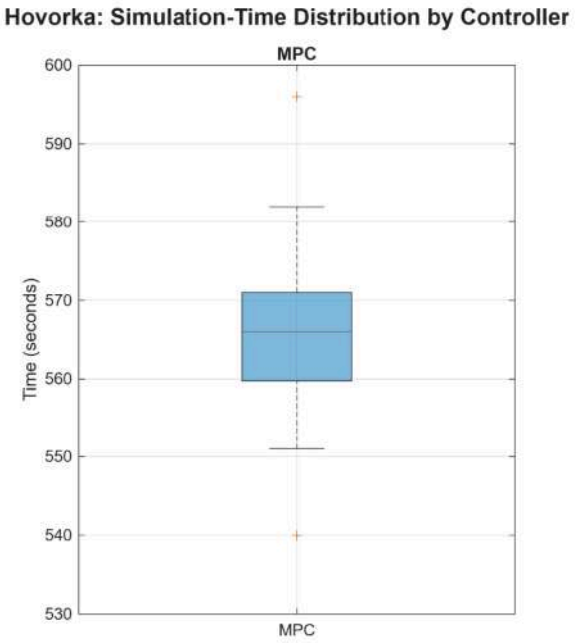


Fig. 11. Distribution of simulation time of the nonlinear MPC on the Hovorka model. The execution time is markedly higher than on the Bergman model, evidencing the computational burden of receding-horizon control on the higher-order physiological model.

The gain-scheduling controller achieved acceptable regula-tion on the Bergman model but failed on the Hovorka model, exposing the inherent limitation of linear reactive strategies against significant pure delays and complex absorption dynam-ics. The nonlinear MPC delivered the best overall compromise on the Bergman model, with the lowest total insulin, the most balanced hypo/hyperglycemic distribution, and lower integral error metrics under the high-CHO disturbance. However, it required careful tuning of constraints, horizons, and weights to preserve robustness under extreme hyperglycemic initial conditions on the Hovorka model, where an overly aggressive optimization induced severe hypoglycemia. Finally, all auto-matic strategies substantially reduced insulin consumption and monetary cost relative to open-loop manual administration, confirming the economic benefit of closed-loop control for T1D therapy. The results are based on a single deterministic 24-h simulation per scenario per model; inter-patient variabil-ity and sensor-noise effects were not assessed, so the reported metrics should be interpreted as nominal performance indica-tors for a representative virtual patient rather than population-level guarantees.

REFERENCES

[1].World Health Organization, “Diabetes,” 2025. [Online]. Avail-able: https://www.who.int/es/news-room/fact-sheets/detail/diabetes [Ac-cessed: Nov. 6, 2025].

[2].PAHO, “Ante el aumento en el nu´mero de casos en todo el mundo... es necesario tomar medidas urgentes contra la diabetes,” 2024. [Online]. Available: https://www.paho.org/es/noticias/14-11-2024.

[3].B. W. Bequette, “Challenges and recent progress in the development of a closed-loop artificial pancreas,” Annu. Rev. Control, vol. 36, no. 2, pp. 255–266, 2012.

[4].I. S. Mughal, L. Patane`, and R. Caponetto, “A comprehensive review of models and nonlinear control strategies for blood glucose regulation in artificial pancreas,” Annu. Rev. Control, vol. 57, Art. no. 100937, 2024.

[5].C. Dalla Man et al., “The UVA/PADOVA Type 1 Diabetes Simulator: New Features,” J. Diabetes Sci. Technol., vol. 8, no. 1, pp. 26–34, 2014.

[6].S. Mehmood et al., “Artificial pancreas control strategies used for type 1 diabetes control and treatment: A comprehensive analysis,” Appl. Syst. Innov., vol. 3, no. 3, p. 31, 2020.

[7].B. W. Bequette, “A Critical Assessment of Algorithms and Challenges in the Development of a Closed-Loop Artificial Pancreas,” Diabetes Technology & Therapeutics, vol. 7, no. 1, pp. 28–47, 2005.

[8].G. Marchetti et al., “An improved PID switching control strategy for type 1 diabetes,” IEEE Trans. Biomed. Eng., vol. 55, no. 3, pp. 857–865, 2008.

[9].R. Hovorka et al., “Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes,” Physiol. Meas., vol. 25, no. 4, pp. 905–920, 2004.

[10].R. N. Bergman, L. S. Phillips, and C. Cobelli, “Physiologic evaluation of factors controlling glucose tolerance in man,” J. Clin. Invest., vol. 68, no. 6, pp. 1456–1467, 1981.

[11].C. Zamarro´n Sobrinos, “Modelos matema´ticos del sistema regulador glucosa-insulina en pacientes diabe´ticos,” Masters thesis, Univ. Pontificia Comillas, Madrid, Espan˜a, 2021.

[12].S. Hu, M. Dunlavey, S. Guzy, and N. Teuscher, “A distributed delay approach for modeling delayed outcomes in pharmacokinetics and pharmacodynamics studies,” J. Pharmacokinet. Pharmacodyn., vol. 45, no. 2, pp. 285–308, 2018.

[13].G. Freckmann, U. Kamecke, D. Waldenmaier, C. Haug, and R. Ziegler, “Accuracy of bolus and basal rate delivery of different insulin pump systems,” Diabetes Technol. Ther., vol. 21, no. 4, pp. 201–208, 2019.

[14].R. Yang, T. J. Tarn, and M. Zhang, “Event-based feedforward control of a drug delivery system for diabetes care,” IFAC Proc. Vol., vol. 41, no. 2, pp. 5393–5398, 2008.

[15].A. Bertachi, L. Biagi, A. Beneyto, and J. Veh´ı, "Dynamic rule-based algorithm to tune insulin-on-board constraints for a hybrid artificial pancreas system," J. Healthc. Eng., vol. 2020, p. 1414597, 2020.

[16].American Diabetes Association Professional Practice Committee, "6. Glycemic targets: Standards of Medical Care in diabetes—2022," Dia-betes Care, vol. 45, no. Suppl. 1, pp. S83–S96, 2022.

[17].T. Battelino et al., "Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range," Diabetes Care, vol. 42, no. 8, pp. 1593–1603, 2019.

[18].American Diabetes Association Professional Practice Committee, "6. Glycemic goals, hypoglycemia, and hyperglycemic crises: Standards of care in diabetes—2026," Diabetes Care, vol. 49, no. Suppl. 1, pp. S132–S149, 2026.

[19].P. E. Cryer, L. Axelrod, A. B. Grossman, et al., "Evaluation and management of adult hypoglycemic disorders: An Endocrine Society clinical practice guideline," J. Clin. Endocrinol. Metab., vol. 94, no. pp. 709–728, 2009.

[20].The Diabetes Control and Complications Trial Research Group, "The effect of intensive treatment of diabetes on the development and progres-sion of long-term complications in insulin-dependent diabetes mellitus," New Engl. J. Med., vol. 329, no. 14, pp. 977–986, 1993.

[21].M. Mouri and M. Badireddy, "Hyperglycemia," in StatPearls. Treasure Island, FL, USA: StatPearls Publishing, 2025

[22].L. Monnier, C. Colette, and D. R. Owens, "Glycemic variability: The third component of the dysglycemia in diabetes. Is it important? How to measure it?," J. Diabetes Sci. Technol., vol. 2, no. 6, pp. 1094–1100, 2008.

[23].K. Ogata, Modern Control Engineering, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[24].D. Gotham, M. J. Barber, and A. Hill, "Production costs and potential prices for biosimilars of human insulin and insulin analogues," BMJ Glob. Health, vol. 3, no. 5, p. e000850, 2018.

[25].A. W. Mulcahy and D. Schwam, "Comparing insulin prices in the United States to other countries: Updated results using 2022 data," Rand Health Q., vol. 11, no. 3, p. 3, 2024.

[26].M. Ewen, H.-J. Joosse, D. Beran, and R. Laing, "Insulin prices, avail-ability and affordability in 13 low-income and middle-income countries," BMJ Glob. Health, vol. 4, no. 3, p. e001410, 2019.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

A Predictive Model for Energy Consumption in P2P Networks Based on Knowledge Discovery and Artificial Intelligence

Un modelo predictivo para el consumo energético en redes P2P basado en el descubrimiento de conocimiento y la inteligencia artificial

Kandel L. Yandar[1]*, Oscar Revelo Sánchez [2] Manuel Bolaños González[3]

[123]Department of Systems Engineering, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia. klyandar23b@udenar.edu.co, orevelo@udenar.edu.co, mbolanos@udenar.edu.co

*Autor de correspondencia: klyandar23b@udenar.edu.co

Citación: K. L. Yandar, O. Revelo Sánchez y M. Bolaños González, "A Predictive Model for Energy Consumption in P2P Networks Based on Knowledge Discovery and Artificial Intelligence," Revista de Ingeniería, vol. 4, no. 1, pp. 45-60, Nov. 2026.

***Abstract*—** Electricity is essential for social and economic development; however, the integration of renewable energy sources and decentralized peer-to-peer (P2P) networks introduces new challenges in consumption management. This research applied the Knowledge Discovery in Databases (KDD) process through four phases: a systematic literature review to identify AI techniques and relevant prediction factors, an analysis of these techniques within KDD, predictive model development, and validation using real-world data. Multivariate time series combining electrical, meteorological, and temporal variables were collected from five institutions in Nariño, Colombia, at two-minute intervals. Two deep learning architectures were implemented: a hybrid LSTM–CNN model and a Transformer model based on attention mechanisms. Performance was analyzed for each node and validated using unseen temporal data, with statistical comparison through the Wilcoxon test. Both models achieved R^2 values above 0.86 and a mean absolute error (MAE) of 0.0072 kWh, demonstrating their ability to capture complex consumption patterns across different operational contexts. These results support the use of AI for energy planning and management in decentralized power grids.

***Keywords*—** Deep learning, Knowledge discovery, Energy consumption prediction, Peer-to-peer networks, Multivariate time series

***Resumen*—** La electricidad es esencial para el desarrollo social y económico; sin embargo, la integración de energías renovables y redes descentralizadas de intercambio entre pares (P2P) plantea nuevos desafíos en la gestión del consumo. Esta investigación aplicó el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD) en cuatro fases: revisión sistemática de literatura, análisis de técnicas de inteligencia artificial, desarrollo de modelos predictivos y validación con datos reales. Se recopilaron series temporales de variables eléctricas, meteorológicas y temporales de cinco instituciones de Nariño, Colombia, a intervalos de dos minutos. Se implementaron dos arquitecturas de aprendizaje profundo: un modelo híbrido LSTM-CNN y un Transformer basado en mecanismos de atención. El rendimiento se evaluó por nodo con datos temporales no vistos y la prueba estadística de Wilcoxon. Ambos modelos obtuvieron valores de R^2 superiores a 0,86 y un error absoluto medio (MAE) de 0,0072 kWh, demostrando su capacidad para identificar patrones complejos de consumo y apoyar la planificación y gestión de redes eléctricas descentralizadas.

***Palabras clave*—** Aprendizaje profundo, descubrimiento de conocimiento, predicción del consumo energético, redes entre pares (P2P), series temporales multivariadas.

I. INTRODUCTION

Electricity is a fundamental resource for economic and social development, as it supports a wide range of productive, industrial, and domestic activities [1]. However, reliance on fossil fuels as the primary source of power generation poses risks of resource depletion and has a significant environmental impact associated with emissions that contribute to climate change [2].

Against this backdrop, the transition to renewable energy sources has emerged as a necessary strategy for moving toward more sustainable, resilient, and efficient energy systems. This process is not limited solely to transforming electricity generation mechanisms; it also involves rethinking traditional distribution and consumption models, promoting decentralized systems capable of adapting to the inherent variability of renewable sources and to growing energy demand [3].

In light of this transformation of the energy sector, electric systems are now integrating new renewable sources. Unlike traditional centralized generation systems, these environments exhibit greater variability and uncertainty due to the intermittent nature of resources such as solar and wind power. This situation complicates operational planning, the balance between supply and demand, and the efficient management of available energy. Although artificial intelligence techniques have been widely investigated for electricity consumption and generation forecasting [4], their application in emerging energy-sharing schemes, such as peer-to-peer (P2P) networks, presents specific challenges. Previous studies have identified challenges in the implementation and scaling of P2P energy systems [5], while research integrating short-term load forecasting into P2P energy trading highlights the need to further investigate the relationship between predictive models and decentralized energy management [6]. Consequently, it is necessary to develop predictive models capable of anticipating energy consumption patterns and supporting decision-making in these new distributed environments.

In this context, electricity consumption plays a key role in the planning and operation of electric power systems, as it reflects the demand that grids must meet at any given moment. This behavior is not constant, as it depends on various factors, including weather conditions, user habits, the availability of electrical equipment, and the economic and social characteristics of each region. For this reason, understanding how energy demand evolves and anticipating its variations is essential for improving the management of the electric system and optimizing grid operations.

In recent years, P2P networks have begun to establish themselves as an innovative approach in the energy sector. This model allows users to actively participate in energy exchange, so that consumers can also act as producers, or prosumers. This facilitates direct energy transactions among participants, reducing dependence on centralized intermediaries [5]. However, for these types of networks to function efficiently, tools capable of anticipating energy consumption patterns are necessary, as accurate demand forecasting helps maintain the system's balance and stability.

Researchers have widely addressed energy consumption forecasting using artificial intelligence and machine learning techniques because they can model complex relationships in energy time series. Commonly investigated approaches include artificial neural networks (ANNs), support vector machines (SVMs), decision trees, Random Forests, and hybrid methods that combine different learning algorithms. Previous studies have applied these techniques to energy forecasting in residential buildings, large facilities, hospitals, and other power consumption scenarios, reporting improvements in predictive performance depending on the model and application context [7]–[10].

In recent years, deep learning models have become particularly relevant in this field. Architectures such as Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), and hybrid CNN-LSTM models have demonstrated strong predictive performance in modeling temporal patterns and dependencies in energy consumption time series. Research conducted in university microgrids, residential environments, and other energy-related contexts has reported promising results using these architectures for electricity consumption forecasting [11]–[14].

More recently, advanced models based on attention mechanisms and deep learning architectures have been investigated for multivariate and multi-step energy forecasting. Notably, Informer and hybrid models have been explored to improve the modeling of complex patterns in long time series and enhance prediction accuracy over extended forecasting horizons [15]. These developments reflect the increasing use of advanced deep learning architectures for the analysis and forecasting of energy time series.

Despite these advances, most research has been conducted in smart buildings, residential homes, university campuses, or conventional power grids. Studies focused on P2P networks, where users can act simultaneously as both energy consumers and producers, are less common. Among the studies identified, Manchalwar et al. [6] analyzed short-term load forecasting in a P2P network using techniques such as SVM, decision trees, and ANNs, finding that ANNs performed best. However, evidence remains limited on applying artificial intelligence-based knowledge discovery models to predict energy consumption in real P2P networks. Furthermore, few studies employ multivariate information from multiple measu-

rement nodes and compare deep learning architectures within a single methodological framework. This gap limits our understanding of how these models behave in distributed energy environments and motivates the development of approaches tailored to P2P networks.

Within this framework, the project "Desarrollo de un modelo transaccional de energía no convencional de múltiples agentes para el Departamento de Nariño (MTE)" lays the groundwork for implementing a P2P network pilot in the region[16]. In line with this project, the present research proposes an artificial intelligence-based knowledge discovery model for predicting energy consumption in a P2P network. The model integrates electrical, meteorological, and temporal variables into a KDD process and evaluates the performance of LSTM-CNN and Transformer architectures using real-world data collected from five institutions in the department of Nariño. The aim is to identify consumption patterns and provide evidence regarding the applicability of deep learning techniques in decentralized energy environments.

The remainder of this article is organized as follows. Section 2 describes the proposed methodology. Section 3 presents the results and discussion from the study. Finally, Section 4 presents the conclusions of this research.

II. METHODOLOGY

The research methodology was structured into four main phases: exploration, analysis, construction, and validation.

A. Exploration Phase

During the exploratory phase, a Systematic Literature Review (SLR) was conducted to identify the main approaches and techniques used in predicting electricity consumption. The results of this stage were previously reported by the authors in an earlier study [4], which served as the conceptual and methodological foundation for this research.

The SLR was conducted following the guidelines proposed by Kitchenham and Charters [17] and included the following activities:

1. **Planning:** definition of research questions, inclusion and exclusion criteria, search strings, and selection of scientific databases.
2. **Implementation:** search for studies, removal of duplicate records, primary and secondary screening, full-text review, and quality assessment of the articles.
3. **Report:** extraction of relevant information, comparative analysis, and consolidation of findings related to artificial intelligence techniques applied to energy forecasting.

B. Analysis Phase

During the analysis phase, the artificial intelligence techniques reported in the literature were identified and classified according to the stages of the KDD process: selection, preprocessing, transformation, data mining, and evaluation. This analysis identified the most commonly used and effective strategies for predicting energy consumption, forming the basis for the proposed model.

C. Construction Phase

During the construction phase, the artificial intelligence-based knowledge discovery model was developed by systematically applying the stages of the KDD process proposed by Fayyad et al. [18].

The main activities carried out were:

1. **Data selection:** the dataset used in this research comes from the project "Desarrollo de un modelo transaccional de energía no convencional de múltiples agentes para el Departamento de Nariño (MTE)," an initiative that began in late 2024 with the implementation of a monitoring and generation infrastructure based on non-conventional energy sources (FNCE). This infrastructure included the installation of smart meters, photovoltaic inverters, and weather stations at five institutions in the municipality of Pasto (Colombia): the Universidad de Nariño (UDENAR), the Universidad Cooperativa de Colombia (UCC) Pasto campus, Universidad Mariana (UNIMAR), Universidad CESMAG (San Damián Campus), and the Hospital Universitario Departamental de Nariño (HUDN). For this study, records collected between January and May 2025 were used, with a sampling frequency of two minutes. Following the integration, cleaning, and consolidation of data from the different measurement nodes, a multivariate dataset was obtained, consisting of 34 predictor variables and one target variable corresponding to electricity consumption (energy_consumption_kWh). The variables included continuous numerical data associated with electrical and meteorological quantities, as well as temporal and categorical variables derived from the operational context of each node.
2. **Preprocessing:** this was performed independently for each node in the P2P network to identify differences in data quality and availability. Initially, the records from smart meters, inverters, and weather stations were checked for missing values and inconsistencies. Subsequently, formats and units of measurement were standardized, converting the recorded power variables from watts (W) to kilowatts (kW). The different data sources were integrated using the `merge_asof` function from the Pandas library, with a maximum tolerance of 120 seconds to

merge temporally close records. Missing values generated during this process were handled using time-based linear interpolation, preserving the continuity of the time series and minimizing information loss.

3. **Transformation:** during this stage, new features were constructed from the original data, including the target variable energy_consumption_kWh, temporal variables derived from the date and time of recording, cyclical representations using sine and cosine functions, a weekend indicator variable, and the P2P network node identifier. Subsequently, feature selection was performed using correlation analysis and the Recursive Feature Elimination (RFE) method with Random Forest, resulting in a final set of 34 predictor variables. The data were divided chronologically into training (70%), validation (15%), and test (15%) sets, preserving the time-series nature of the data. The variables were then normalized using RobustScaler, whose parameters were calculated solely on the training set to avoid information leakage. Finally, time series were generated using sliding windows of 120 observations (4 hours of history), which were used as input for the LSTM–CNN and Transformer models.
4. **Data mining:** implementation of two deep learning architectures: a hybrid LSTM–CNN architecture and a model based on Transformer attention mechanisms. The models were trained using supervised learning to predict multivariate time series of energy consumption one step ahead, and the experiments were conducted in Python 3.9.7 using the libraries TensorFlow 2.6.0, Scikit-learn 1.6.1, Pandas 1.3.5, NumPy 1.19.5, and Matplotlib 3.6.2, within a JupyterHub environment.
5. **Evaluation:** model performance was evaluated using five metrics widely used in energy time series forecasting problems: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE), and adjusted MAPE.

MAE measures the average magnitude of the absolute differences between observed and predicted values. This metric quantifies the average error made by the model in the same units as the target variable (kWh), with values close to zero being desirable.

$$\mathrm{MAE} = \frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} |y_i - \hat{y}_i|$$

Where yi represents the actual value, $\hat{y}_i$ the predicted value, and n is the total number of observations.

RMSE evaluates the average magnitude of the error by assigning a greater penalty to large errors. This characteristic makes it a suitable metric for identifying models that are sensitive to significant deviations

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} (y_i - \hat{y}_i)^2}$$

Like the MAE, its results are expressed in kWh, and lower values indicate better performance.

R^2 measures the proportion of the variability in the actual data that can be explained by the predictive model. Values close to 1 indicate a high explanatory power, while values close to 0 reflect a low ability to represent the behavior of the analyzed variable

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Where $\bar{y}$ corresponds to the average of the observed values.

The MAPE expresses the average error as a percentage relative to the true value, allowing for a comparison of performance between datasets with different consumption levels.

$$\mathrm{MAPE} = \frac{100}{\mathbf{n}} \sum_{i=1}^{\mathbf{n}} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Lower values indicate better predictive accuracy. However, when actual values are close to zero, the MAPE may yield excessively high or even undefined results. For this reason, an adjusted MAPE was also calculated, excluding observations with consumption below 0.1 kWh. This variant provides a more representative measure of relative error in scenarios where energy demand is very low.

Taken together, these metrics allow for a complementary evaluation of the accuracy, explanatory power, and robustness of the proposed models, providing a comprehensive view of their predictive performance.

D. Validation Phase

Validation was performed using two complementary

strategies.

1. **Performance Validation:** an independent dataset corresponding to June was used, strictly adhering to the time sequence of the series to preserve chronological dependence and prevent information leakage. This validation made it possible to measure the generalization ability of the LSTM–CNN and Transformer models when applied to data not used during training.
2. **Computational performance validation:** the computational performance of both models was evaluated during the inference stage using different execution environments. This comparison allowed us to analyze the efficiency, robustness, and portability of the implemented architectures.

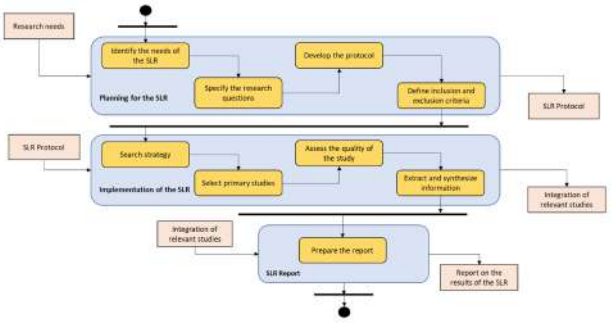


Fig 1. SLR Process. Source: Adapted from [19]

To determine whether the differences observed between the LSTM–CNN and Transformer models were statistically significant, a hypothesis test based on the nonparametric Wilcoxon signed-rank test was applied.

The analysis focused on the error metrics used to evaluate the models' predictive accuracy: MAE, RMSE, and Adjusted Mean Absolute Percentage Error (MAPEaj).

These metrics allow for quantifying the differences between actual and predicted values in both absolute and relative terms.

The null hypothesis (H_0) stated that there are no statistically significant differences between the results obtained by the LSTM–CNN and Transformer models, while the alternative hypothesis (H_1) posited the existence of significant differences between the two approaches.

The Wilcoxon test was selected because the results correspond to paired samples obtained from the same time series and nodes of the P2P network, and because it does not require the assumption of normality in the error distribution. This characteristic makes it suitable for datasets with small sample sizes and possible asymmetries.

The test was conducted independently for each evaluation metric during the evaluation and temporal validation phases. A significance level of ($\alpha = 0.05$) was adopted for all analyses. Statistical calculations were performed using the SciPy library in Python.

The null hypothesis was rejected when the p-value was less than the established significance level ($\alpha = 0.05$).

III. RESULTS

A. Exploratory Phase

In this initial phase, a literature review was conducted to identify and analyze the main AI techniques used in electricity forecasting. To this end, the guidelines proposed by Kitchenham and Charters [17], widely used in systematic review studies, were followed and adapted in accordance with the methodological proposal presented by Zapata and Barón[19].

This process allowed for the structured organization of the search, selection, and analysis of the most relevant studies for the development of the research. Figure 1 shows the literature review process carried out in this study.

1. Planning the SLR

The SLR was designed to address two research questions:

RQ1: What AI techniques are used for electricity forecasting?

RQ2: What are the key factors used in electricity forecasting?

In addition, inclusion and exclusion criteria were defined, covering the period 2013–2024, articles in English and Spanish, and studies applying AI to energy forecasting. Duplicates, off-topic articles, and secondary reviews were excluded.

2. Conducting the SLR

The search was conducted in the SCOPUS, ACM Digital Library, IEEE Xplore, and Google Scholar databases, using search terms such as: ("artificial intelligence techniques") AND ("energy management" OR "energy consumption" OR "energy prediction").

Initially, 216 records were identified; after screening and applying filters, this number was reduced to 100 articles. Subsequently, titles and abstracts were reviewed, and 68 documents were selected for full-text review. Finally, after applying quality criteria, 47 primary studies were included and formed the basis of the analysis.

3. SLR report

Based on the selected studies and in response to the defined research questions, it was found that the most commonly used techniques for predicting electricity consumption are:

- Artificial Neural Networks (ANNs), widely used for their ability to capture nonlinear temporal relations-

hips.

- Support Vector Machines (SVM and SVR), applied especially in scenarios involving short time series and meteorological variables.
- Hybrid models (LSTM-CNN, SARIMA-LSSVR, EEMD-Informer), which demonstrated greater accuracy by combining complementary techniques.

In addition, the SLR helped identify key factors in the design of prediction models, including:

- Data collection and quality: historical consumption time series and exogenous variables (meteorological and seasonal).
- Preprocessing: normalization, outlier removal, and handling of missing data.
- Data splitting: training, validation, and testing to ensure model generalization.

The findings of the SLR demonstrate that AI-based models are central to energy consumption forecasting. Among the techniques reviewed, neural networks and hybrid models showed the best performance, while SVM and SVR stood out for their effectiveness in scenarios with lower data volumes. These results served as the basis for the analysis phase, in which the most appropriate techniques for implementing the proposed knowledge discovery model were defined.

B. Analysis Phase

The KDD process was used as a framework to classify and synthesize how the literature addresses the prediction of electricity consumption. Based on the studies identified in the Systematic Literature Review, the findings were organized according to the main stages of the KDD process: data selection, preprocessing, transformation, data mining, and evaluation. This classification allowed for a structured analysis of the methodologies used in previous research. The results obtained for each of these stages are presented below.

1. Data Selection

Analysis of the reviewed studies reveals a notable diversity in the methods used to obtain energy consumption data. This variety allows for working with different levels of temporal and contextual granularity, depending on the information source used. Among the most common approaches are the following:

- Data collection via smart meters, which allows for the recording of consumption profiles with hourly or sub-hourly resolution in homes, buildings, or university campuses.
- Extraction of information from institutional systems, which integrate consumption histories with external variables such as weather conditions or calendar data.
- Monitoring in real or experimental environments through measurement campaigns conducted in hospitals, offices, or homes, which facilitates the incorporation of variables related to facility usage or occupancy patterns.
- Use of physical sensors or IoT devices, which allow for the recording of electrical variables, such as current, voltage, and power, as well as environmental variables, such as temperature, occupancy, or carbon dioxide levels, providing greater context to explain variations in consumption.
- Computational simulation, used when sufficient real-world data is not available, allows for the generation of synthetic time series for the initial evaluation of models, although with potential limitations in representing actual behavior.
- Use of historical repositories or databases, which facilitate access to extensive time series and allow for the analysis of consumption patterns over longer time horizons.

2. Data Preprocessing

In the studies analyzed, data preprocessing is recognized as an important stage for ensuring the quality and consistency of the data used in energy prediction models. Particular attention is given to common issues in energy-related datasets, including outliers, incomplete records, and other anomalies that may affect subsequent modeling stages [20], [21].

Based on the literature review, several strategies for improving data quality prior to model development were identified. One of the most common practices involves the detection and treatment of outliers. Saleh et al. [20], for example, incorporated a specific preprocessing stage for the identification and rejection of anomalous values in electrical load data. Similarly, Baseer et al. [21] considered the identification of outliers, the removal of incomplete entries, and the treatment of anomalous data as part of the preprocessing of renewable energy generation datasets. These procedures contribute to improving the consistency of the data and reducing the potential influence of anomalous observations on model training.

Another relevant aspect at this stage is the handling of missing data. The reviewed studies employ different imputation strategies depending on the characteristics of the dataset and the forecasting context. Among these approaches, temporal interpolation is used to estimate missing values from available observations within the time series. Sadeghian Broujeny et al. [22], for instance, employed interpolation techniques to impute missing values in an energy consumption dataset. This type of approach helps preserve the continuity of the time series while maintaining the temporal structure required for subsequent predictive modeling.

Overall, these preprocessing strategies contribute to improving the quality and consistency of the datasets used for energy forecasting. Their application is particularly relevant in energy-related time series, where incomplete or anomalous observations may affect the learning process and, consequently, the predictive performance of the models.

3. Transformation

The transformation techniques identified allow for the conversion of preprocessed data into representations better suited for machine learning algorithms, thereby facilitating the identification of relevant patterns in consumption time series.

Based on the systematic literature review, four types of transformations were identified that appear most frequently in the analyzed studies: normalization, aggregation, feature selection, and feature construction. Table I presents a summary of the studies that apply each of these techniques within the data preparation process.

TABLE I.
DATA TRANSFORMATION TECHNIQUES USED IN SLR STUDIES

Normalization	Aggregation	Feature Selection	Feature Construction
[23], [24], [8], [10], [25], [26], [27], [28], [6], [29], [11], [30], [31], [13], [12], [22]	[23], [24], [7], [9], [10], [26], [11], [32], [12]	[24], [8], [9], [26], [6], [20], [30], [33], [12], [21], [22]	[23], [24], [8], [7], [25], [34], [35], [36], [33], [12], [14]

Source: Author's own work

4. Data Mining

Within the scope of the literature review conducted for the first objective of this research, various artificial intelligence techniques applied to electricity forecasting were identified, many of which belong to the fields of machine learning and deep learning. These techniques have proven to be highly effective in complex contexts involving large volumes of data, such as energy systems, where it is necessary to model nonlinear relationships and temporal dependencies among multiple variables.

Figure 2 presents a summary of the main data mining techniques identified in the studies analyzed during the literature review.

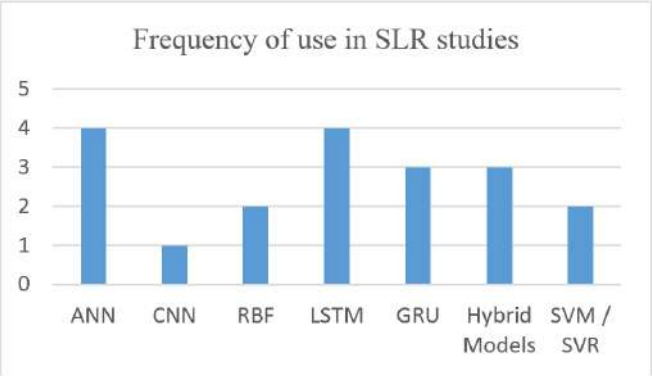


Fig 2. Techniques/models used for energy consumption forecasting. Source: Compiled by the author.

ANNs are among the most widely used techniques in the literature due to their ability to learn complex patterns from data. Inspired by the functioning of the human brain, these networks consist of layers of interconnected neurons that process information and adjust their weights during the learning process to improve the model's performance. Within this category are various architectures used in energy forecasting, such as multilayer perceptron (MLP) networks, convolutional neural networks (CNNs), radial basis function (RBF) networks, and recurrent networks like LSTM and GRU, which are particularly well-suited for time series analysis as they allow for the capture of long-term dependencies in the data.

Furthermore, the literature also highlights a growing use of hybrid models, which combine two or more machine learning techniques to leverage the strengths of each approach. In the context of energy forecasting, these models typically integrate statistical methods with artificial intelligence algorithms or combine different neural architectures. Examples of this include combinations such as ARIMA-ANN or CNN-LSTM, as well as models that integrate neural networks with fuzzy logic or metaheuristic optimization. These types of approaches have been shown to improve prediction accuracy by capturing both linear and nonlinear patterns present in the data.

Finally, support vector machines (SVMs) represent another technique widely used in the reviewed studies. This supervised learning method seeks to find an optimal hyperplane that allows modeling relationships between variables in high-dimensional spaces. By using kernel functions, SVMs can address nonlinear problems, making them suitable for predicting electricity consumption or generation when multiple influencing variables exist, such as weather conditions or demand patterns. However, some studies note that their performance may be limited when working with very large datasets or

those with high temporal frequency.

5. Evaluation and Interpretation

The SLR shows that studies on electricity forecasting use a diverse set of metrics to evaluate the accuracy and robustness of the developed models. The selection of these metrics is essential for evaluating model performance and comparing different forecasting approaches.

Among the most commonly used metrics are the coefficient of determination (R^2), which measures the model's explanatory power, and the root mean square error (RMSE), which quantifies the average magnitude of the error between observed and predicted values, penalizing larger errors more heavily.

Furthermore, the reviewed studies show a trend toward the combined use of metrics based on absolute and relative error, such as the mean absolute error (MAE) and the mean absolute percentage error (MAPE), which allows for a more balanced evaluation of model performance, since while some metrics reflect the magnitude of the error in absolute terms, others allow the error to be interpreted in relation to the actual values of the analyzed variable.

C. Construction Phase

In this phase, the energy consumption prediction model was built by systematically following the stages of the KDD process. This process ranged from data collection and selection to the development, training, and evaluation of the final prediction model, thereby ensuring a structured approach to information processing and the generation of knowledge from the available data.

1. Data Selection

The experimental dataset used to build the model comes from the project "Desarrollo de un modelo transaccional de energía no convencional de múltiples agentes para el Departamento de Nariño (MTE)" [16], which implemented a metering system across five interconnected nodes of a peer-to-peer (P2P) power grid: the Universidad de Nariño (UDENAR), the Universidad Cooperativa de Colombia (UCC), Universidad Mariana (UNIMAR), Universidad CESMAG (UCESMAG), and the Hospital Universitario Departamental de Nariño (HUDN).

Each node is equipped with smart meters, inverters, and weather stations connected to a central monitoring platform that records data in real time. The resulting dataset consists of time series that include electrical variables (active power, reactive power, power factor, current, and voltage), temporal variables (date and time), and meteorological variables—such as temperature, solar irradiance, relative humidity, and wind speed—which influence energy consumption and generation patterns. Table II presents the details of the data collected at each node of the analyzed P2P network.

TABLE II.
BREAKDOWN OF COLLECTED DATA.

Point	Device	Measurement period	Number of records
UDENAR	Meter	28/01/25 - 31/05/25	87941
	Investor		87916
	Station		124337
UCC	Meter	28/02/25 – 31/05/25	66550
	Investor		36196
	Station		43352
UNIMAR	Meter	05/02/25 – 31/05/25	80657
	Investor		43678
	Station		76579
HUDN	Meter	04/04/25 – 31/05/25	41048
	Investor		22407
	Station		15693
UCESMAG	Meter	27/02/25 – 31/05/25	66699
	Investor		36110
	Station		61760

Source: Compiled by the author

2. Preprocessing

Preprocessing was performed on a node-by-node basis to ensure data quality, temporal consistency, and unit homogeneity prior to energy consumption modeling. Three main procedures were applied to all nodes: i) conversion of the power recorded by the inverters from W to kW (acPower and dcPower), ii) temporal synchronization between the different data sources using merge_asof with a tolerance of 120 seconds, using the electricity meter as a reference, and iii) handling of missing values, striving to maintain the continuity of the consumption series and exogenous variables.

At the UDENAR node, no missing values were identified in the individual datasets; however, after integration, missing values appeared in some meteorological variables due to station discontinuities, which were handled through temporal interpolation.

At UCC and UNIMAR, the original datasets also did not contain missing values; however, interruptions were identified at the weather stations and the absence of nighttime records at the inverters, due to the lack of generation during those hours. In these cases, temporal interpolation was applied to the meteorological variables, and a zero value was assigned to the inverter variables during the night.

For HUDN, the meter-maintained continuity in the

records, while the weather station recorded longer periods without data and the inverter again presented nighttime gaps. The same criteria were applied: temporal interpolation of the meteorological variables and assignment of zero to the inverter during the night. At UCESMAG, where the meter and weather station records were mostly continuous, only the nighttime gaps in the inverter data were corrected using the same procedure.

Finally, the data from the five nodes were integrated into a single dataset through temporal concatenation by timestamp, incorporating a node identifier. The resulting dataset contains 364,286 records and 81 features, preserving the temporal heterogeneity across nodes and maintaining the consumption series as the primary reference for modeling. These preprocessing decisions ensured the continuity and consistency of the data, avoiding over-imputation biases and providing a suitable foundation for subsequent stages of data transformation and predictive modeling.

3. Transformation

In this stage, the consolidated dataset was prepared for use in energy consumption modeling. First, the target variable `energy_consumption_kWh` was constructed based on the active power recorded by the smart meters, enabling the representation of energy consumption in two-minute intervals. In addition, temporal variables describing the cyclical nature of demand were incorporated, such as the sine and cosine encodings of the hour and day (hour_sin, hour_cos, day_sin, day_cos) and the weekend indicator (is_weekend).

Subsequently, features were selected using Pearson's correlation and the Recursive Feature Elimination (RFE) method, with a Random Forest-based estimator. This process identified the variables most influential on energy consumption and reduced the set to 34 relevant predictor variables.

The dataset was divided into training (70%), validation (15%), and test (15%) sets, respecting the chronological order of the data to preserve the time series structure. Next, normalization was applied using RobustScaler, trained solely on the training data to prevent information leakage.

Finally, time series were generated using sliding windows of 120 steps (approximately four hours), which serve as input for the deep learning model and allow us to capture the evolution of consumption over time.

4. Data Mining

The data mining phase involves applying artificial intelligence algorithms to identify patterns in energy consumption and generate predictions based on the transformed dataset. In this research, two deep learning architectures were implemented and compared: a hybrid LSTM–CNN network and a multi-site Transformer model, both aimed at predicting electricity consumption in the P2P network of institutions in the municipality of Pasto.

• Hybrid LSTM–CNN Model

The LSTM–CNN model was designed to combine two complementary capabilities: LSTM networks, which allow for the capture of long-term temporal dependencies, and CNNs, which facilitate the identification of local patterns in consumption time series.

The network receives as input sequences of 120-time steps (approximately 4 hours) with 34 variables per time point, corresponding to electrical, meteorological, and temporal measurements. Additionally, a node embedding is incorporated, allowing the model to recognize the specific characteristics of each site in the network.

The architecture is organized into two branches: a recurrent LSTM branch, which learns the temporal dynamics of consumption, and a convolutional Conv1D branch, which identifies local variations in the time series. Subsequently, the outputs of both branches are combined into a vector that is processed through a dense layer with ReLU activation and a dropout mechanism to reduce overfitting. Finally, a linear output layer predicts energy consumption in kWh.

• Transformer Model

As a second approach, a Transformer model was implemented, adapted for the analysis of multi-site time series. This model allows capturing relationships between different time points using self-attention mechanisms, without relying on recurrent structures.

The Transformer receives the same input structure as the previous model: sequences of 120 steps with the 34 selected variables and the node embedding. Its architecture is based on an Encoder block, which includes layer normalization, a Multi-Head Attention mechanism to learn dependencies between different points in the series, and a feed-forward network with ReLU activation and dropout to improve generalization ability.

Finally, the representation generated by the encoder is transformed into a dense vector that, through a linear output layer, produces the energy consumption prediction in kWh.

Both models were trained and evaluated under the same experimental conditions to compare their performance in predicting consumption within the analyzed power grid.

5. Evaluation and Interpretation

In this phase, we analyzed the predictive performance

and computational efficiency of the proposed models (LSTM–CNN and Transformer) when applied to predicting energy consumption (kWh) in the P2P network.

• Model Performance Evaluation

For the quantitative evaluation, the metrics R^2, MAE, RMSE, MAPE, adjusted MAPE, and MaxAE were used, which allow for the analysis of the model's fit, the magnitude of the error, and the stability of the predictions. The results obtained are presented in Table III.

TABLE III.
EVALUATION METRICS RESULTS

Metric\ Model	LSTM-CNN	Transformer
R^2	0.8883	0.8696
MAE	0.0075	0.00956
RMSE	0.0156	0.01684
MAPE	43.17%	47.94%
Adjusted MAPE	11.27%	11.96%
MaxAE	0.2368	0.24623

Source: Compiled by the author

Overall, both models performed well, with R^2 values exceeding 0.86, indicating a good fit between the actual and predicted values. The LSTM–CNN model had the lowest average errors (MAE = 0.0075 and RMSE = 0.0156), suggesting greater accuracy in estimating energy consumption. For its part, the Transformer model achieved similar results (MAE = 0.0096 and RMSE = 0.0168), maintaining stable performance across the network's various nodes.

Analysis of relative errors showed that the traditional MAPE tends to overestimate the error when actual values are close to zero, which is common during periods of low demand. To reduce this effect, an adjusted MAPE was calculated by excluding observations below 0.1 kWh. With this correction, both models recorded relative errors close to 11%, reflecting adequate performance for energy forecasting applications. Regarding the maximum absolute error (MaxAE), the values were similar in both models (≈0.24 kWh).

Overall, the results indicate that both approaches adequately capture electricity consumption patterns. The LSTM–CNN model stands out for its greater accuracy and stability in predictions, while the Transformer model shows comparable performance with a greater ability to adapt to variability across sites.

• Performance Analysis by Node

Following the overall performance analysis, we proceed to examine the results obtained for each node within the P2P network. This disaggregated analysis allows us to identify how operational conditions, the continuity of records, and specific consumption patterns at each institution influence the accuracy of the models. The results are presented in Tables IV and V.

TABLE IV.
METRICS RESULTS BY NODE - LSTMCNN MODEL.

Metric\ Model	UDENAR	UCC	UNIMAR	HUDN	UCESMAG
R^2	0.8419	0.8436	0.7368	0.4484	0.9660
MAE	0.0095	0.0094	0.0080	0.0075	0.0017
RMSE	0.0198	0.0197	0.0139	0.0099	0.0021
MAPE	63.00%	61.95%	32.66%	7.11%	26.88%
Adjusted MAPE	13.15%	12.79%	24.61%	8.53%	NaN
MaxAE	0.1607	0.1616	0.2368	0.0504	0.0127

TABLE V.
METRICS RESULTS BY NODE - TRANSFORMER MODEL.

Metric\ Model	UDENAR	UCC	UNIMAR	HUDN	UCESMAG
R^2	0.8206	0.8200	0.6789	0.4168	0.8131
MAE	0.0125	0.0127	0.0091	0.0077	0.0033
RMSE	0.0211	0.0211	0.0153	0.0102	0.0048
MAPE	64.23%	67.57%	35.52%	7.42%	40.23%
Adjusted MAPE	19.29%	17.62%	24.56%	8.05%	NaN
MaxAE	0.1654	0.1629	0.2462	0.0561	0.0287

Source: Author's own work

a) UDENAR Node

At this node, both models achieved R^2 values greater than 0.82, indicating a strong correlation between the predictions and the actual values. Absolute errors remained low (MAE ≈ 0.009–0.013 kWh; RMSE ≈ 0.020–0.021 kWh), while the maximum error (MaxAE) was around 0.16 kWh. After applying the adjusted MAPE, the values ranged between 13% and 19%, reflecting stable behavior consistent with institutional consumption patterns.

b) UCC Node

Performance at this node was consistent with the results observed at UDENAR. The R^2 values ranged between 0.82 and 0.84, with similar absolute errors (MAE ≈ 0.009–0.013 kWh; RMSE ≈ 0.020–0.021 kWh). The adjusted MAPE ranged between 13% and 18%, indicating stable performance and adequate predictive capability in an environment with high operational regularity.

c) UNIMAR Node

At this node, a slight decrease in the final explanatory power was observed among the models, with R^2 values between 0.68 and 0.74. However, absolute errors remained low (MAE ≈ 0.008–0.009 kWh; RMSE ≈ 0.014–0.015 kWh). The adjusted MAPE was around 24%, indicating acceptable performance, influenced by greater

variability in consumption patterns.

d) HUDN Node

The node corresponding to the Nariño Departmental University Hospital presented the lowest R^2 values (≈ 0.41–0.45), which is mainly explained by the reduced training time window (two months) and by the highly variable nature of hospital consumption, associated with non-stationary operational dynamics. Despite this, the absolute errors were low (MAE ≈ 0.0075–0.0077 kWh; RMSE ≈ 0.010 kWh), and the adjusted MAPE remained close to 8%, demonstrating good accuracy in absolute terms.

e) UCESMAG Node

This node showed the best performance in terms of absolute error, especially for the LSTM–CNN model (MAE = 0.0017 kWh; RMSE = 0.0021 kWh), with an R^2 value of 0.966. Meanwhile, the Transformer model achieved an R^2 of 0.813, remaining within a satisfactory range.

In this case, the adjusted MAPE could not be calculated (NaN) for both models. This is because, after applying the exclusion threshold (< 0.1 kWh), there were no valid observations left for calculating the metric. This behavior is associated with the low magnitude of energy consumption at this node, where most records are close to zero. Therefore, the absence of this indicator does not reflect a limitation of the model, but rather a condition inherent to the data, characterized by high stability and low variability.

In general terms, the analysis by node shows that those sites with greater regularity in their records and lower variability in consumption patterns (UDENAR, UCC, and UCESMAG) achieve the highest levels of performance. In contrast, in scenarios characterized by greater operational dynamics or lower availability of historical data (UNIMAR and HUDN), a slight reduction in the models' explanatory power is observed, although absolute errors remain within narrow ranges.

From a comparative perspective, the LSTM–CNN model shows a slight advantage in accuracy and fit at the local level. For its part, the Transformer model maintains stable performance and greater adaptability to variability across nodes. Taken together, these findings demonstrate that model performance depends largely on the specific characteristics of each measurement point and highlight the complementarity of both architectures as viable alternatives for predicting energy consumption in distributed electrical environments.

• Performance Evaluation

In addition, the computational efficiency of the models was evaluated in two execution environments: a high-performance server from the project and a general-purpose personal computer. The server is equipped with Intel® Xeon® Platinum 8452Y processors, 512 GB of RAM, and NVIDIA L4 GPUs optimized for parallel computing, while the personal computer features an Intel® Core i7 processor, 8 GB of RAM, and no dedicated GPU.

The computational results, presented in Table VI, show that on the server, both models run efficiently, with low CPU usage and adequate GPU utilization. In this environment, the Transformer model demonstrated faster training, requiring less total time than the LSTM–CNN model.

TABLE VI.

COMPUTATIONAL PERFORMANCE EVALUATION

Aspect evaluated	LSTM-CNN		Transformer	
	Server	Personal computer	Server	Personal computer
Training time by season	1.6 min	26.65 min	30 s	18.9 min
Total training time	38 min	639.6 min	12.5 min	474.16 min
RAM usage	5.5%	90%	6.8%	81%
GPU/CPU usage	GPU 40% / CPU 0.1%	CPU 65%	GPU 58.5% / CPU 0.3 %	CPU 70%

Source: Compiled by the author

In contrast, when run on a personal computer, training times increased significantly due to the absence of a GPU. In this scenario, the Transformer once again demonstrated greater efficiency, completing the process in less time than the LSTM–CNN.

In summary, the results show that the computing environment significantly influences training performance. While the LSTM–CNN model offers a slight advantage in accuracy, the Transformer stands out for its efficiency and scalability, especially when specialized infrastructure with GPUs is available to process large volumes of energy data.

D) Validation Phase

Following the variations observed during the evaluation phase, primarily associated with the specific characteristics of each institution in the P2P network, we proceeded to test the models' generalization ability using unseen data. To this end, a temporal validation was implemented, using the data collected during the month of June as an independent test set. This approach ensured that the information used in the validation was completely new to the models, replicating the same KDD

pipeline used during training.

The results are analyzed at three complementary levels: (i) global validation results, which integrate the behavior of all nodes in the P2P network; (ii) validation results by node, which allow for the identification of variations in model performance based on the specific characteristics of each institution; and (iii) analysis of computational performance during the inference stage, to evaluate the efficiency of the models in different execution environments.

1) Global Validation Results

Table VII presents the results of the global validation of the models. Overall, both approaches maintained solid performance. The LSTM–CNN model achieved an R^2 = 0.9146, with low absolute errors (MAE = 0.0072 kWh and RMSE = 0.0136 kWh). Meanwhile, the Transformer achieved an R^2 of 0.8991, with similarly low error values (MAE = 0.0088 kWh and RMSE = 0.0148 kWh).

TABLE VII.

OVERALL VALIDATION RESULTS.

Metric\ Model	LSTM-CNN	Transformer
R2	0.9146	0.8991
MAE	0.0072	0.0088
RMSE	0.0136	0.0148
MAPE	52.80%	73.01%
Adjusted MAPE	10.78%	12.85%
MaxAE	0.17707	0.17578

Source: Compiled by the author

These results show that both architectures are able to accurately estimate actual consumption values, with average errors remaining below 0.01 kWh. Regarding relative error, the classic MAPE showed high values due to records with very low consumption. To reduce this effect, the adjusted MAPE was used, which excludes observations below 0.1 kWh. With this correction, relative errors were reduced to 10.78% for the LSTM–CNN model and 12.85% for the Transformer, reflecting a more representative performance.

Finally, the maximum absolute error (MaxAE) was less than 0.18 kWh for both models, indicating that even the predictions furthest from the actual value remain within an acceptable margin.

The graphical representation of the validation results (Figures 3 and 4) shows the comparison between the actual and predicted values for the month of June 2025. The superimposition of both series—actual consumption and estimated consumption shows a high degree of correspondence between the curves, indicating that the models successfully reproduce the temporal dynamics of energy consumption, capturing both daily variations and demand peaks during the analyzed period.

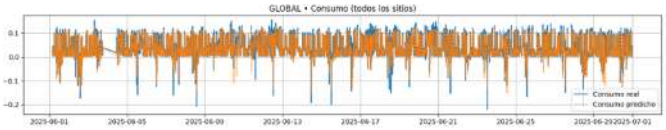


Fig 3. Energy consumption forecasts - LSTMCNN model. Source: Author's own work

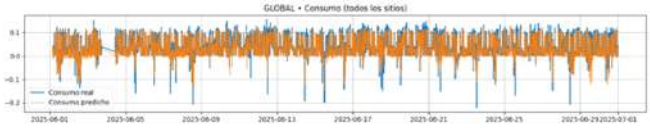


Fig 4. Energy consumption forecasts - Transformer model. Source: Author's own work

2) Computational Performance Results in the Validation Phase

In addition to predictive performance, the computational performance of the models was evaluated during the validation phase to analyze their efficiency and feasibility in different processing environments. This analysis is important given that, in real-world applications, computational capacity can vary across institutions. The tests were conducted in the same two scenarios defined previously: a high-performance server and a personal computer. The results of the inference process are presented in Table VIII.

TABLE VIII.

COMPUTATIONAL PERFORMANCE RESULTS FOR INFERENCE.

Aspect evaluated	Server		Personal computer	
	LSTM-CNN	Transformer	LSTM-CNN	Transformer
Inference time	0.60 min	0.73 min	2.9 min	4.8 min
RAM usage	2.3%	3.5%	50%	20%
CPU usage	1%	0.28%	100%	100%
GPU usage	8.6%	6.5%	N/A	N/A

Source: Author's own work

On the server, both models performed efficiently, with low inference times and minimal resource usage. The LSTM–CNN model completed validation in 0.60 minutes, while the Transformer model did so in 0.73 minutes. Memory consumption was minimal (between 2.3% and 3.5%), and CPU load remained low, as processing relied primarily on the GPU.

In contrast, on the personal computer, where execution was performed solely on the CPU, runtimes

increased significantly: 2.9 minutes for LSTM–CNN and 4.8 minutes for Transformer. Furthermore, CPU utilization reached 100% in both cases, accompanied by higher memory consumption, indicating a greater computational load in environments without GPU acceleration.

Overall, these results show that model performance depends largely on the execution environment. While both approaches are efficient and suitable for near-real-time applications on specialized infrastructure, execution remains possible on devices with limited resources, albeit with longer processing times

3) Hypothesis Testing

To determine whether the differences observed between the LSTM–CNN and Transformer models are statistically significant in predicting energy consumption, the nonparametric Wilcoxon signed-rank test was applied.

This method allows for a comparison of the performance of both models under the same input conditions, making it appropriate in this case due to the small sample size (five nodes) and the potential non-normality of the data. For the analysis, the metrics MAE, RMSE, and adjusted MAPE were considered, both in the evaluation phase and in the temporal validation.

The results are presented in Table IX, where the errors of both models are compared for each of the metrics analyzed, allowing us to determine whether the differences found are statistically significant.

TABLE IX.
RESULTS OF THE WILCOXON STATISTICAL TEST.

Phase	Metric	n_ pairs	Statistician	p-value	Conclusion (e±=0.05)
Evaluation	MAE	5	0.0	0.0625	SED
Evaluation	RMSE	5	0.0	0.0625	SED
Evaluation	MAPEaj	4	3.0	0.625	SED
Validation	MAE	5	0.0	0.0625	SED
Validation	RMSE	5	0.0	0.0625	SED
Validation	MAPEaj	4	4.0	0.875	SED

Source: Author's own work

Where:

- Phase: indicates the stage of the process at which the statistical test was applied.
- Metric: corresponds to the error metric used to compare the models.
- n_pairs: number of pairs of observations used in the test.
- Statistic: value calculated by the Wilcoxon test.
- p_value: probability associated with the statistic; if $p < 0.05$, the null hypothesis is rejected, whereas if $p \geq 0.05$, there is insufficient evidence to reject it.
- Conclusion: interpretation of the test at the 5% significance level. The abbreviation SED stands for "No Evidence of Difference."

The results show that, in all cases, the p-values were greater than 0.05, so no statistically significant differences were found between the models. Although the LSTM–CNN model exhibited slightly lower average errors in several scenarios, these differences are not sufficient to claim better performance from a statistical standpoint.

This result suggests that both models perform equivalently in estimating energy consumption within the P2P network. Consequently, the choice between one or the other can be based on practical criteria, such as the availability of computational resources, training efficiency, or ease of implementation, rather than on differences in predictive accuracy.

E) Discussion of the Results

The results obtained demonstrate that the LSTM–CNN and Transformer architectures are capable of adequately modeling energy consumption behavior in a real P2P network, achieving determination coefficients greater than 0.90 during global validation and absolute errors less than 0.01 kWh. These results demonstrate the effectiveness of deep learning approaches in capturing the temporal dependencies present in multivariate energy consumption time series.

One of the main contributions of this research lies in the comprehensive application of the KDD process to build predictive models in a P2P network composed of multiple institutional nodes. Unlike most studies reported in the literature, which focus on individual buildings, smart homes, or conventional microgrids, this work incorporates electrical, meteorological, and temporal information from different institutions within a real-world distributed energy exchange environment. Additionally, a node-by-node disaggregated analysis and temporal validation using completely unseen data are performed, allowing for a more rigorous evaluation of the models' generalization ability.

The results obtained are consistent with previous research reporting better performance of deep learning-based architectures for energy consumption forecasting. Studies such as [9] found that RNN- and LSTM-based models outperform traditional techniques such as ANN, SVM, and XGBoost in energy consumption scenarios, due to their ability to capture complex temporal dependencies. Similarly, [14] reported that artificial neural networks achieve better results than other machine lear-

ning approaches for load forecasting in P2P networks. The findings of this research reinforce these conclusions and demonstrate that both LSTM–CNN and Transformer are viable alternatives for this type of application.

From a practical perspective, the results suggest that the developed models can be used as support tools for energy planning and decision-making in P2P networks. The ability to anticipate energy consumption facilitates resource scheduling, the balancing of supply and demand, and the optimization of energy transactions among network participants. Furthermore, the computational performance analysis showed that both models can run in different technological environments, although the use of GPU-based infrastructure significantly reduces processing times.

However, the study has some limitations. First, the research was conducted on a network consisting of five institutional nodes located in the same geographic region; therefore, the results should be generalized to other contexts with caution. Second, some nodes, such as HUDN and UCESMAG, exhibited specific conditions associated with periods of low occupancy or limited data availability, which affected certain evaluation metrics. Finally, although two widely used deep learning architectures were evaluated, future research could incorporate more recent attention-based models, additional hybrid approaches, or federated learning techniques for distributed energy scenarios.

IV. CONCLUSIONS

The results showed that both models successfully capture consumption patterns across the network's various nodes, maintaining stable behavior and demonstrating good generalization ability. The variations observed across institutions are primarily related to the specific characteristics of each environment, such as changes in occupancy, operational dynamics, or specific periods that affect demand. However, further analysis would be required to establish the contribution of each factor to the observed differences.

The results showed that both models successfully capture consumption patterns across the network's various nodes, maintaining stable behavior and demonstrating good generalization ability. The variations observed across institutions are primarily related to the specific characteristics of each environment, such as changes in occupancy, operational dynamics, or specific periods that affect demand.

Regarding computational performance, it was evident that the execution environment has a significant impact on the models' efficiency. The use of GPU-based infrastructure significantly reduced processing times and optimized resource utilization, whereas on systems without graphics acceleration, processing times increased and the load on the CPU and memory was higher. This indicates that the practical implementation of these models must consider both the design and the availability of technological resources.

Finally, the hypothesis testing confirmed that there are no statistically significant differences between the two evaluated architectures. In this regard, the LSTM–CNN and Transformer models can be considered equivalent in terms of accuracy, so the choice between them can be based on criteria such as computational efficiency or scalability. Overall, the results support the use of deep learning techniques as an effective alternative for predicting energy consumption in distributed power grids.

One limitation of this study is the duration of the data collection period used to construct the dataset, which covered five months, from January to May 2025. Although an independent dataset corresponding to June was used for temporal validation, the observation period may not fully capture seasonal variations in electricity consumption and generation or longer-term changes in energy-use patterns. Consequently, the extent to which the results generalize to other temporal conditions remains to be established. Future research should extend the data collection period and evaluate the models under a wider range of operating conditions to further assess their robustness and generalization capability.

REFERENCES

[1].C. J. Khare, H. K. Verma, and V. Khare, Optimal power generation and power flow control using artificial intelligence techniques. 2021. doi: 10.1016/B978-0-12-820004-9.00028-0.

[2].S. Gandomi, M. Barzegar, S. Zolghadri, and M. R. Rahimpour, "The Future of Fossil Fuels Supply and Impacts on Sustainability," in Encyclopedia of Renewable Energy, Sustainability and the Environment, Elsevier, 2024, pp. 121–130. doi: 10.1016/B978-0-323-93940-9.00270-X.

[3].A. Akila, R. C. Bansal, and A. Abo-Khalil, "Advancements in Peer-to-Peer Energy Management Systems – An Overview," in 2024 7th International Conference on Electric Power and Energy Conversion Systems (EPECS), IEEE, Nov. 2024, pp. 167–172. doi: 10.1109/EPECS62845.2024.10805521.

[4].K. L. Yandar, O. Revelo Sánchez, and M. E. Bolaños-González, "Machine learning para la predicción de energía eléctrica: una revisión sistemática de literatura," Ing. y Compet., vol. 26, no. 2, Jul. 2024, doi: 10.25100/iyc.v26i2.13875.

[5].W. Tushar et al., "Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges," Appl. Energy, vol. 282, p. 116131, Jan. 2021, doi: 10.1016/j.apenergy.2020.116131.

[6].A. D. Manchalwar, N. R. Patne, B. V. S. Vardhan, and M. Khedkar, "Peer-to-peer energy trading in a distribution network considering the impact of short-term load forecasting," Electr. Eng., vol. 105, no. 4, pp. 2069–2081, Aug. 2023, doi: 10.1007/s00202-023-01793-8.

[7].J.-S. Chou and D.-S. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," Energy, vol. 165, pp. 709–726, Dec. 2018, doi: 10.1016/j.energy.2018.09.144.

[8].W. Waheed and Q. Xu, "Optimal Short Term Power Load Forecasting Algorithm by Using Improved Artificial Intelligence Technique," in 2020 2nd International Conference on Computer and Information Sciences (ICCIS), IEEE, Oct. 2020, pp. 1–4. doi: 10.1109/ICCIS49240.2020.9257675.

[9].L. Cáceres, J. I. Merino, and N. Díaz-Díaz, "A Computational Intelligence Approach to Predict Energy Demand Using Random Forest in a Cloudera Cluster," Appl. Sci., vol. 11, no. 18, p. 8635, Sep. 2021, doi: 10.3390/app11188635.

[10].O. Timur, K. Zor, Ö. Çelik, A. Teke, and T. İbrikçi, "Application of Statistical and Artificial Intelligence Techniques for Medium-Term Electrical Energy Forecasting: A Case Study for a Regional Hospital," J. Sustain. Dev. Energy, Water Environ. Syst., vol. 8, no. 3, pp. 520–536, Sep. 2020, doi: 10.13044/j.sdewes.d7.0306.

[11].I. Hajjaji, H. El Alami, M. R. El-Fenni, and H. Dahmouni, "Evaluation of Artificial Intelligence Algorithms for Predicting Power Consumption in University Campus Microgrid," in 2021 International Wireless Communications and Mobile Computing (IWCMC), IEEE, Jun. 2021, pp. 2121–2126. doi: 10.1109/IWCMC51323.2021.9498891.

[12].Zuitao Ma, "Neural Networks Used for Time Series Prediction of Power Consumption," University of Oslo, 2019.

[13].G. Kim, G. Lee, S. An, and J. Lee, "Forecasting future electric power consumption in Busan New Port using a deep learning model," Asian J. Shipp. Logist., vol. 39, no. 2, pp. 78–93, Jun. 2023, doi: 10.1016/j.ajsl.2023.04.001.

[14].S. B. Martinez, "Energy management systems for smart homes and local energy communities based on optimization and artificial intelligence techniques," Universidad Politécnica de Cataluña, 2023.

[15].F. Li et al., "Improving the accuracy of multi-step prediction of building energy consumption based on EEMD-PSO-Informer and long-time series," Comput. Electr. Eng., vol. 110, 2023, doi: 10.1016/j.compeleceng.2023.108845.

[16]."PROYECTO SIGP CON CÓDIGO DE REGISTRO 89530, DESARROLLO DE UN MODELO TRANSACCIONAL DE ENERGÍA NO CONVENCIONAL DE MÚLTIPLES AGENTES PARA EL DEPARTAMENTO DE NARIÑO." 2022.

[17].S. Kitchenham, B., & Charters, "Guidelines for performing systematic literature reviews in software engineering," Tech. report, Ver. 2.3 EBSE Tech. Report. EBSE, no. January 2007, pp. 1–57, 2007.

[18].U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., vol. 17, no. 3, pp. 37–53, 1996.

[19].A. Barón and C. Zapata, "Estrategia Metodológica para la Elaboración de Síntesis Conceptuales en Ingeniería de Software: una Aplicación al Caso del Constructo Teórico de Práctica," 4th Int. Conf. Softw. Eng. Res. Innov. CONISOFT, 2016.

[20].A. I. Saleh, A. H. Rabie, and K. M. Abo-Al-Ez, "A data mining based load forecasting strategy for smart electrical grids," Adv. Eng. Informatics, vol. 30, no. 3, pp. 422–448, Aug. 2016, doi: 10.1016/j.aei.2016.05.005.

[21].M. Abdul Baseer, A. Almunif, I. Alsaduni, and N. Tazeen, "Electrical Power Generation Forecasting from Renewable Energy Systems Using Artificial Intelligence Techniques," Energies, vol. 16, no. 18, 2023, doi: 10.3390/en16186414.

[22].R. Sadeghian Broujeny, S. Ben Ayed, and M. Matalah, "Energy Consumption Forecasting in a University Office by Artificial Intelligence Techniques: An Analysis of the Exogenous Data Effect on the Modeling," Energies, vol. 16, no. 10, 2023, doi: 10.3390/en16104065.

[23].G. Grigoraș and B.-C. Neagu, "An Advanced Decision Support Platform in Energy Management to Increase Energy Efficiency for Small and Medium Enterprises," Appl. Sci., vol. 10, no. 10, p. 3505, May 2020, doi: 10.3390/app10103505.

[24].Y. Chen and P. Phelan, "Predicting peak energy demand for an office building using artificial intelligence (ai) approaches," in American Society of Mechanical Engineers, Power Division (Publication) POWER, 2021. doi: 10.1115/POWER2021-64492.

[25].H. R. O. Rocha, I. H. Honorato, R. Fiorotti, W. C. Celeste, L. J. Silvestre, and J. A. L. Silva, "An Artificial Intelligence based scheduling algorithm for demand-side energy management in Smart Homes," Appl. Energy, vol. 282, p. 116145, Jan. 2021, doi: 10.1016/j.apenergy.2020.116145.

[26].T. Vantuch and M. Prilepok, "An ensemble of multi-objective optimized fuzzy regression models for short-term electric load forecasting," in2017 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, Nov. 2017, pp. 1–7. doi: 10.1109/SSCI.2017.8285348.

[27].M. Rizwan and Y. R. Alharbi, "Artificial Intelligence Based Approach for Short Term Load Forecasting for Selected Feeders at Madina, Saudi Arabia," Int. J. Electr. Electron. Eng. Telecommun., vol. 10, no. 5, pp. 300–306, 2021, doi: 10.18178/ijeetc.10.5.300-306.

[28].I. Alsaidan, M. Rizwan, and M. Alaraj, "Solar energy forecasting using intelligent techniques: A step towards sustainable power generating system," J. Intell. Fuzzy Syst., vol. 42, no. 2, pp. 885–896, Jan. 2022, doi: 10.3233/JIFS-189757.

[29].J. R. Gómez Sarduy, J. P. Monteagudo Yanes, M. E. Granado Rodríguez, J. L. Quiñones Ferreira, and Y. M. Torres,

"Determining cement ball mill dosage by artificial intelligence tools aimed at reducing energy consumption and environmental impact," Ing. e Investig., vol. 33, no. 3, pp. 49–54, Sep. 2013, doi: 10.15446/ing.investig.v33n3.41043.

[30].D.-H. Tran, D.-L. Luong, and J.-S. Chou, "Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings," Energy, vol. 191, p. 116552, Jan. 2020, doi: 10.1016/j.energy.2019.116552.

[31].H. Hussein, "An Optimal Design Methodology of Adaptive Neuro-Fuzzy Inference System for Energy Load Forecasting - Hail city case study (Saudi Arabia)," in Proceedings of the Fourth International Conference on Engineering & MIS 2018, New York, NY, USA: ACM, Jun. 2018, pp. 1–7. doi: 10.1145/3234698.3234765.

[32].J. Olatomiwa Lanre, "Optimal planning and design of hybrid renewable energy system for rural healthcare facilities / Olatomiwa Lanre Joseph.," 2016.

[33].S. Fathi and R. Srinivasan, "Climate Change Impacts on Campus Buildings Energy Use: An AI-Based Scenario Analysis," in Proceedings of the 1st ACM International Workshop on Urban Building Energy Sensing, Controls, Big Data Analysis, and Visualization, New York, NY, USA: ACM, Nov. 2019, pp. 112–119. doi: 10.1145/3363459.3363540.

[34].L. Raju, P. V K, S. A. A. S, V. V, and B. V, "Building Energy Management and Conservation using Internet of Things," in 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), IEEE, Apr. 2022, pp. 970–974. doi: 10.1109/ICSCDS53736.2022.9760907.

[35].K. W. Mui, L. T. Wong, M. K. Satheesan, and A. Balachandran, "A Hybrid Simulation Model to Predict the Cooling Energy Consumption for Residential Housing in Hong Kong," Energies, vol. 14, no. 16, p. 4850, Aug. 2021, doi: 10.3390/en14164850.

[36].Y. Sasaki et al., "A Robust Economic Load Dispatch in Community Microgrid Considering Incentive-based Demand Response," IFAC-PapersOnLine, vol. 55, no. 9, pp. 389–394, 2022, doi: 10.1016/j.ifacol.2022.07.068.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

Control predictivo adaptativo para regulación del estrés mediante estímulos musicales basados en señales HRV

Adaptive Predictive Control for Stress Regulation Using HRV-Based Musical Stimuli

Jonathan Cabrera [1], Alvaro A. Jiménez-Ocaña [2], Andrés Pantoja [3] *,

[1] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. jdcabrera21A@udenar.edu.co
[2] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. andresjimenez@udenar.edu.co
[3] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. ad_pantoja@udenar.edu.co

* Autor de correspondencia: ad_pantoja@udenar.edu.co

Citación: J. Cabrera, A. A. Jiménez-Ocaña y A. Pantoja, "Control predictivo adaptativo para regulación del estrés mediante estímulos musicales basados en señales HRV," Revista de Ingeniería, vol. 4, no. 1, pp. 61-75, Nov. 2026.

Resumen—Este artículo aborda el problema de la regulación personalizada del estrés mediante retroalimentación fisiológica, centrándose en las limitaciones de las intervenciones musicales en lazo abierto ante la alta variabilidad inter e intra-sujeto. Se propone un esquema de control adaptativo basado en señales de variabilidad del ritmo cardíaco (HRV), en el cual estímulos musicales ajustables se regulan mediante estimación del estrés basada en aprendizaje automático y control predictivo adaptativo. Las señales fisiológicas se procesan para estimar una probabilidad continua de relajación, utilizada como variable controlada. La dinámica del sujeto se modela mediante identificación ARX recursiva, lo que permite la adaptación en línea ante respuestas psicofisiológicas variables en el tiempo. Se implementa una estrategia de Control Predictivo Basado en Modelo (MPC) adaptativo para regular el estímulo musical, mientras que un esquema de Control por Búsqueda de Extremos (ESC) ajusta dinámicamente la referencia de relajación. El enfoque propuesto se evalúa mediante simulaciones numéricas y pruebas experimentales con sujetos humanos bajo condiciones de inducción de estrés. Los resultados experimentales muestran un incremento promedio del 11.8% en el Tiempo en Rango (TIR) durante el control activo, en comparación con las condiciones de música en lazo abierto y sin música, demostrando el potencial del control adaptativo en lazo cerrado para la modulación personalizada del estrés.

Palabras claves—Control adaptativo, Control basado en HRV, Control por búsqueda de extremos (ESC), Control predictivo basado en modelo (MPC), Regulación de estrés.

Abstract—This article addresses the problem of personalized stress regulation through physiological feedback, focusing on the limitations of open-loop musical interventions in the presence of high inter- and intra-subject variability. An adaptive control scheme based on heart rate variability (HRV) signals is proposed, in which adjustable musical stimuli are regulated through machine learning-based stress estimation and adaptive predictive control. Physiological signals are processed to estimate a continuous probability of relaxation, used as the controlled variable. Subject dynamics are modeled using recursive ARX identification, allowing online adaptation to time-varying psychophysiological responses. An adaptive Model Predictive Control (MPC) strategy is implemented to regulate the musical stimulus, while an Extremum Seeking Control (ESC) scheme dynamically adjusts the relaxation reference. The proposed approach is evaluated through numerical simulations and experimental testing with human subjects under stress-induction conditions. Experimental results show an average increase of 11.8% in Time in Range (TIR) during active control compared to open-loop music and no-music conditions, demonstrating the potential of closed-loop adaptive control for personalized stress modulation.

Keywords—Adaptive control, Extremum seeking control (ESC), Heart rate variability (HRV)-based control, Model predictive control (MPC), Stress regulation.

I. INTRODUCCIÓN

La variabilidad del ritmo cardíaco (Heart Rate Variability, HRV) es un indicador fisiológico ampliamente utilizado para evaluar el equilibrio del sistema nervioso autónomo y la capacidad adaptativa del organismo frente a estímulos de estrés físico y emocional [1]. Una reducción en la HRV se ha asociado con niveles elevados de estrés, dificultades en la regulación emocional y una menor capacidad de adaptación fisiológica, lo que ha motivado su uso en aplicaciones de monitoreo y regulación del estrés.

Entre las estrategias no invasivas más estudiadas se encuentran las técnicas de biofeedback basadas en HRV, las cuales permiten a los individuos modular sus respuestas fisiológicas mediante retroalimentación visual o auditiva. Este tipo de enfoques ha demostrado ser efectivo para incrementar la variabilidad cardíaca y mejorar la regulación fisiológica [2]. No obstante, la mayoría de estos sistemas operan bajo esquemas de lazo abierto o semi–abierto, donde los parámetros de intervención permanecen fijos o dependen de reglas heurísticas, limitando su capacidad de adaptación frente a la variabilidad inter- e intra-sujeto.

En este contexto, la música ha emergido como un estímulo sensorial con alto potencial para la modulación de respuestas fisiológicas asociadas al estrés. Estudios previos han evidenciado que la música puede reducir marcadores fisiológicos y subjetivos de estrés, tales como la ansiedad percibida, la presión arterial y la frecuencia cardíaca [3]–[5]. Asimismo, intervenciones musicales han sido empleadas en poblaciones clínicas, mostrando efectos positivos sobre métricas de HRV y bienestar general [6]. Sin embargo, cuando los estímulos musicales no se adaptan dinámicamente a la respuesta fisiológica del sujeto, su efectividad puede ser limitada o altamente dependiente de preferencias individuales.

Más recientemente, se han propuesto enfoques basados en inteligencia artificial y aprendizaje automático para la estimación del estrés a partir de señales fisiológicas, incluyendo HRV, frecuencia cardíaca y otras variables biomédicas. Estos métodos han mostrado buenos resultados en tareas de clasificación y estimación del estado de estrés [9], [10]. No obstante, la mayoría de estos trabajos se centran en la detección del estado fisiológico y no integran explícitamente un esquema de control que permita regular activamente dicho estado mediante estímulos externos.

Desde una perspectiva de ingeniería de control, la regulación del estrés plantea retos significativos debido a la naturaleza no lineal, variante en el tiempo y altamente incierta de los sistemas psicofisiológicos. Recientes revisiones del estado del arte destacan que la mayoría de las estrategias existentes se basan en esquemas de lazo abierto o adaptaciones limitadas, evidenciando la necesidad de enfoques de control adaptativo en lazo cerrado que integren estimación fisiológica y control dinámico [7].

El presente trabajo se enmarca dentro de la línea de investigación desarrollada en [7], donde se presenta una revisión exhaustiva de técnicas de monitoreo y regulación del estrés basadas en señales fisiológicas, incluyendo enfoques de biofeedback y métodos inteligentes. En particular, este estudio aprovecha recursos generados en el marco de [8], incluyendo los datos fisiológicos adquiridos durante los experimentos con sujetos, el sistema de adquisición de señales biomédicas y los códigos base para la extracción de índices y características de la variabilidad del ritmo cardíaco (HRV). Sobre esta infraestructura previamente desarrollada, el presente trabajo propone una contribución original al integrar un esquema de control adaptativo en lazo cerrado, combinando estimación basada en aprendizaje automático, identificación recursiva y control predictivo, orientado específicamente a la regulación dinámica del estrés mediante estímulos musicales personalizados.

En este trabajo se propone un esquema de control adaptativo en lazo cerrado para la regulación del estrés basado en señales de HRV, donde estímulos musicales ajustables actúan como entradas de control. A partir del procesamiento de señales fisiológicas, se estima una probabilidad continua de relajación que se utiliza como variable controlada. La dinámica del sistema se aborda mediante identificación ARX recursiva y se regula a través de una estrategia de Control Predictivo Basado en Modelo (MPC) adaptativo, complementada con un esquema de Control de Búsqueda Extrema (ESC) para el ajuste dinámico de la referencia de relajación.

II. METODOLOGÍA DEL SISTEMA DE CONTROL

El sistema se estructura en tres componentes principales: el módulo de estimación de estado (XGBoost), el módulo de identificación de planta (ARX Recursivo) y el módulo de control (MPC Adaptativo con ESC).

A. Experimento Preliminar y Adquisición de Datos

Los datos empleados en la fase preliminar de este trabajo fueron adquiridos durante un experimento con sujetos descrito en [8]. En dicho experimento se aplicaron estímulos musicales mientras se registraban señales fisiológicas, con el fin de caracterizar la respuesta psicofisiológica asociada al estrés y la relajación. La adquisición de datos se llevó a cabo mediante un sistema de electrocardiografía (ECG) de una sola derivación, utilizando una tarjeta de adquisición, lo que permitió

obtener señales en tiempo real de forma no invasiva.

Durante el experimento, el estímulo musical empleado fue la pieza *Canon in D Major* del compositor Johann Pachelbel, seleccionada con base en estudios previos que reportan efectos positivos de la música clásica lenta y estructurada sobre la reducción del estrés y la modulación de respuestas fisiológicas asociadas a la relajación [3]–[5]. De manera simultánea, se registraron señales de electrocardiograma (ECG), a partir de las cuales se extrajeron los intervalos R–R para el análisis de la variabilidad del ritmo cardíaco (HRV).

Siguiendo la metodología descrita en [11], se calcularon índices fisiológicos en distintos dominios. En el dominio del tiempo se extrajeron métricas como la frecuencia cardíaca media (mHR), el porcentaje de intervalos NN mayores a 50 ms (pNN50) y el valor RMSSD. En el dominio de la frecuencia se calculó la potencia relativa en la banda respiratoria (Pr,r). Asimismo, se obtuvieron métricas no lineales derivadas del diagrama de Poincaré, incluyendo SD1, SD2 y la relación SD1/SD2.

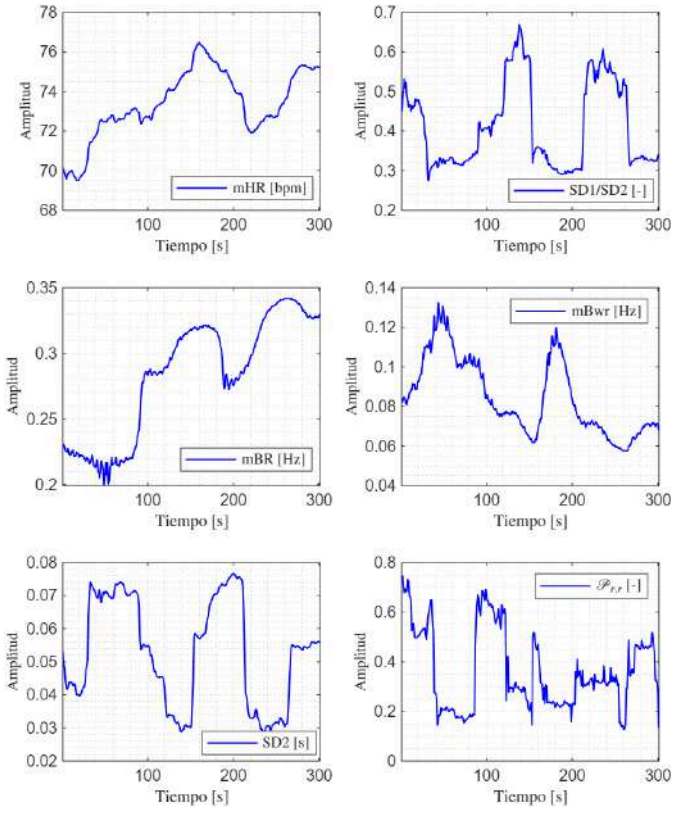


Fig. 1. Evolución temporal de señales fisiológicas y características HRV obtenidas durante el experimento preliminar con estímulo musical.

El protocolo experimental se estructuró en distintas etapas claramente definidas, que incluyen un periodo inicial de reposo, una fase de inducción de estrés y una fase de intervención musical. Cada etapa tuvo una duración preestablecida, permitiendo observar la evolución temporal de los índices fisiológicos ante diferentes condiciones. Durante la fase de intervención, la pieza musical fue reproducida bajo condiciones controladas y posteriormente modulada en amplitud durante los experimentos de control.

La Fig. 1 presenta un conjunto representativo de señales fisiológicas y características HRV extraídas durante la escucha del estímulo musical para un paciente. La evolución temporal de estas señales evidencia variaciones fisiológicas coherentes con las distintas etapas del protocolo experimental y con la intervención musical aplicada, sirviendo como base para la estimación posterior del estado continuo de estrés–relajación.

B. Estimación de la Probabilidad de Estrés–Relajación

Diversos estudios han demostrado que la variabilidad de la frecuencia cardíaca (HRV) constituye un marcador fisiológico confiable del estado de estrés y relajación, reflejando la actividad del sistema nervioso autónomo [14], [15]. En particular, la disminución de la HRV se asocia con estados de estrés agudo y crónico, mientras que valores elevados se relacionan con una mayor capacidad de autorregulación fisiológica. Estas propiedades han sido ampliamente explotadas en sistemas de detección de estrés en tiempo real basados en señales fisiológicas [17], [18].

No obstante, la relación entre las múltiples características HRV y el estado psicofisiológico del sujeto es inherentemente no lineal y dependiente del contexto temporal, lo que dificulta su modelado mediante enfoques analíticos clásicos. En consecuencia, resulta necesario emplear técnicas de aprendizaje automático capaces de capturar dichas relaciones complejas y proporcionar una estimación robusta del nivel de estrés–relajación.

Con este objetivo, además de las características HRV extraídas, se empleó un modelo de aprendizaje automático basado en Extreme Gradient Boosting (XGBoost), siguiendo exactamente la metodología propuesta en [11]. XGBoost es un algoritmo supervisado basado en ensambles de árboles de decisión y técnicas de gradient boosting, diseñado para capturar relaciones no lineales complejas y ampliamente utilizado en tareas de clasificación y regresión con señales fisiológicas [19].

La elección de este modelo se fundamenta en los resultados reportados en [11], donde XGBoost demostró un desempeño robusto en la detección de estrés en tiempo real a partir de señales ECG de una sola derivación. En dicho trabajo se destaca su eficiencia computacional, su capacidad de generalización inter–sujeto y su idoneidad para aplicaciones en línea, lo cual lo hace particularmente adecuado para ser integrado

dentro de un esquema de control adaptativo en lazo cerrado, como el propuesto en este estudio.

Las entradas del modelo XGBoost corresponden a un conjunto de 21 índices fisiológicos extraídos de la señal ECG de una sola derivación, calculados sobre ventanas deslizantes de 60 segundos con un corrimiento de 1 segundo. Dichos índices incluyen métricas de HRV en el dominio del tiempo (por ejemplo, mHR, SDNN, RMSSD, pNN50), del dominio de la frecuencia y métricas no lineales (SD1, SD2, SD1/SD2), así como características respiratorias derivadas del ECG, siguiendo la metodología descrita en [11].

La salida del modelo XGBoost se define como una probabilidad continua asociada al estado de estrés del sujeto, obtenida mediante la función predict proba. En este trabajo, dicha salida se reinterpreta como una probabilidad de relajación, denotada como y(t) y acotada en el intervalo [0, 1], donde valores cercanos a 0 representan un estado de estrés y valores cercanos a 1 indican un estado de relajación. Esta representación probabilística permite transformar información fisiológica de alta dimensionalidad en una única variable escalar adecuada para ser utilizada como señal de salida dentro de un esquema de control en lazo cerrado.

El modelo fue entrenado utilizando la base de datos del proyecto ES3, que contiene registros ECG de una sola derivación correspondientes a 40 participantes bajo diferentes condiciones experimentales de estrés y reposo [11]. El proceso de entrenamiento y validación se realizó siguiendo una estrategia de validación cruzada leave-one-subject-out (LOSO), con el fin de evaluar la capacidad de generalización inter-sujeto del clasificador, tal como se reporta en el artículo de referencia.

C. Modelo ARX nominal del paciente sintético

La identificación del modelo ARX no recursivo se realizó a partir de los datos obtenidos en el experimento preliminar, utilizando como entradas las características musicales extraídas de la señal de audio (NPitch, Short-Time Energy, Zero Crossing Rate, Spectral Centroid y Spectral Kurtosis) y como salida la probabilidad de relajación estimada mediante el modelo basado en XGBoost. Con el fin de obtener una estructura dinámica adecuada que represente el comportamiento promedio del sujeto, se llevó a cabo un proceso de selección automática de los órdenes del modelo.

1) Características musicales de entrada:

Las entradas del modelo dinámico corresponden a cinco características musicales extraídas de la señal de audio reproducida durante el experimento: *NPitch, Short-Time Energy (STE), Zero Crossing Rate (ZCR), Spectral Centroid y Spectral Kurtosis*. Estas características se obtuvieron a partir de la señal musical segmentada en ventanas de 1 segundo y calculadas de forma sincronizada con la estimación de la probabilidad de relajación, permitiendo establecer una relación dinámica entre estímulo musical y respuesta fisiológica.

Desde el punto de vista físico–perceptual, cada una de estas características describe un atributo específico del contenido musical. La característica NPitch representa la altura tonal dominante del fragmento musical y está asociada a la percepción de melodía, la cual ha sido reportada como un factor influyente en la respuesta emocional del oyente. La ShortTime Energy cuantifica la energía instantánea de la señal y está directamente relacionada con la intensidad o volumen percibido, variable que puede inducir estados de activación o relajación.

Por su parte, el Zero Crossing Rate mide la tasa de cruces por cero de la señal temporal y se asocia con el contenido de altas frecuencias y la textura del sonido. El Spectral Centroid describe la distribución espectral de la energía y se interpreta como una medida del "brillo" del sonido, mientras que la Spectral Kurtosis caracteriza la forma del espectro, permitiendo diferenciar señales con componentes espectrales más concentrados o dispersos.

La selección de estas características se fundamentó en dos criterios principales. En primer lugar, todas ellas pueden ser calculadas de manera eficiente en tiempo real, lo cual resulta compatible con los requerimientos del esquema de control propuesto. En segundo lugar, se realizó un análisis preliminar de la incidencia de cada característica sobre la salida del modelo, evaluando las ganancias estáticas estimadas entre las entradas musicales y la probabilidad de relajación. Dicho análisis permitió identificar que estas cinco características presentan una influencia significativa y diferenciable sobre la dinámica del sistema, mientras que otras características musicales evaluadas mostraron contribuciones marginales o altamente correlacionadas entre sí.

En consecuencia, el conjunto seleccionado constituye un compromiso adecuado entre relevancia fisiológica, interpretabilidad perceptual y viabilidad computacional, permitiendo capturar de forma efectiva la relación entre el estímulo musical y la respuesta psicofisiológica modelada mediante la probabilidad de relajación estimada.

2) Optimización de los órdenes del modelo ARX:

Con el objetivo de obtener una representación dinámica adecuada del paciente sintético para simulación y diseño de control, se implementó un algoritmo de optimización de tipo evolutivo, similar a un Algoritmo Genético (GA), orientado a la búsqueda de los órdenes óptimos del modelo ARX. En este esquema, cada individuo de la población representa una combinación de órdenes (na, nb, nk) correspondientes al denominador,

numeradores y retardos de entrada del modelo.

La población inicial se generó de forma aleatoria dentro de un rango predefinido, y la calidad de cada individuo se evaluó mediante el porcentaje de ajuste del modelo sobre un conjunto de datos de validación independiente. Como función de aptitud (fitness) se utilizó el índice de ajuste proporcionado por la función compare de MATLAB, el cual cuantifica la capacidad del modelo para reproducir la dinámica observada en los datos experimentales bajo simulación libre (sin predicción).

Durante cada generación, el individuo con mejor desempeño fue preservado, mientras que el resto de la población se actualizó mediante perturbaciones aleatorias controladas, análogas a un operador de mutación. El paso de mutación se redujo progresivamente a lo largo de las generaciones, permitiendo una exploración amplia del espacio de soluciones en las etapas iniciales y una refinación local en las iteraciones finales.

Como resultado del proceso de optimización, se seleccionó la estructura ARX que maximizó el ajuste del modelo sobre los datos de validación. El modelo nominal quedó definido por el siguiente vector de órdenes:

$$\text{orden}_{\text{base}} = [27, 4, 4, 4, 4, 4, 1, 1, 1, 1, 1],$$

donde el primer término corresponde al orden del denominador, los siguientes cinco términos a los órdenes de los numeradores asociados a cada una de las características musicales consideradas como entradas, y los últimos cinco términos a los retardos de entrada respectivos.

Este modelo ARX no recursivo fue adoptado como la planta nominal del paciente sintético, ya que permite capturar de manera adecuada la dinámica promedio entre las características musicales y la probabilidad de relajación estimada. Sobre esta representación se realizaron las simulaciones numéricas y el ajuste inicial del esquema de control predictivo adaptativo.

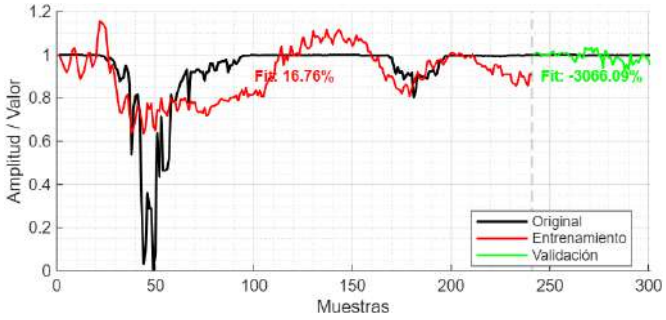


Fig. 2. Comparación de Modelo: Entrenamiento vs Validación

La Fig. 2 muestra el desempeño del modelo ARX no recursivo seleccionado, evaluado en modo de simulación libre (sin predicción), de acuerdo con el criterio empleado durante el proceso de optimización. En dicha figura se comparan la salida real del sistema con las salidas simuladas del modelo sobre los conjuntos de entrenamiento y validación.

Como puede observarse, el modelo presenta un ajuste razonable tanto en el conjunto de entrenamiento como en el de validación. Visualmente, la respuesta del modelo sobre los datos de validación reproduce adecuadamente la tendencia de la señal original, incluso con un seguimiento más consistente que en el conjunto de entrenamiento.

No obstante, el índice de ajuste porcentual (Fit) reportado para la validación presenta un valor muy bajo. Este resultado no refleja un mal comportamiento del modelo, sino que se debe a la forma en que MATLAB calcula dicha métrica, la cual es altamente sensible a la varianza del conjunto de datos evaluado. En particular, cuando la señal de validación presenta una variabilidad reducida, pequeños errores absolutos pueden traducirse en valores de Fit extremadamente negativos.

Adicionalmente, dado que el propósito del modelo ARX no recursivo es actuar como planta nominal del paciente sintético para simulación y ajuste inicial del controlador, y no como un predictor generalizable, el nivel de ajuste obtenido resulta suficiente para los objetivos del presente trabajo. La compensación de estas discrepancias se aborda posteriormente mediante identificación ARX recursiva, integrada dentro del lazo de control adaptativo.

D. Identificación de la Planta (ARX Recursivo)

La dinámica del sujeto, considerada como la planta del sistema de control, se modela mediante un esquema de identificación ARX recursiva de múltiples entradas y una salida (MISO). Este modelo tiene como objetivo capturar, en línea, la relación dinámica entre las características musicales del estímulo aplicado y la respuesta fisiológica del sujeto, expresada en términos de la probabilidad de relajación estimada.

La salida del modelo corresponde a la probabilidad de relajación $y(k) \in [0, 1]$, obtenida a partir del modelo de aprendizaje automático basado en características HRV descrito previamente.

Las entradas del modelo ARX corresponden a las características musicales extraídas de la señal de audio, las cuales se clasifican según su rol dentro del esquema de control:

- Variable manipulada (MV): la Short-Time Energy, la cual es directamente afectada por la ganancia musical u(k) aplicada por el controlador MPC. Esta

variable representa el principal canal de actuación del sistema de control sobre el estímulo musical.

- Perturbaciones medidas (MD): las restantes características musicales, a saber: NPitch, Zero Crossing Rate, Spectral Centroid y Spectral Kurtosis. Estas señales no son controladas directamente, pero influyen en la respuesta fisiológica del sujeto y se consideran explícitamente como entradas exógenas medidas.

El modelo ARX recursivo se define por los órdenes na, nb y nk. Tras un proceso de experimentación y ajuste empírico, se seleccionó la siguiente configuración para la identificación en línea:

$$n_a = 5,\ n_b = 5 \times [1, 1, 1, 1, 1],\ n_k = 8 \times [1, 1, 1, 1, 1],$$

donde na corresponde al orden del denominador del modelo, mientras que nb y nk representan los órdenes y retardos asociados a cada una de las cinco entradas musicales.

Para permitir la adaptación del modelo frente a cambios inter e intra-sujeto en la respuesta psicofisiológica, se emplea un algoritmo de identificación ARX recursiva con factor de olvido. El factor de olvido utilizado fue

$$\lambda = 1 - \frac{1}{50},$$

lo que permite dar mayor peso a las mediciones recientes, favoreciendo una rápida adaptación del modelo ante variaciones dinámicas del sistema, sin comprometer la estabilidad numérica del estimador.

El modelo ARX recursivo identificado se convierte posteriormente a una representación en espacio de estados, incorporando explícitamente la asignación de señales para su uso dentro del esquema de Control Predictivo Basado en Modelo. En particular, la Short-Time Energy se define como variable manipulada (MV), mientras que las demás características musicales se consideran perturbaciones medidas (MD).

Este modelo ARX recursivo constituye el núcleo adaptativo del esquema de control, permitiendo actualizar en línea la dinámica de la planta y mejorar el desempeño del controlador MPC frente a la variabilidad inherente de la respuesta fisiológica humana.

E. Ciclo de Control MPC Adaptativo

El controlador principal es un MPC que opera en un lazo de control adaptativo, optimizando la amplitud musical ($u(t)$) en cada iteración.

1) Configuración del MPC Adaptativo:

El controlador predictivo basado en modelo (MPC) fue seleccionado frente a estrategias clásicas como PID, LQR o control por ganancia programada debido a su capacidad para predecir el comportamiento futuro del sistema a partir de un modelo dinámico y de información histórica, sin depender de una estructura específica del modelo. El MPC formula el problema de control como una optimización en línea sobre un horizonte temporal móvil, minimizando una función de costo que penaliza el error de seguimiento y el esfuerzo de control, lo que le permite adaptarse continuamente a cambios en la dinámica del sistema y a la presencia de incertidumbres.

Dadas estas características, el MPC es particularmente adecuado para sistemas complejos que presentan retardos, restricciones y variaciones dinámicas, como el sistema analizado en este trabajo. En este contexto, controladores clásicos como el PID pueden resultar insuficientes, mientras que el MPC permite manejar explícitamente estas condiciones, garantizando un mejor seguimiento de la referencia y mayor robustez del sistema.

El controlador MPC se configuró con los siguientes parámetros, extraídos directamente de la implementación:

- Horizonte de Predicción *(P): P = 8*
- Horizonte de Control *(M): M = 5*
- Restricciones de la Variable Manipulada (Amplitud): La amplitud musical se restringió al rango $u_{min} = 0$ y $u_{max} = 5$. Las restricciones en la tasa de cambio fueron $\Delta u_{min} = -0.553$ y $\Delta u_{max} = 0.553$.
- Pesos de la Función de Costo: La función de costo se ponderó para priorizar el seguimiento de la referencia sobre la acción de control. Se utilizó un peso de $W_{Output} = 4$ para la variable de salida (probabilidad de relajación), $W_{MV} = 0.01$ para la acción de control y $W_{MV\,Rate} = 0.0$ para la tasa de cambio de la acción de control.

F. Extremum Seeking Control (ESC)

Dado que la relación entre la estimulación musical y la respuesta fisiológica del sujeto es altamente no lineal, dependiente del tiempo y sujeta a variaciones inter e intrasujeto, se empleó un esquema de Extremum Seeking Control (ESC). Este enfoque permite optimizar en línea una función de costo desconocida, utilizando únicamente mediciones de salida del sistema, sin requerir un modelo explícito de la dinámica subyacente [12].

En el esquema ESC implementado, la variable de control corresponde a la referencia $\hat{u}(t)$, la cual ajusta indirectamente la ganancia aplicada a la señal musical reproducida al sujeto. Esta ganancia modula la amplitud del estímulo sonoro, afectando principalmente la energía de la señal musical y, en consecuencia, la respuesta fisiológica del usuario. La salida del sistema es la probabilidad de relajación $y(t)$, estimada a partir de las

características HRV mediante el modelo de aprendizaje automático descrito previamente.

El principio de funcionamiento del ESC se basa en la inyección de una señal de perturbación sinusoidal (dither) sobre la referencia de control, lo que permite estimar en línea el gradiente de la función de costo asociada al nivel de relajación. Esta estimación se obtiene mediante un proceso de demodulación y filtrado, seguido de una integración que ajusta la referencia en la dirección que maximiza la relajación del sujeto.

Los parámetros fundamentales del esquema ESC son:

- Frecuencia de la señal de perturbación: $f = 0.1$ *Hz*.
- Amplitud de la señal de perturbación: $A = 0.1$.
- Ganancia de integración: $K = 2.3$.

La selección de estos parámetros se realizó mediante un proceso de ajuste progresivo, considerando las características temporales del sistema y el tiempo de muestreo de 1 segundo. Dado que la evaluación confiable de la respuesta fisiológica requiere observar su evolución en ventanas temporales suficientemente largas, se determinó que ventanas de al menos 10 segundos eran necesarias para capturar adecuadamente la dinámica del sistema. En pruebas preliminares con ventanas de 5 segundos se observó una respuesta ruidosa e inconsistente, lo que motivó la elección final de una frecuencia de perturbación de 0.1 Hz, compatible con el horizonte temporal de análisis.

La amplitud del dither y la ganancia de integración fueron ajustadas de manera que la referencia pudiera modificarse de forma progresiva y segura, evitando cambios abruptos en la estimulación musical que pudieran generar respuestas fisiológicas adversas o inestabilidad en el sistema.

La función de costo empleada para guiar el proceso de optimización se define como:

$$J(k) = \left(\mathrm{ref}_{\mathrm{base}} - y(k)\right)^2, \qquad [1]$$

donde $y(k)$ representa la probabilidad de relajación estimada en el instante k, y $ref_{base} = 1$ corresponde al estado deseado de máxima relajación. Bajo este planteamiento, el esquema ESC no solo permite el seguimiento de una referencia predefinida, sino que además posibilita la búsqueda activa, en tiempo real, del punto de máxima relajación del sujeto, adaptándose dinámicamente a su respuesta fisiológica.

El ciclo de control adaptativo se ejecuta en cada iteración de la siguiente manera:

1. Extracción de Características: Se extraen características musicales en ventanas de un segundo.
2. Actualización de la Planta: El estimador de estados y el modelo ARX recursivo actualizan los parámetros del sujeto en tiempo real, generando un nuevo modelo de planta $\hat{G}(z, t)$.
3. Optimización MPC: El controlador MPC utiliza el modelo $\hat{G}$ (z, t) para calcular la secuencia óptima de acciones de control (u_{MPC}) que minimiza una función de costo. La referencia de control es generada por el esquema ESC.
4. Acción de Control: La acción de control final u(t) se aplica a la amplitud musical.

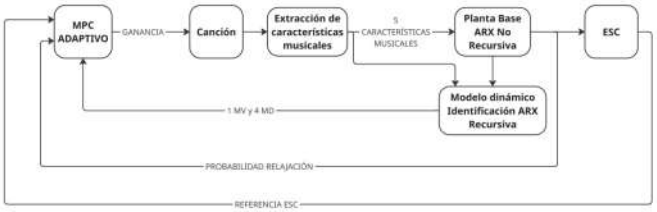


Fig. 3. Diagrama de bloques del sistema de control simulado sobre el paciente sintético (Planta Base)

El sistema realiza una verificación de estabilidad del modelo de la planta en cada iteración antes de aplicar el control adaptativo.

III. RESULTADOS EN SIMULACIÓN

Con el fin de evaluar el desempeño del esquema de control propuesto en un entorno controlado, se realizaron simulaciones numéricas utilizando el paciente sintético y el controlador MPC adaptativo con referencia generada mediante ESC, de acuerdo con el diagrama de bloques de la Fig. 3. En esta sección se presentan los resultados más relevantes.

A. Evolución de la Referencia Ideal y Función de Costo

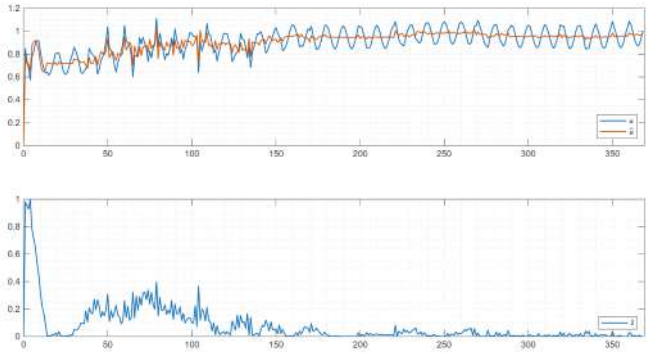


Fig. 4. Evolución de la referencia ideal estimada $\hat{u}$ y disminución del costo del controlador MPC.

En la Fig. 4 se muestra la evolución de la referencia ideal estimada (û) junto con la señal de control aplicada. Este comportamiento representa el escenario deseado del sistema, donde la referencia no es fija, sino que se ajusta dinámicamente de acuerdo con el estado del

sujeto. Se ilustra la evolución del costo asociado al controlador MPC, donde se observa una disminución progresiva del mismo a lo largo del tiempo. Este comportamiento indica que el controlador logra minimizar de forma efectiva el error de seguimiento y el esfuerzo de control, reflejando un proceso de adaptación adecuado del sistema ante la dinámica variante del sujeto.

B. Seguimiento de la Referencia de Relajación

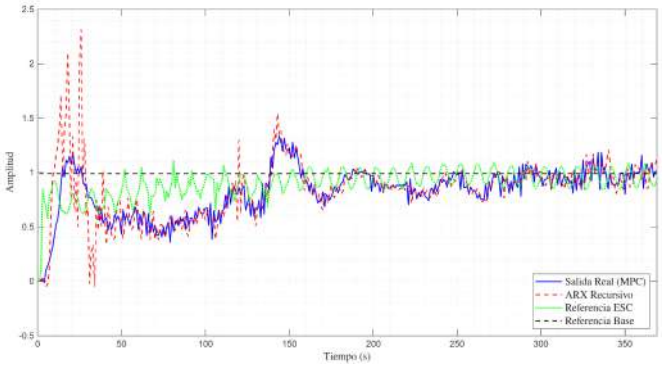


Fig. 5. Seguimiento de la referencia de relajación generada por el ESC: salida real del MPC, modelo ARX recursivo y referencia base.

En la Fig. 5 se presenta la comparación entre la salida real del sistema, la salida estimada mediante el modelo ARX recursivo y la referencia generada por el esquema ESC. Puede observarse que la señal de referencia incrementa de manera gradual, representando la capacidad progresiva del sujeto para alcanzar estados de mayor relajación.

A medida que transcurre el tiempo, la referencia converge hacia un valor cercano a 1, el cual representa un estado estable asociado a la relajación del sistema. La salida controlada sigue dicha referencia de manera progresiva, presentando transiciones suaves y evitando variaciones abruptas que podrían resultar fisiológicamente incómodas o incluso contraproducentes. Este comportamiento pone de manifiesto la capacidad del controlador MPC adaptativo para gestionar adecuadamente las restricciones del sistema y priorizar una evolución segura y controlada de la señal.

Adicionalmente, se observa en la Fig. 5 que el proceso de identificación presenta una fase transitoria durante los instantes iniciales del experimento. En particular, durante los primeros 50 segundos, los parámetros del modelo ARX recursivo aún no han convergido a valores estables, debido a la limitada información disponible en esta etapa inicial. Por esta razón, durante dicho intervalo no se implementa el esquema de MPC adaptativo. En su lugar, se emplea un controlador MPC basado en el mejor modelo previamente identificado mediante un proceso de identificación no recursiva, garantizando así un comportamiento estable del sistema mientras el modelo recursivo completa su proceso de adaptación.

Una vez transcurrido este periodo y alcanzada una mayor estabilidad en la estimación de los parámetros, entra en funcionamiento el MPC adaptativo apoyado en el modelo ARX identificado de forma recursiva. Esta transición garantiza un desempeño más robusto del control, al evitar la utilización de modelos inestables o poco representativos de la dinámica real del sistema. Este comportamiento se aprecia con mayor claridad en la Fig. 7, donde se evidencia la convergencia y estabilización de los parámetros del modelo identificado.

C. Comportamiento de la Señal de Control

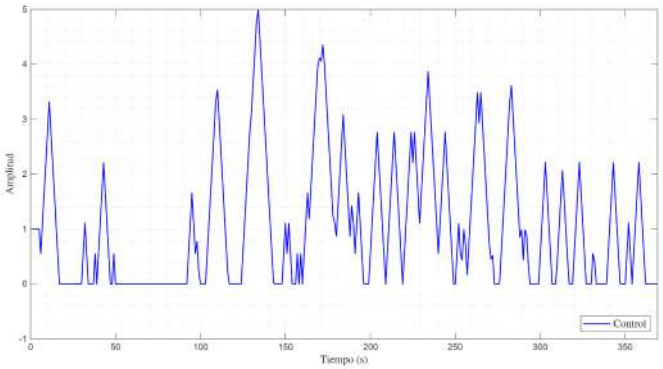


Fig. 6. Evolución temporal de la señal de control correspondiente a la ganancia musical.

La Fig. 6 muestra la evolución temporal de la señal de control u(t), correspondiente a la ganancia aplicada a la señal musical. Se observa que la acción de control presenta variaciones significativas a lo largo del experimento, reflejando la naturaleza adaptativa del esquema de control y la necesidad de compensar cambios en la dinámica estimada del sistema.

Si bien la señal de control no es constante, sus variaciones se mantienen dentro de un rango acotado y no presentan cambios abruptos de alta frecuencia. Este comportamiento es consecuencia directa de la formulación del MPC, el cual penaliza variaciones excesivas de la entrada y prioriza una evolución progresiva de la ganancia, compatible con restricciones perceptuales y fisiológicas.

Adicionalmente, la variabilidad observada en u(t) está asociada al proceso de exploración y ajuste continuo introducido por el esquema de Extremum Seeking Control (ESC), cuyo objetivo es desplazar la referencia hacia regiones que maximicen el nivel de relajación estimado. En este contexto, las oscilaciones de la señal de control son esperadas y necesarias para permitir la adaptación del sistema ante condiciones dinámicas y no estacionarias del sujeto.

En conjunto, estos resultados evidencian que el controlador MPC adaptativo es capaz de modular de forma efectiva la amplitud del estímulo musical, manteniendo la estabilidad del sistema y evitando variaciones abruptas que podrían resultar fisiológicamente incómodas o contraproducentes.

D. Evolución de los Parámetros del Modelo ARX Recursivo

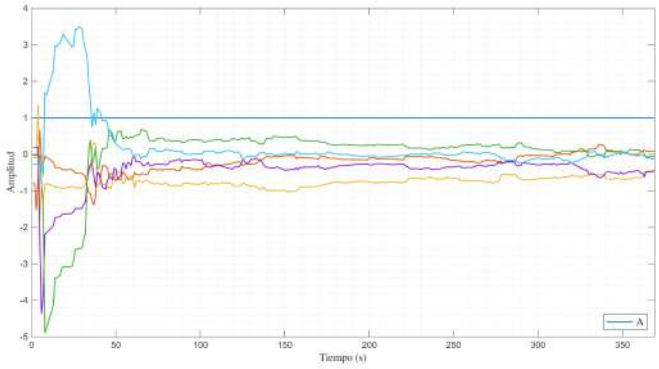


Fig. 7. Evolución temporal de los coeficientes del denominador del modelo ARX recursivo durante el proceso de identificación adaptativa.

Finalmente, se analizó la evolución temporal de los coeficientes del denominador del modelo ARX recursivo. Inicialmente, los parámetros presentan una distribución dispersa y poco definida, como se puede ver en la Fig. 7, lo cual es consistente con una fase de identificación temprana y alta incertidumbre del modelo.

Sin embargo, se observa que los coeficientes comienzan a ajustarse progresivamente y muestran variaciones cada vez menores. Este comportamiento indica que el algoritmo de identificación recursiva converge hacia un conjunto de parámetros estables, lo que permite una representación más precisa de la dinámica del sistema y mejora el desempeño del controlador adaptativo.

IV. RESULTADOS EXPERIMENTALES Y DISCUSIÓN

Los datos experimentales analizados en esta sección corresponden a experimentos con sujetos descritos en [8]. El estudio contó con la participación de 43 sujetos voluntarios, estudiantes universitarios del Programa de Ingeniería Electrónica de la Universidad de Nariño, ubicada en Pasto, Colombia. La muestra fue seleccionada siguiendo criterios de inclusión específicos, con el fin de garantizar condiciones basales homogéneas y minimizar la influencia de variables confusoras en la respuesta fisiológica al estrés.

1) Criterios de Inclusión:

Los participantes debían cumplir con los siguientes requisitos:

- Edad comprendida entre 19 y 36 años
- Ausencia de diagnóstico previo de trastornos psiquiátricos o neurológicos
- Ausencia de consumo regular de medicamentos psicotrópicos
- No consumo de cafeína, alcohol ni sustancias estimulantes en las 24 horas previas

2) Características Demográficas:

La población experimental presentó las siguientes características generales:

- Distribución por sexo: 32 hombres (74.4%) y 11 mujeres (25.6%)
- Edad promedio: 23 ± 3 años.

A. Grupos Experimentales

Los 43 participantes fueron asignados aleatoriamente a tres grupos experimentales, procurando mantener proporciones similares entre hombres y mujeres, con el objetivo de evitar sesgos asociados a diferencias de género en la respuesta al estrés [21]. La distribución resultante se presenta en la Tabla correspondiente:

TABLE I
DISTRIBUCIÓN DE PARTICIPANTES POR GRUPO EXPERIMENTAL

Grupo	N	Hombres	Mujeres
GC1 (Sin música)	15	11	4
GC2 (Música abierta)	14	11	3
GE (Música controlada)	14	10	4
Total	**43**	**32**	**11**

1) Descripción de los Grupos:

- GC1 (Grupo de Control 1 - Sin música): Los participantes de este grupo realizaron el protocolo experimental completo sin estímulo musical alguno durante ninguna de las etapas. Este grupo sirve como línea base para cuantificar el efecto puro del estresor cognitivo sobre las variables fisiológicas y psicométricas, sin la influencia de ninguna intervención externa.
- GC2 (Grupo de Control 2 - Música en lazo abierto): Los participantes escucharon la pieza musical Canon in D Major de Johann Pachelbel con ganancia fija (amplitud constante igual a 1.0) durante la etapa de estresor con estímulo (E4). Este grupo permite aislar el efecto de la presencia de música sobre la respuesta al estrés, independientemente de cualquier estrategia de control adaptativo, y sirve como referencia para evaluar la mejora introducida por el sistema de control propuesto.

- GE (Grupo Experimental - Música con control adaptativo): Los participantes escucharon la misma pieza musical que GC2, pero modulada dinámicamente por el controlador MPC adaptativo a través de la ganancia en la etapa E4. En este grupo se implementó el esquema completo de control en lazo cerrado, integrando:
  – Estimación continua del estado de relajación mediante el modelo XGBoost
  – Identificación recursiva ARX de la dinámica del sujeto
  – Control predictivo adaptativo (MPC)
  – Ajuste dinámico de la referencia de relajación

B. Protocolo Experimental y Etapas

Cada sesión experimental tuvo una duración total de 670 segundos (aproximadamente 11 minutos) y se llevó a cabo en un ambiente controlado. Durante todo el experimento, se registraron de forma continua las señales de electrocardiograma (ECG) de una sola derivación mediante el sistema de adquisición biomédica desarrollado en el proyecto base [11].

El protocolo se estructuró en cuatro etapas secuenciales claramente diferenciadas, diseñadas para inducir, medir y modular progresivamente el estado de estrés de los participantes:

1) E1 (1-240 s): Relajación Basal: Duración: 240 segundos (4 minutos)

Esta etapa inicial tiene como objetivo establecer el estado de relajación basal de cada participante, el cual sirve como referencia individual para todo el análisis posterior. Durante este periodo:

- Los participantes permanecieron sentados en una posición cómoda, en un ambiente silencioso y con iluminación tenue
- Se les instruyó para mantener los ojos cerrados y respirar de manera natural y relajada
- Se reprodujo un audio con instrucciones de relajación progresiva basadas en el método de Schultz [16]
- No se aplicó ningún estresor ni estímulo musical
- Se registraron continuamente las señales ECG para establecer los valores basales de HRV y demás índices fisiológicos

Al finalizar esta etapa, se administró el primer conjunto de cuestionarios psicométricos (PSS, STAI-E, STAI-R) para establecer las condiciones psicológicas iniciales de cada participante.

2) E2 (241-250 s): Transición: Duración: 10 segundos

Breve periodo de transición entre el estado de relajación basal y la fase de inducción de estrés. Durante este intervalo:

- Se informó verbalmente al participante sobre la próxima tarea a realizar
- Se presentaron las instrucciones para la prueba de estresor cognitivo

Esta etapa de transición permite que el participante procese cognitivamente la información sobre la tarea inminente, lo cual puede generar una respuesta anticipatoria de estrés observable en las señales fisiológicas.

3) E3 (251-370 s): Estresor Puro: Duración: 120 segundos (2 minutos)

Fase de inducción controlada de estrés agudo mediante una tarea cognitiva estándar. Esta etapa es idéntica para los tres grupos experimentales y tiene como objetivo generar un nivel moderado pero significativo de estrés psicológico. La tarea aplicada consistió en la resolución continua de operaciones aritméticas mentales presentadas a través de una interfaz desarrollada en MATLAB. El programa mostraba de forma secuencial problemas aritméticos de dificultad moderada, los cuales debían ser resueltos bajo restricción temporal, sin retroalimentación inmediata durante la ejecución.

Al finalizar la tarea, el sistema informaba al participante una puntuación asociada a su desempeño, lo cual introduce un componente adicional de evaluación cognitiva y presión psicológica.

La selección de este tipo de estresor se fundamenta en estudios previos que demuestran su capacidad para inducir cambios significativos en marcadores fisiológicos (frecuencia cardíaca, HRV) y psicométricos (STAI-E, VASS) de estrés agudo [17]. Adicionalmente, esta tarea genera una carga cognitiva y emocional comparable a la del Trier Social Stress Test (TSST), pero con mayor control experimental y un menor componente de estrés social.

4) E4 (371-670 s): Estresor con Estímulo: Duración: 300 segundos (5 minutos)

Fase de intervención diferencial entre grupos, donde se mantiene activo el estresor cognitivo mientras se introduce el estímulo musical según el grupo experimental. Esta es la etapa crítica donde se evalúa la efectividad del controlador propuesto.

Este diseño experimental permite evaluar de forma controlada y reproducible el efecto de la música, así como el impacto de un esquema de control adaptativo sobre la respuesta fisiológica al estrés, manteniendo constantes las condiciones del estresor y de adquisición fisiológica.

C. Estrategia de Referencia Adaptativa

Durante la fase de simulación se implementó un esquema de Extremum Seeking Control (ESC) para el

ajuste dinámico de la referencia de relajación [12]. Sin embargo, en los experimentos se identificaron limitaciones prácticas, principalmente asociadas al tiempo de convergencia y al uso de perturbaciones periódicas que podían afectar la comodidad del paciente. Por esta razón, en la implementación experimental se adoptó una estrategia de referencia adaptativa incremental, basada en la respuesta observada del sujeto.

Aunque difiere en su implementación, esta estrategia comparte con el ESC el objetivo fundamental de optimización en línea, buscando maximizar el estado de relajación sin conocer previamente la función fisiológica del sistema. En este sentido, puede interpretarse como una variante heurística y discreta del ESC, más adecuada para aplicaciones biomédicas reales.

D. Análisis Psicológico y Validación Experimental

Con el objetivo de analizar el efecto de cada condición experimental y validar la efectividad del protocolo de inducción de estrés, se evaluó el cambio promedio en las variables psicométricas entre las mediciones pre-intervención (post-E1) y post-intervención (post-E4). La diferencia se calculó como Post–Pre, donde un valor positivo indica un incremento en los niveles de estrés o ansiedad, mientras que un valor negativo representa una reducción.

La Tabla II resume las diferencias promedio obtenidas para cada grupo. La Ansiedad Estado (STAI-E) constituye la métrica más sensible al estresor agudo inducido durante el experimento, ya que refleja cambios transitorios en la activación emocional del sujeto [21].

TABLE II
ESTADÍSTICA DESCRIPTIVA DE LAS VARIABLES PSICOLÓGICAS (MEDIA ± DE)

Grupo	N	Variable	Pre	Post
GC1	15	PSS	18.80 ± 4.09	19.80 ± 5.78
		STAI-R	44.07 ± 8.54	47.47 ± 9.01
		STAI-E	37.47 ± 6.16	48.33 ± 10.26
GC2	14	PSS	18.21 ± 4.17	17.93 ± 4.41
		STAI-R	43.71 ± 8.59	42.64 ± 8.22
		STAI-E	37.50 ± 7.63	43.93 ± 9.66
GE	14	PSS	14.43 ± 7.02	17.29 ± 5.55
		STAI-R	36.36 ± 14.70	39.79 ± 11.49
		STAI-E	34.36 ± 13.56	39.50 ± 10.35

El grupo GC1 (sin música) presentó el mayor incremento en Ansiedad Estado ($\Delta = +10.86$), estableciendo una línea base del efecto del estresor matemático sin intervención, consistente con reportes previos sobre tareas cognitivas estresantes [17].

En el grupo GC2 (música abierta), el incremento fue menor ($\Delta = +6.43$), lo que sugiere que la presencia de música, aun sin regulación, ejerce un efecto mitigador parcial sobre la respuesta al estrés, en concordancia con estudios que reportan efectos ansiolíticos generales de la música [4], [5].

Finalmente, el grupo experimental GE (música controlada mediante MPC adaptativo) presentó el menor aumento en Ansiedad Estado ($\Delta = +5.14$). Comparado con GC1, este resultado representa una reducción del impacto del estresor del 52.7%, lo que evidencia la capacidad del sistema de control para atenuar de forma significativa la respuesta de ansiedad aguda, en línea con enfoques de regulación fisiológica asistida por control y sistemas inteligentes [11].

En contraste, las escalas PSS y STAI-Rasgo mostraron variaciones menos consistentes entre los grupos. Este comportamiento es esperable, dado que ambas métricas evalúan componentes más estables del estrés percibido y la ansiedad disposicional, respectivamente, y no están diseñadas para capturar cambios inmediatos inducidos por un estresor agudo [20], [21].

No obstante, los valores observados en el grupo experimental sugieren que la intervención musical adaptativa no incrementa el estrés basal ni la ansiedad rasgo, lo que refuerza la seguridad del esquema de control propuesto.

E. Tiempo en Rango (TIR) en Estado de Relajación

La principal métrica fisiológica utilizada fue el Tiempo en Rango (TIR) en estado de relajación, definido como el porcentaje de tiempo en el cual la salida del sistema permanece dentro del intervalo $0.5 \leq y(t) \leq 1$. Esta métrica fue evaluada a lo largo de cuatro etapas del experimento: E1 (Basal), E2 (Transición), E3 (Estresor Puro) y E4 (Control Activo).

Con el fin de evaluar la dispersión y robustez de los resultados, se realizó un análisis estadístico descriptivo del TIR por grupo y etapa. La Tabla III presenta la media, mediana y desviación estándar del TIR en estado relajado.

TABLE III
ESTADÍSTICAS DESCRIPTIVAS POR ETAPA Y GRUPO EN ESTADO RELAJADO

Etapa	Grupo	Media (%)	Mediana (%)	DE (%)
E1	GC1	72.22	90.00	35.93
	GC2	66.49	77.71	32.81
	GE	83.72	94.17	22.71
E2	GC1	72.67	100.00	45.43
	GC2	88.57	100.00	26.85
	GE	78.57	100.00	37.80
E3	GC1	42.39	25.83	39.09
	GC2	35.30	29.17	27.70
	GE	46.61	47.92	30.65
E4	GC1	48.38	51.33	39.06
	GC2	49.81	53.33	37.19
	GE	55.74	50.00	38.29

Durante la fase basal (E1), todos los grupos presentaron TIR elevados (>66%), sin diferencias marcadas entre ellos, lo que confirma condiciones iniciales comparables y valida la aleatorización de los participantes. En la fase de transición (E2), se observó un incremento generalizado del TIR en todos los grupos, posiblemente debido a la familiarización con el entorno experimental y la ausencia de estresores activos.

La fase de estresor puro (E3) indujo una reducción drástica del TIR en todos los grupos (GC1: 42.39%, GC2: 35.30%, GE: 46.61%), validando la efectividad del protocolo de inducción de estrés mediante la tarea. Notablemente, el GE mantuvo un TIR ligeramente superior incluso sin intervención activa, aunque esta diferencia se debe a variabilidad aleatoria dado que aún no existía una acción de control diferenciada.

1) Desempeño en la Fase de Control Activo (E4):

En la fase de control activo (E4), el GE alcanzó un TIR promedio de 55.74%, superando tanto el umbral de relajación (50%) como a los grupos de control (GC1: 48.38%, GC2: 49.81%). Esta mejora del +11.79% respecto a GC1 y +11.95% respecto a GC2 valida la hipótesis de que el MPC adaptativo, al modular la amplitud musical en función del estado del sujeto, genera un impacto positivo y sostenido en la regulación del estrés.

Sin embargo, es importante analizar la relación entre la media y la mediana para comprender la naturaleza de esta mejora. La mediana del GE alcanzó el 50.00%, valor comparable al de los grupos de control (GC1: 51.33%, GC2: 53.33%). El hecho de que la media (55.74%) sea superior a la mediana (50.00%) indica una distribución asimétrica positiva, es decir, que existen sujetos con respuestas particularmente favorables al control adaptativo que elevan el promedio.

El controlador MPC incrementó el TIR promedio y amplificó las respuestas en sujetos altamente controlables, evidenciando un mecanismo de personalización implícito (como se observa en la Tabla IV, donde 6 sujetos alcanzaron TIR > 85%). Simultáneamente, para sujetos con menor controlabilidad, el sistema mantuvo un desempeño cercano al 50%, evitando deterioros significativos.

En contraste, los grupos de control presentan medianas ligeramente superiores pero medias inferiores, sugiriendo distribuciones más simétricas pero con menor capacidad de alcanzar estados de alta relajación. Este patrón evidencia que, aunque el controlador no benefició uniformemente a todos los sujetos, logró mejoras sustanciales en una fracción significativa de la población sin comprometer la estabilidad del resto.

2) Variabilidad Inter-Sujeto y Justificación del Control Adaptativo:

Las elevadas desviaciones estándar observadas en todas las etapas (rangos de 22-45%) reflejan la alta variabilidad inter-sujeto característica de las respuestas psicofisiológicas. Esta variabilidad no disminuyó en la fase de control activo (E4: GE=38.29%, GC1=39.06%, GC2=37.19%), lo cual es consistente con las diferencias individuales en controlabilidad documentadas en la literatura de sistemas biológicos.

Esta variabilidad justifica precisamente el enfoque adaptativo del controlador propuesto, diseñado específicamente para personalizar la intervención según la respuesta individual de cada sujeto. El análisis por sujeto individual (Tabla IV) confirma que el desempeño del sistema depende fuertemente de la capacidad de identificación del modelo ARX y de la adaptabilidad fisiológica del participante, aspectos que varían considerablemente entre individuos.

Adicionalmente, se observó que la salida del modelo XGBoost presenta un comportamiento coherente con las distintas etapas del protocolo experimental. Durante la fase basal (E1), la probabilidad de relajación estimada alcanza valores elevados, mientras que en la fase de estresor puro (E3) se evidencia una reducción significativa de dicha variable. En la etapa de control activo (E4), la probabilidad de relajación muestra un incremento progresivo, consistente con la aplicación del esquema de control adaptativo.

Este comportamiento sugiere que el modelo de estimación captura adecuadamente las transiciones fisiológicas inducidas por el protocolo experimental, proporcionando una variable de salida interpretable y adecuada para su uso dentro del lazo de control.

F. Análisis por Sujeto y Robustez del Control

Con el fin de evaluar la robustez del esquema propuesto frente a la variabilidad fisiológica intersujeto, se

realizó un análisis individual considerando conjuntamente las métricas de desempeño ITAE y el tiempo en región de interés (TIR).

La Tabla IV resume los patrones de comportamiento observados y permite clasificar a los sujetos según su grado de controlabilidad fisiológica.

TABLE IV

RESULTADOS DEL CONTROL POR SUJETO SEGÚN MÉTRICAS ITAE Y TIR

Sujeto	Comportamiento observado	Resultado (ITAE / TIR)
15, 18, 32, 34, 39, 42	Alta adaptabilidad fisiológica y correcta identificación del modelo (Fig. 8).	ITAE bajo con una mediana de 2676.8, TIR Alto Estrés bajo (252–295 s) que quivale a una mediana en porcentaje de tiempo relajado del 95.6667 Control exitoso.
11, 12, 13, 23, 26	Seguimiento parcial de la referencia con adaptación progresiva (Fig. 9).	ITAE con una mediana de 13246, TIR medio (65–144 s), que quivale a una mediana en porcentaje de tiempo relajado del 38.6667 Control dinámico.
14, 20, 35	Respuesta fisiológica rígida o no adecuadamente identificable (Fig. 10).	ITAE alto con una mediana de 23755, TIR bajo (2–23 s), que quivale a una mediana en porcentaje de tiempo relajado de 2 Baja controlabilidad fisiológica.

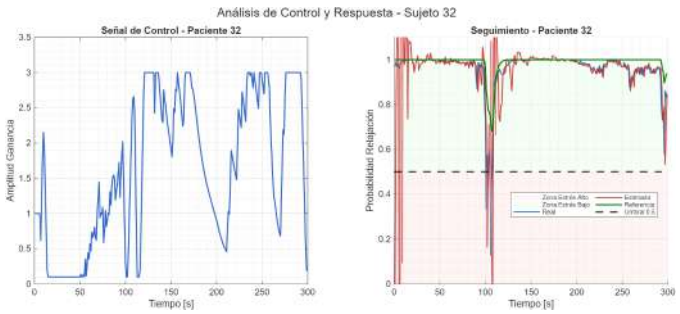


Fig. 8. Paciente 32 – Caso de mejor desempeño.

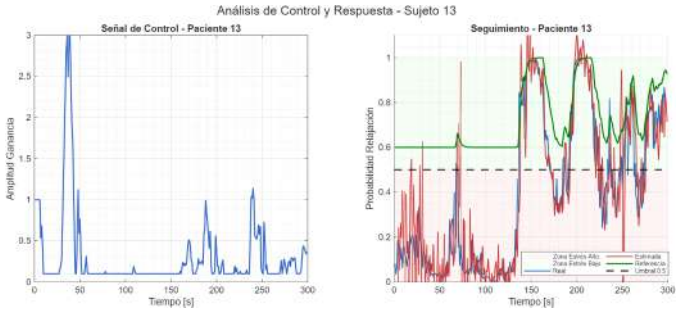


Fig. 9. Paciente 13 – Caso de desempeño intermedio.

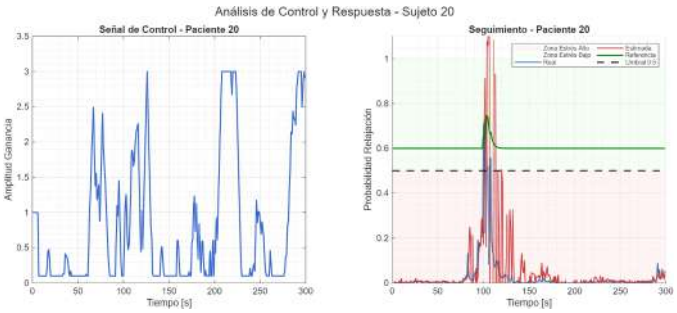


Fig. 10. Paciente 20 – Caso fisiologicamente más desafiante.

El análisis por sujeto evidencia que el desempeño del controlador no es uniforme y depende fuertemente de la capacidad de adaptación fisiológica individual, así como de la calidad del modelo identificado en línea. Los sujetos con alta adaptabilidad presentan valores bajos de ITAE y medianas elevadas de TIR (ver Fig. 8), lo que indica un seguimiento efectivo y sostenido de la referencia de relajación.

En los sujetos con adaptación parcial, como el mostrado en la Fig. 9 el sistema logra mantener la estabilidad y reducir el impacto del estresor, aunque con un seguimiento incompleto de la referencia. Este comportamiento se refleja en valores intermedios de ITAE y TIR, lo que sugiere que el esquema MPC adaptativo conserva robustez ante la incertidumbre moderada del modelo.

Por el contrario, los sujetos con respuestas fisiológicas rígidas o con procesos de identificación no confiables (Fig. 10) exhiben valores elevados de ITAE y tiempos mínimos en la región de relajación, evidenciando un control fisiológico limitado. Este resultado pone de manifiesto las limitaciones inherentes al control de sistemas biológicos altamente variables y resalta la necesidad de incorporar mecanismos de detección temprana de baja controlabilidad y estrategias de personalización en trabajos futuros.

V. CONCLUSIONES

En este trabajo se propuso y evaluó un sistema de control en lazo cerrado para la regulación del estrés mediante estímulos musicales ajustables, integrando identificación recursiva de la dinámica psicofisiológica del sujeto y un esquema de Control Predictivo Basado en Modelo (MPC) adaptativo con generación dinámica de referencia mediante Control de Búsqueda Extrema (ESC).

Los resultados en simulación demostraron que el controlador MPC adaptativo es capaz de seguir referencias variables de relajación de manera estable, respetando las restricciones del sistema y minimizando el error de seguimiento, incluso ante variaciones en la dinámica del modelo identificado. La identificación ARX recursiva permitió capturar cambios temporales en la respuesta

del sistema, proporcionando modelos adecuados para el control en línea.

En las pruebas experimentales, el sistema propuesto logró un incremento promedio del 11.8% en el Tiempo en Rango (TIR) en estado de relajación durante la fase de control activo, en comparación con las condiciones de música abierta y la ausencia de estímulo musical. Estos resultados evidencian que la modulación musical adaptativa, guiada por un esquema de control en lazo cerrado, puede generar mejoras sostenidas en la regulación del estrés bajo condiciones de estrés inducido.

El análisis por sujeto reveló una alta variabilidad interindividual, característica inherente de los sistemas biológicos. El desempeño del controlador dependió fuertemente de la controlabilidad fisiológica del sujeto y de la calidad de la identificación del modelo, lo que resalta la importancia de enfoques adaptativos y personalizados. En sujetos con alta adaptabilidad fisiológica, el sistema alcanzó estados prolongados de relajación, mientras que en sujetos con baja controlabilidad, el controlador mantuvo estabilidad sin inducir efectos adversos.

Entre las principales limitaciones del estudio se encuentran la dependencia del desempeño del modelo de identificación y la ausencia de mecanismos explícitos para detectar sujetos con baja controlabilidad fisiológica. Como trabajo futuro, se propone la incorporación de estrategias de detección temprana de controlabilidad, así como la extensión del sistema a múltiples variables musicales y métricas fisiológicas adicionales, con el fin de mejorar la robustez y generalidad del enfoque.

REFERENCIAS

[1].F. Shaffer and J. P. Ginsberg, “An Overview of Heart Rate Variability Metrics and Norms,” Frontiers in Public Health, vol. 5, no. 258, pp. 1–17, Sep. 2017.

[2].P. M. Lehrer, E. Vaschillo, and B. Vaschillo, “Resonant Frequency Biofeedback Training to Increase Cardiac Variability: Rationale and Manual for Training,” Applied Psychophysiology and Biofeedback, vol. 25, no. 3, pp. 177–191, Sep. 2000.

[3].N. Yehuda, “Music and Stress,” Journal of Adult Development, vol. 18, no. 2, pp. 85–94, Jun. 2011.

[4].M. V. Thoma, R. La Marca, R. Brönnimann, L. Finkel, U. Ehlert, and U. M. Nater, “The Effect of Music on the Human Stress Response,” PLoS ONE, vol. 8, no. 8, Art. no. e70156, Aug. 2013.

[5].W. E. J. Knight and N. S. Rickard, “Relaxing music prevents stressinduced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females,” Journal of Music Therapy, vol. 38, no. 4, pp. 254–272, 2001.

[6].J. Du, P. Shi, F. Fang, and H. Yu, “Effect of music intervention on subjective scores, heart rate variability, and prefrontal hemodynamics in patients with chronic pain,” Frontiers in Human Neuroscience, vol. 16, p. 1057290, 2022.

[7].A. Jiménez-Ocaña, A. Pantoja, M. A. Valderrama, and L. F. Giraldo, “A Systematic Review of Technology-Aided Stress Management Systems: Automatic Measurement, Detection and Control,” IEEE Access, vol. 11, pp. 116109–116126, 2023, doi: 10.1109/ACCESS.2023.3325763.

[8].A. A. Jiménez-Ocaña, “Stress regulation by an adaptive closed-loop system using physiological signal analysis with machine learning,” Ph.D. dissertation, Universidad de los Andes, Bogotá, Colombia, 2026.

[9].S. Can, N. Chalabianloo, D. Ekiz, and A. Ersoy, “Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study,” Sensors, vol. 19, no. 8, pp. 1849, 2019.

[10].A. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, “Continuous stress detection using a wrist device: In laboratory and real life,” Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1185–1193, 2016.

[11].A. Jiménez-Ocaña, A. Pantoja, R. Bailón, and L. F. Giraldo, “Real-Time Stress Detection Using Single-Lead ECG Signal Analysis,” in Proc. 3rd Int. Congress of Biomedical Engineering and Bioengineering (CIIBBI), Cali, Colombia, 2024, pp. 1–6, doi: 10.1109/CIIBBI63846.2024.10784783.

[12].M. Krstić and H.-H. Wang, “Stability of extremum seeking feedback for general nonlinear dynamic systems,” Automatica, vol. 36, no. 4, pp. 595–601, 2000.

[13].M. Krstić, “Extremum seeking control,” in Encyclopedia of Systems and Control, Springer, 2015.

[14].J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, “A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health,” Neuroscience & Biobehavioral Reviews, vol. 36, no. 2, pp. 747–756, 2012.

[15].Z. Visnovcova, M. Mestanik, M. Javorka, D. Mokra, M. Gala, A. Jurko, et al., “Complexity and time asymmetry of heart rate variability are altered in acute mental stress,” Physiological Measurement, vol. 35, no. 7, pp. 1319–1334, 2014.

[16].F. Stetter and P. Kupper, “Autogenic training: A meta-analysis of clinical outcome studies,” Applied Psychophysiology and Biofeedback, vol. 27, no. 1, pp. 45–98, Mar. 2002.

[17].J. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[18].A. de Santos-Sierra, C. Ávila, J. Guerra-Casanova, and G. del Pozo, “Real-time stress detection by means of physiological signals,” in Proc. International Conference on Recent

Advances in Biometrics, 2011.

[19].T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794.

[20].S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," Journal of Health and Social Behavior, vol. 24, no. 4, pp. 385–396, 1983.

[21].A. P. Allen et al., "Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test," Neuroscience & Biobehavioral Reviews, vol. 38, pp. 94–124, 2014.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

Estimación de la generación fotovoltaica mediante modelos lineales basados en datos satelitales: estudio comparativo en Pasto, Colombia

Photovoltaic Generation Estimation Using Linear Models Based on Satellite Data: A Comparative Study in Pasto, Colombia

Didier F. Villaquiran Samboni [1]*, Javier Revelo Fuelagán [2]

[1] Departamento de Electrónica, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia. didier.v@udenar.edu.co

[2] Departamento de Electrónica, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia. javierrevelof@udenar.edu.co

* Autor de correspondencia: didier.v@udenar.edu.co

Citación: D. F. Villaquiran Samboni y J. Revelo Fuelagán, "Estimación de la generación fotovoltaica mediante modelos lineales basados en datos satelitales: estudio comparativo en Pasto, Colombia," Revista de Ingeniería, vol. 4, no. 1, pp. 76-87, Nov. 2026.

Resumen—En este trabajo se presenta un análisis comparativo de dos modelos lineales para la estimación de la energía fotovoltaica generada en la ciudad de San Juan de Pasto, departamento de Nariño, Colombia, utilizando datos satelitales de irradiancia tomados de la plataforma SOLCAST. El estudio se realiza sobre dos sistemas solares fotovoltaicos (SSFV) on-grid implementados: el SSFV CEDENAR (de 12,3 kW) y el SSFV UDENAR (de 16,5 kW). El primer modelo, basado en un factor de eficiencia global aplicado sobre la irradiancia global horizontal (GHI), estima la energía generada con un único parámetro de ajuste, mientras que, para el segundo modelo, se aplica una regresión lineal multivariable que incorpora variables ambientales relevantes seleccionadas mediante la magnitud de la correlación de Spearman ($|\rho| \geq 0{,}25$). Los resultados muestran eficiencias globales óptimas de $\eta = 76{,}81$ % y $\eta = 73{,}04$ % para CEDENAR y UDENAR respectivamente, con errores acumulados de energía total inferiores al 2,1 %. El modelo multivariable aporta mejoras marginales en R^2 de +1,52 y +2,07 puntos porcentuales. Adicionalmente, la comparación de las variables ambientales estimadas por SOLCAST con las mediciones de la estación meteorológica local del SSFV UDENAR revela que únicamente la irradiancia ($\rho = 0{,}60$) y la velocidad del viento ($\rho = 0{,}56$) presentan correlaciones significativas entre ambas fuentes, lo que limita la interpretación causal del modelo lineal multivariable. Se concluye que el modelo simple basado en GHI constituye una herramienta práctica y confiable para estimaciones de generación a largo plazo con una eficiencia global promedio de 75 % para la ciudad de Pasto.

Palabras claves—Eficiencia global, Generación de energía solar, Irradiancia global horizontal, Modelo lineal multivariable, Sistema fotovoltaico, SOLCAST.

Abstract—This paper presents a comparative analysis of two linear models for estimating photovoltaic energy generation in the city of San Juan de Pasto, Nariño, Colombia, using satellite irradiance data from the SOLCAST platform. The study was conducted on two implemented on-grid solar photovoltaic systems (PVSS): the CEDENAR PVSS (12.3 kW) and the UDENAR PVSS (16.5 kW). The first model, based on a global efficiency factor applied to global horizontal irradiance (GHI), estimates generated energy using a single fitting parameter, whereas the second model applies multivariable linear regression incorporating relevant environmental variables selected based on the magnitude of Spearman's correlation ($|\rho| \geq 0{,}25$). Results show optimal global efficiencies of $\eta = 76{,}81$ % and $\eta = 73{,}04$ % for CEDENAR and UDENAR, respectively, with cumulative total energy errors below 2.1%. The multivariable model yields marginal improvements in R^2 of +1.52 and +2.07 percentage points. Additionally, comparing SOLCAST-estimated environmental variables with local weather station measurements from the UDENAR PVSS reveals that only irradiance ($\rho = 0{,}60$) and wind speed ($\rho = 0{,}56$) exhibit significant correlations between sources, limiting the causal interpretation of the multivariable linear model. It is concluded that the simple GHI-based model serves as a practical and reliable tool for long-term generation estimates, with an average global efficiency of 75% for the city of Pasto.

Keywords—Global efficiency, Global horizontal irradiance, Multivariable linear model, Photovoltaic system, Solar energy generation, SOLCAST.

I. INTRODUCCIÓN

La energía solar fotovoltaica se ha consolidado como una de las tecnologías de mayor crecimiento dentro del sector de las energías renovables a nivel global. La capacidad instalada de plantas fotovoltaicas pasó de 0.23 GW en 1996 a 2.25 TW en 2024, con tasas de crecimiento anual superiores al 20 % [1]. Este crecimiento ha sido impulsado por la reducción de costos de los módulos, el aumento en la eficiencia de conversión y los marcos regulatorios que favorecen la integración de fuentes renovables [1] [2]. En Colombia, la Ley 1715 de 2014 estableció incentivos para la integración de fuentes no convencionales de energía renovable (FNCER), promoviendo la instalación de sistemas fotovoltaicos conectados a red (On-Grid) tanto a escala residencial como institucional [3].

En desarrollo de esta ley, la Resolución CREG 030 de 2018 reguló la autogeneración a pequeña escala y la generación distribuida, y la Resolución CREG 174 de 2021 actualizó dicho marco estableciendo el esquema de créditos de energía para el reconocimiento de excedentes inyectados al sistema.

La estimación precisa de la generación eléctrica de un SSFV es fundamental para evaluar la viabilidad técnica y económica de proyectos solares, planificar la integración con la red eléctrica y verificar el desempeño operativo de las instalaciones [4] [5]. La producción real de un SFV depende de múltiples factores: la irradiancia incidente, la temperatura de operación de las celdas, las pérdidas en cableado e inversor, la suciedad acumulada y las condiciones atmosféricas locales como nubosidad, humedad relativa, régimen de vientos, entre otras [6] [7].

Para la estimación de la generación fotovoltaica se han desarrollado diversos enfoques metodológicos. Los modelos basados en regresión lineal, tanto simple como múltiple, han demostrado ser herramientas efectivas y de bajo costo computacional [8] [5] [6]. En los estudios [8] y [5] se emplean modelos de regresión múltiple para la predicción de potencia fotovoltaica, encontrando que variables como la irradiancia, la temperatura y la velocidad del viento son los predictores más significativos. Conjuntamente, técnicas de aprendizaje automático como redes neuronales artificiales (MLP), XGBoost y modelos híbridos CNN-LSTM han sido aplicadas para mejorar la precisión de la estimación, aunque con incrementos marginales respecto a modelos lineales cuando la relación entre las variables es predominantemente lineal [9] [10] [11].

En la mayoría de las instalaciones fotovoltaicas no se dispone de estaciones meteorológicas locales, por lo que los modelos de estimación recurren a datos de irradiancia provenientes de plataformas satelitales. SOLCAST es un servicio de modelamiento satelital que proporciona series temporales de irradiancia global horizontal (GHI) y otras variables ambientales para cualquier punto geográfico, con resoluciones temporales desde 5 hasta 60 minutos [12]. En el estudio de validación realizado en [12] se ha demostrado que SOLCAST presenta un buen desempeño global, con un sesgo medio (MBE) del 0.33 % y una desviación estándar de sesgos del 1.7 % en 48 estaciones BSRN. No obstante, comparaciones recientes en regiones tropicales han evidenciado que la precisión varía según las condiciones climáticas y geográficas locales [10]. Sin embargo, estudios de adaptación al sitio para la ciudad de Salta realizados en [13] [7] han mostrado que las estimaciones de SOLCAST presentan un RMSE del 23.6 %, requiriendo procedimientos de corrección para mejorar su representatividad local.

La confiabilidad de las bases de datos satelitales para la estimación de generación fotovoltaica ha sido objeto de múltiples estudios comparativos. En [14] se compararon bases de datos de NASA, SOLCAST y PVGIS contra mediciones terrestres en Irlanda, encontrando que las mediciones en tierra ofrecen mayor precisión espacial. En [10] se evaluaron SOLCAST, NASA POWER y Open-Meteo en cuatro sitios tropicales mediante XGBoost, concluyendo que SOLCAST presenta el mejor desempeño general con un R^2 promedio de 0.91. En [11] evaluaron tres modelos de aprendizaje automático para adaptación al sitio en el noroeste argentino, encontrando que modelos simples como la regresión lineal ofrecen resultados comparables a modelos más complejos, sugiriendo que la calidad de los datos de entrada impone un límite superior a las mejoras alcanzables.

En el contexto colombiano, el estudio abordado en [15] muestra que ya se han realizado estudios sobre la predicción de la radiación solar global en el departamento de Nariño, contribuyendo al conocimiento del recurso solar en esta región andina ecuatorial. En [16] se desarrolló un modelo de estimación de potencia generada por un SFV a partir de variables de bases de datos de recurso energético solar, evaluando la relación entre variables satelitales y la generación real. Sin embargo, en la revisión bibliográfica realizada, no se han reportado estudios que comparen modelos lineales de estimación de generación fotovoltaica aplicados sobre datos satelitales de SOLCAST en la región de Pasto, ni la evaluación de correspondencia entre las variables ambientales estimadas por esta plataforma y mediciones meteorológicas reales en un punto geográfico específico para un SSFV determinado.

La ciudad de San Juan de Pasto, presenta condiciones climáticas particulares: su proximidad a la línea ecuatorial garantiza una variación mínima del ángulo de incidencia solar a lo largo del año, pero la elevada nubosidad y precipitación características de la zona andina ecuatorial introducen una alta variabilidad intradía en la irradiancia disponible [15]. Estas condiciones hacen que

la estimación de generación basada exclusivamente en datos satelitales requiera una validación rigurosa con mediciones locales.

En este contexto, el presente trabajo tiene como objetivo comparar el desempeño de dos modelos lineales para la estimación de la energía fotovoltaica generada en la ciudad de Pasto: un modelo de eficiencia global basado en GHI y un modelo de regresión lineal multivariable que incorpora variables ambientales adicionales. El estudio se aplica sobre dos sistemas fotovoltaicos conectados a red (SSFV CEDENAR de 12.3 kW y SSFV UDENAR de 16.5 kW), utilizando datos de irradiancia de la plataforma SOLCAST. Adicionalmente, se evalúa la correspondencia entre las variables ambientales estimadas por SOLCAST y las mediciones de la estación meteorológica instalada en el SSFV UDENAR, con el fin de determinar la confiabilidad de los datos satelitales como insumo para modelos de estimación en esta región geográfica.

II. METODOLOGÍA

La metodología para la definición de los parámetros en los modelos lineales de estimación de potencia generada se presenta en la Fig.1. Dado que se tienen tres fuentes de datos diferentes, correspondientes a i) datos estimados a partir de imágenes satelitales obtenidos a través de SOLCAST, ii) variables ambientales medidas por una estación meteorológica instalada en el bloque 1 de la Universidad de Nariño y iii) la potencia activa en AC generada por los SSFV. Al cruzar las variables disponibles, se cuenta con dos conjuntos de datos para el ajuste de los modelos: uno que relaciona las variables ambientales de SOLCAST con la potencia AC medida, y otro que relaciona las variables ambientales de la estación meteorológica con esa misma potencia AC. Para cada conjunto de datos se realiza un proceso de limpieza y selección de variables, en donde de SOLCAST se extraen la irradiancia, temperatura, humedad relativa, velocidad y dirección del viento y precipitaciones, de la estación meteorológica se extraen las mismas variables medidas localmente, y de la plataforma de monitoreo Modelo Transaccional de Energía (MTE) se extrae la potencia activa AC como variable objetivo.

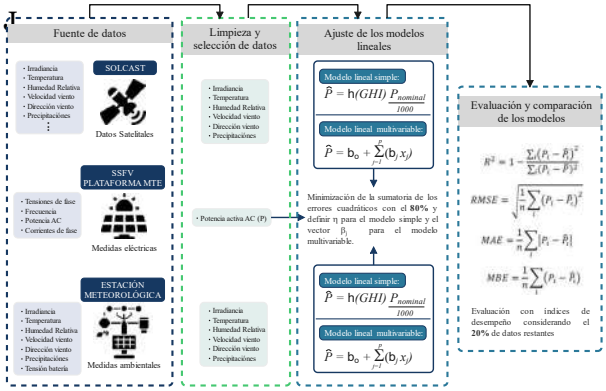


Figura 1. Diagrama metodológico para la definición de los modelos de estimación de potencia fotovoltaica generada.

Una vez preparados los datos, se definen los dos modelos lineales a comparar. El primero es un modelo lineal simple que estima la potencia generada a partir exclusivamente de la irradiancia global horizontal, escalada por la potencia nominal del sistema y ponderada por un factor de eficiencia global que condensa en un único parámetro el conjunto de pérdidas inherentes al sistema, incluyendo las pérdidas por temperatura de operación, eficiencia del inversor, pérdidas resistivas en cableado, suciedad acumulada y la diferencia angular entre la irradiancia horizontal y la que efectivamente incide sobre el plano de los módulos. El segundo es un modelo de regresión lineal multivariable que, además de la irradiancia, incorpora las variables ambientales que presentan una correlación de Spearman superior a 0,25 con la potencia AC, con el fin de capturar efectos de segundo orden asociados a las condiciones meteorológicas. Los parámetros de ambos modelos se ajustan mediante la minimización de la suma de errores cuadráticos utilizando el 80 % de los datos disponibles. Finalmente, el desempeño de los modelos se evalúa sobre el 20 % restante de los datos mediante cuatro métricas: el coeficiente de determinación R^2, la raíz del error cuadrático medio RMSE, el error absoluto medio MAE y el sesgo medio del error MBE, lo que permite una comparación cuantitativa entre ambos enfoques de estimación.

II-A. Descripción de los sistemas fotovoltaicos

En la Fig.2 se presenta la ubicación geográfica del SSFV ubicado en la subestación de Pasto de CEDENAR y las referencias del inversor y los paneles que componen al SSFV CEDENAR.



Figura 2. Sistema fotovoltaico On-Grid CEDENAR de 12.3 kW, instalado en la subestación CEDENAR Pasto.

El SSFV CEDENAR se encuentra instalado en uno de los bloques de la subestación CEDENAR, ubicada en la ciudad de Pasto, departamento de Nariño, Colombia. El punto geográfico de instalación corresponde a las coordenadas 1.22727° de latitud norte y -77.28195° de longitud oeste.

El SSFV CEDENAR se compone de un inversor Huawei SUN2000-20KTL-M3, con una potencia nominal de 20 kW. Este equipo cuenta con cuatro entradas MPPT (Maximum Power Point Tracking) independientes y una salida de corriente alterna (AC) trifásica. Cuenta con 20 módulos de 615 W cada uno, distribuidos en dos arreglos de 10 módulos conectados en serie. Cada uno de estos arreglos se encuentra conectado a una entrada MPPT independiente del inversor.

La tensión del panel solar en el punto de máxima potencia (Vmp) es de 40,73 V y una potencia máxima de 615 W bajo condiciones estándar de prueba para cada módulo. Al estar conectados 10 módulos en serie en cada arreglo, se obtiene una tensión máxima de 407,3 V por string, valor que se encuentra dentro del rango de operación admitido por el inversor. En conjunto, los 20 módulos conforman una capacidad instalada total de 12,3 kWp, mientras que el inversor presenta una potencia nominal de 20 kW. Esta configuración resulta en una relación DC/AC de aproximadamente 0,615, indicando que el inversor opera con un margen amplio respecto a la potencia del arreglo fotovoltaico.

En la Fig.3 se presenta la ubicación geográfica correspondiente a la ubicación de los paneles fotovoltaicos en el bloque 1 de la Universidad de Nariño sede Torobajo y las referencias del inversor y los paneles que componen al SSFV UDENAR.



Figura 3. Sistema fotovoltaico On-Grid UDENAR de 16.5kW, instalado en el bloque 1 de la Universidad de Nariño.

El punto geográfico de instalación corresponde a las coordenadas 1,233242° de latitud norte y -77,2929° de longitud oeste. El SSFV UDENAR se compone de un inversor FRONIUS SYMO 15.0-3 208, con una potencia nominal de 15 kW, que cuenta con dos entradas asociadas a módulos MPPT independientes y una salida de corriente alterna (AC) trifásica. Cuenta con 30 módulos de 550 W STC cada uno, que conforman una capacidad instalada total de 16.5 kWp, mientras que el inversor presenta una potencia pico admisible entre 12 kW y 19.5 kW. Esta configuración resulta en una relación DC/AC de aproximadamente 0.8461, un margen de expansión inferior respecto al SSFV CEDENAR.

II-B. Modelo de estimación lineal simple

El modelo lineal simple estima la potencia activa generada por el SSFV a partir de la irradiancia global horizontal y la potencia pico instalada Ppi, aplicando un factor de eficiencia global η que representa la fracción de la energía teórica que efectivamente se convierte en energía AC inyectada a la red. La potencia estimada para cada instante de tiempo se define como se muestra en la Ec.1

$$\hat{P} = \eta(GHI)\frac{P_{pi}}{1000W/m^2} \tag{1}$$

Donde $\hat{P}$ es la potencia estimada en [W], GHI es la irradiancia global horizontal en [W/m²], P_{pi} es la potencia instalada del sistema en [W] (12300 W para el SSFV CEDENAR y 16500 W para el SSFV UDENAR) y el factor 1000 corresponde a la irradiancia en condiciones estándar de prueba (1000 W/m²). El parámetro η agrupa todas las pérdidas relacionadas, incluyendo las pérdidas por temperatura de operación de las celdas, la eficiencia de conversión del inversor, las pérdidas resistivas en el cableado DC y AC, la suciedad acumulada sobre los módulos, el desajuste eléctrico entre módulos y la diferencia angular entre la irradiancia horizontal y la que efectivamente recibe el plano inclinado de los paneles.

Para determinar el valor óptimo de η se formula el problema de optimización que minimiza la suma de errores cuadráticos entre la energía diaria real medida por el inversor y la energía diaria estimada como se muestra en la 2.

$$FO(\eta) = \sum_{i=1}^{N} \left(E_{real,i} - \eta E_{base,i}\right)^2 \tag{2}$$

$$\frac{dFO(\eta)}{d\eta} = 0 \Rightarrow \eta_{opt} = \frac{\sum_{i=1}^{N}(E_{real,i}E_{base,i})}{\sum_{i=1}^{N}(E_{base,i})^2} \tag{3}$$

Dado que la Ec.2 es una funcion cuadrática convexa de una variable, tiene un único mínimo, aplicando el criterio de la derivada de primer orden se obtiene el valor óptimo de forma analítica a través de la Ec.3, en donde $E_{base,i}$ es la energía teórica diaria calculada a partir de los registros de GHI, N es el número de días con datos válidos en ambas fuentes dentro del conjunto de entrenamiento y $E_{base,i}$ es la energía generada por el SSFV medida de la plataforma MTE.

II-C. Modelo de estimación lineal multivariable

El segundo modelo busca mejorar la estimación de la potencia generada incorporando, además de la GHI, otras variables ambientales que aporten información complementaria. Se propone el modelo lineal multivariable mostrado en la Ec.4

$$\hat{P} = \beta_o + \sum_{j=1}^{p} \beta_j x_j \quad (4)$$

donde x_j representa cada una de las p variables ambientales seleccionadas, β_j es el coeficiente de ponderación asociado a cada variable y β_o es el intercepto del modelo. A diferencia del modelo de eficiencia global que opera a escala diaria comparando energías acumuladas, este modelo trabaja a resolución temporal de 10 minutos, estimando la potencia instantánea a partir de las condiciones ambientales de cada intervalo para poder capturar la variación intradía de las variables ambientales más relevantes.

La selección de las variables predictoras se realizó mediante la correlación de Spearman entre cada variable ambiental y la potencia AC medida. Se incluyeron en el modelo aquellas variables cuya magnitud del coeficiente de correlación satisface $|\rho| \geq 0{,}25$. La correlación de Spearman fue elegida por su robustez ante relaciones no lineales y distribuciones no normales. Este criterio resultó en la selección de 12 variables para el SSFV CEDENAR y 8 variables para el SSFV UDENAR.

El vector de coeficientes $\beta = [\beta_0, \beta_1, \ldots, \beta_p]^T$ de la Ec.5 se obtiene al minimizar la suma de los residuos al cuadrado sobre las n observaciones del conjunto de entrenamiento a través del método de mínimos cuadrados ordinarios (OLS).

$$\beta = \arg\min \sum_{i=1}^{n} (P_i - \hat{P}_i)^2 = \arg\min \sum_{i=1}^{n} \left(P_i - \beta_0 - \sum_{j=1}^{p} \beta_j \cdot x_{ij} \right)^2 \quad (5)$$

La solucion analítica en notacion matricial es:

$$\beta = (X^T X)^{-1} X^T y \quad (6)$$

En donde X es la matriz de diseño de dimensión $n(p + 1)$, cuya primera columna es un vector de unos para el intercepto y las columnas restantes contienen los valores de las p variables seleccionadas, e y es el vector de potencias reales medidas por el inversor.

Dado que las variables seleccionadas tienen unidades y escalas diferentes (la GHI en W/m², la temperatura en °C, la humedad en porcentaje), los coeficientes β_j no son directamente comparables entre sí. Para determinar la importancia relativa de cada variable se estimó el mismo modelo sobre las variables estandarizadas (restando el valor medio y dividiendo sobre la desviación estándar de cada variable).

II-D. Métricas de evaluación

El desempeño de ambos modelos se evaluó sobre el conjunto de prueba independiente (20 % de los datos) mediante las siguientes métricas:

El coeficiente de determinación R^2:

$$R^2 = 1 - \frac{\sum_i (P_i - \hat{P}_i)^2}{\sum_i (P_i - \bar{P})^2} \quad (7)$$

La raíz del error cuadrático medio RMSE expresa el error típico del modelo en [W], penalizando con mayor peso los errores de gran magnitud:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_i (P_i - \hat{P}_i)^2} \quad (8)$$

El error absoluto medio MAE calcula el promedio de los errores absolutos, tratando todas las desviaciones con igual peso:

$$\mathrm{MAE} = \frac{1}{n} \sum_i |P_i - \hat{P}_i| \quad (9)$$

El sesgo medio del error MBE es la unica métrica que conserva el signo, permitiendo identificar si el modelo tiende a sobreestimar (MBE negativo) o subestimar (MBE positivo) la potencia:

$$\mathrm{MBE} = \frac{1}{n} \sum_i (P_i - \hat{P}_i) \quad (10)$$

donde Pi es la potencia real medida por el inversor en el instante i, $\hat{P}_i$ es la potencia estimada por el modelo, $\bar{P}$ es el promedio de las potencias reales y n es el número total de observaciones en el conjunto de prueba. Adicionalmente, se comparó la energía mensual estimada por cada modelo con la energía mensual medida, calculando el error porcentual para cada mes y el error total acumulado en todo el período de análisis.

III. RESULTADOS

Para el SSFV UDENAR se cuenta con datos de las tres fuentes disponibles (SOLCAST, Estación meteorológica y mediciones de potencia activa AC generada) desde el 01 de julio de 2023 hasta el 31 de diciembre de 2023, mientras que para el SSFV CEDENAR se dispone con datos de 2 de las tres fuentes debido a la ausencia de registros medidos mediante estación meteorológica en este sitio, correspondientes a los datos de SOLCAST y las mediciones de la potencia activa AC generada desde el 01 de febrero de 2026 hasta el 10 de agosto de 2026.

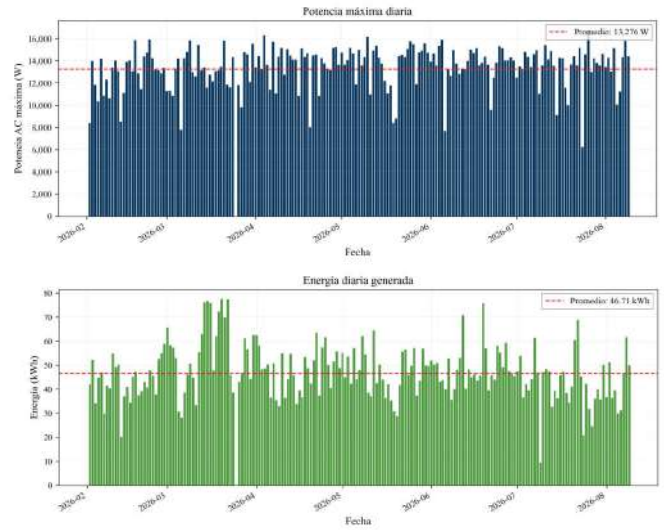


Figura 4. Potencia máxima y energía diaria generada por el SSFV CEDENAR.

En la Fig.4 se presenta la potencia máxima reportada en cada día de operación para el periodo analizado en el SSFV CEDENAR, en donde se aprecia que, la potencia máxima que se puede presentar en un día en promedio es de 13.28 kW y la energía total producida en un día promedio es de 46.71 kWh. Dada la potencia nominal del SSFV CEDENAR, las horas efectivas de irradiancia en un día promedio es de 3.80h.

Por otra parte, en la Fig.5 se aprecia que la potencia máxima que se puede presentar en un día en promedio para el SSFV UDENAR es de 13.62 kW y la energía total producida en un día promedio es de 58.55 kWh. Dada la potencia nominal del SSFV UDENAR, las horas efectivas de irradiancia en un día promedio es de 3.55h.

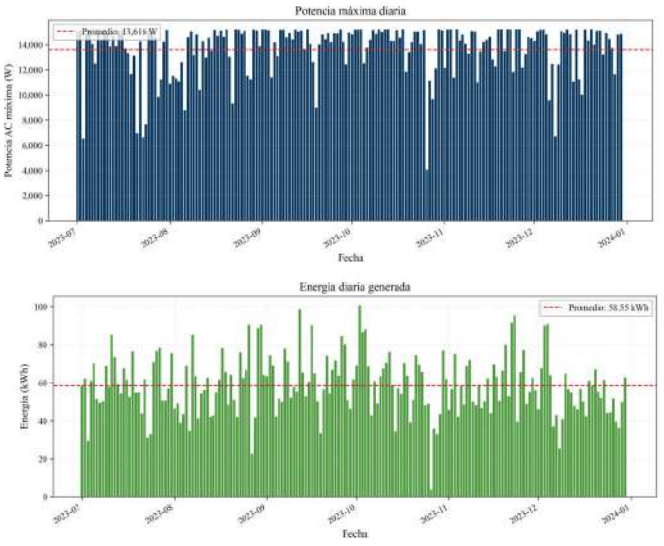


Figura 5. Potencia máxima y energía diaria generada por el SSFV UDENAR.

Mientras la potencia máxima del SSFV CEDENAR supera la potencia pico instalada en algunos días de operación, en el SSFV UDENAR nunca se supera la potencia instalada, este fenómeno se atribuye a los ajustes del inversor Fronius instalado en el SSFV UDENAR, que por criterios de ajuste de factor de potencia o por seguridad en las inyecciones de potencia, limita la inyección de potencia activa por encima de la potencia pico instalada, a pesar de que en horas de mayor irradiancia sí fuera posible superar dicha potencia generada como se aprecia en la Fig.6.

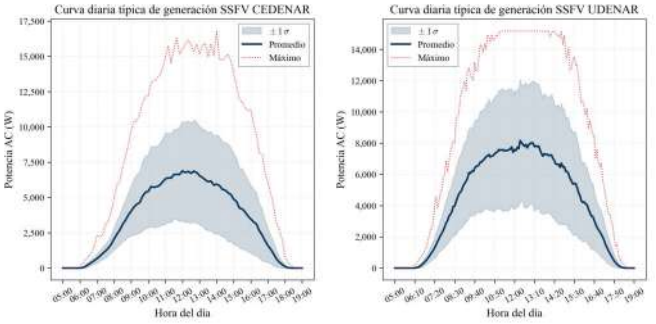


Figura 6. Curvas diarias típicas de potencia activa promedio y máxima generada por los sistemas SSFV CEDENAR y SSFV UDENAR.

Como se muestra en la Fig.6, la curva típica de los valores máximos de potencia activa del SSFV UDENAR detectados en todo el periodo de análisis se saturan en 16.5 kW, mientras que en el SSFV CEDENAR la curva típica de potencias máximas sigue un comportamiento más natural. En la Fig.7 se presentan los valores óptimos de la eficiencia η para el modelo lineal simple en los dos SSFV.

Dado que la raíz y la división entre N son transformaciones estrictamente crecientes y constantes, el valor de η que minimiza el RMSE es exactamente el mismo que minimiza la suma de los errores cuadráticos definida como función objetivo del problema de optimización en la Ec.2.

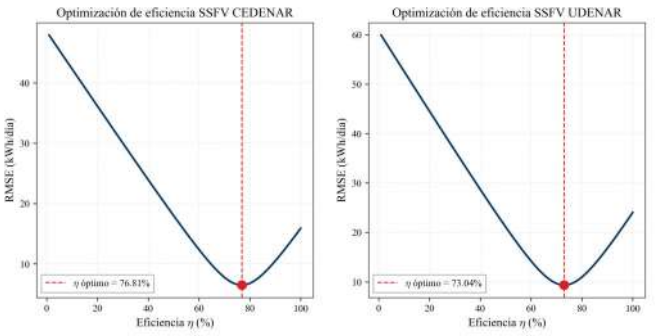


Figura 7. Eficiencia global óptima para el modelo lineal simple respecto a la raíz del error cuadrático promedio.

III-A. Análisis de correlación con la potencia generada

Para definir las variables más relevantes a incluir en el modelo lineal multivariable se evalúan las magnitudes de los coeficientes de correlación de Spearman, En la Fig.8 se presenta la matriz de correlación de la potencia generada con las variables ambientales tomadas de SOLCAST para el SSFV CEDENAR:

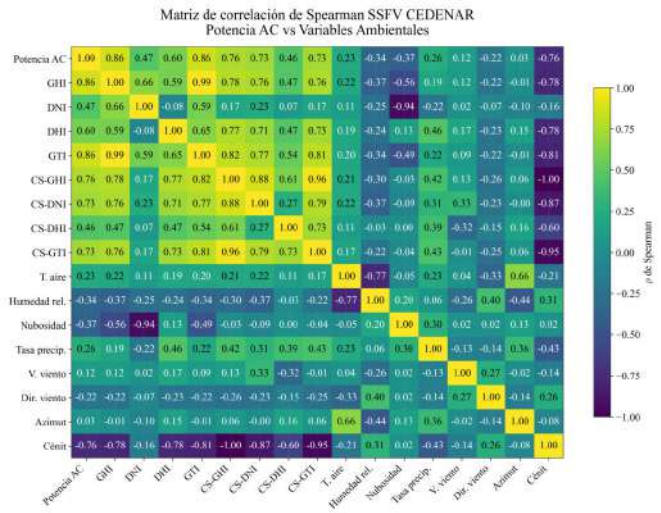


Figura 8. Matriz de correlación de Spearman de la potencia AC generada medida por el SSFV y las variables ambientales definidas por SOLCAST para el SSFV CEDENAR.

La matriz de correlación de Spearman muestra que la potencia AC generada está principalmente asociada con las variables de irradiancia solar. La GHI y la GTI presentan las correlaciones positivas más altas (ρ=0,86 en ambos casos), lo que confirma que la disponibilidad de radiación sobre el plano horizontal y sobre la superficie de los módulos constituyen las variables principales para la producción fotovoltaica. La DHI y la DNI también presentan correlaciones positivas con la potencia generada, mientras que la nubosidad, la humedad relativa, la precipitación y el ángulo cenital presentan relaciones negativas con la potencia generada por el SSFV CEDENAR.

En la Fig.9 se presenta la matriz de correlación de la potencia generada con las variables ambientales tomadas de SOLCAST para el SSFV UDENAR.

La matriz de correlación de Spearman muestra que la potencia AC está principalmente asociada, nuevamente, con las variables de irradiancia solar. La GHI y la GTI presentan las correlaciones positivas más altas (ρ=0,86 y ρ=0,85 respectivamente), lo que confirma que la disponibilidad de radiación sobre el plano horizontal y sobre la superficie de los módulos son las variables más representativas para la generación fotovoltaica. La DHI y la DNI también presentan asociaciones positivas, mientras que la nubosidad, la humedad relativa, la velocidad y la dirección del viento presentan relaciones negativas con la potencia generada. En la matriz de correlación de la Fig.9 se destaca la correlación con la temperatura del aire y la humedad relativa; mientras estas variables en el SSFV CEDENAR tenían una correlación baja, ahora en el SSFV UDENAR presentan una correlación significativa con la potencia generada.

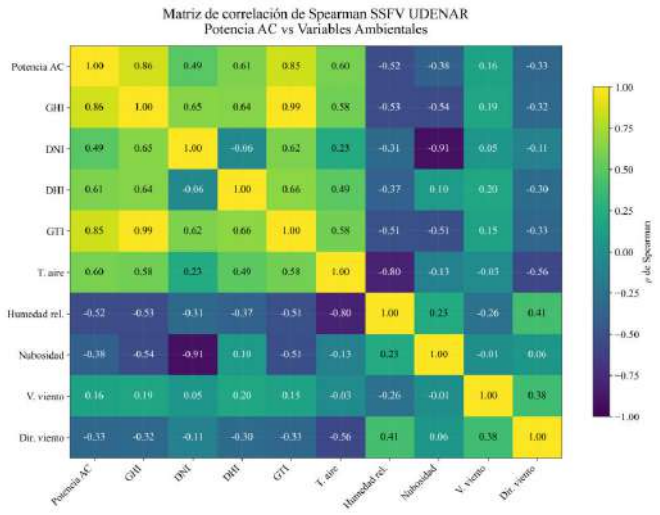


Figura 9. Matriz de correlación de Spearman de la potencia AC generada medida por el SSFV UDENAR y las variables ambientales definidas por SOLCAST.

En la Tabla I se presentan las variables que satisfacen el criterio en el coeficiente de correlación $|\rho| \geq 0{,}25$ utilizadas en el modelo lineal multivariable del SSFV CEDENAR.

TABLA I
COEFICIENTES DE PONDERACIÓN PARA EL MODELO LINEAL MULTIVARIABLE CON LAS VARIABLES MÁS SIGNIFICATIVAS DEL SSFV CEDENAR.

VARIABLE	ρ	β	β_ Estandarizado
GHI	**+0,8631**	**14,4756**	**3709,4354**
GTI	0,8589	-3,9864	-885,7909
CS-GHI	0,7644	-0,807	-304,021
Cénit	-0,7606	-31,652	-749,7469
CS-GTI	0,7328	-0,4051	-99,7087
CS-DNI	0,7259	-0,2363	-55,1743
DHI	0,596	-2,0536	-312,1396
DNI	0,4712	-1,9171	-516,0562
CS-DHI	0,4636	-0,0417	-14,8464
Nubosidad	-0,3673	1,5651	42,9732
Humedad R.	-0,3425	-3,9914	-40,2448
Tasa precipitación	0,2591	245,707	138,6733

Como se aprecia en la Tabla I, la irradiancia global horizontal (GHI) es la variable dominante del modelo, con el mayor coeficiente estandarizado (3709.43). Esto es coherente con su elevada correlación de Spearman (ρ = +0,86) y con la física del proceso de conversión fotovoltaica, donde la irradiancia es la variable principal de generación.

Un aspecto que requiere interpretación cuidadosa es la presencia de coeficientes β negativos en variables que tienen correlación positiva con la potencia, como GTI (ρ = +0,86, β = −3,99), DNI (ρ = +0,47, β = −1,92) y DHI (ρ = +0,60, β = −2,05). Este comportamiento no implica que mayor irradiancia reduzca la generación, sino que es consecuencia de la multicolinealidad entre las variables de radiación solar. GHI, GTI, DNI, DHI y sus respectivas componentes de cielo despejado miden aspectos relacionados de la misma magnitud física, por lo que cuando el modelo ya cuenta con GHI como variable principal, las demás componentes actúan como términos correctivos que ajustan la estimación ante condiciones donde la relación simple con GHI no es suficiente (cielos parcialmente nublados, ángulos solares bajos, predominancia de radiación difusa).

El ángulo cenital presenta un coeficiente β = −31,65 y un coeficiente estandarizado de -749.75, constituyéndose como la segunda variable más influyente en magnitud absoluta. Su signo negativo es físicamente consistente: a mayor ángulo cenital (sol más bajo en el horizonte), menor es la componente de irradiancia que interceptan los módulos, lo que reduce la potencia generada. La tasa de precipitación muestra el mayor coeficiente β no estandarizado (245.71), aunque su coeficiente estandarizado (138.67) y su correlación de Spearman (ρ = +0,26) son moderados. Esto se explica porque la precipitación tiene una escala numérica pequeña (mm/h), por lo que su coeficiente β se amplifica para compensar la escala, pero su influencia real sobre la varianza del modelo es limitada como se aprecia en el modelo estandarizado. El signo positivo puede parecer contraintuitivo, pero indica que los eventos de precipitación en Pasto tienden a concentrarse en las horas de la tarde, cuando la generación ya ha sido alta durante el día, generando una asociación estadística positiva que no es necesariamente causal.

La humedad relativa (β = −3,99, ρ = −0,34) y la nubosidad (β = +1,57, ρ = −0,37) presentan coeficientes estandarizados pequeños (-40.25 y +42.97 respectivamente), lo que indica una contribución marginal al modelo. El signo positivo de la nubosidad en el coeficiente β, a pesar de su correlación negativa, es otro efecto de la multicolinealidad: el modelo ya captura el efecto de las nubes a través de la diferencia entre la irradiancia medida (GHI) y la de cielo despejado (CS-GHI), dejando que la variable de nubosidad actúe como un ajuste residual.

En la Tabla II se presentan las variables que satisfacen el criterio en el coeficiente de correlación |ρ| ≥ 0,25 utilizadas en el modelo lineal multivariable del SSFV UDENAR:

TABLA II
COEFICIENTES DE PONDERACIÓN PARA EL MODELO LINEAL MULTIVARIABLE CON LAS VARIABLES MÁS SIGNIFICATIVAS DEL SSFV UDENAR.

VARIABLE	ρ	β	β_ Estandarizado
GHI	**0,8622**	**11,8011**	**3167,9699**
GTI	0,8542	-0,6464	-201,5808
DHI	0,6141	0,1984	23,6395
Temperatura	0,5983	212,489	596,6328
Humedad relativa	-0,5248	2,0268	53,4909
DNI	0,4909	-0,0272	-22,5175
Nubosidad	-0,3791	5,731	129,5375
Dir. del viento	-0,3284	2,0218	158,1027

Como se aprecia en la Tabla II, la irradiancia global horizontal (GHI) es la variable dominante del modelo, con el mayor coeficiente estandarizado (3167,97); nuevamente, esto es coherente con el valor del coeficiente de correlación de Spearman (ρ = +0,86). La temperatura del aire es la segunda variable más influyente con un β estandarizado = 596,63 y un β = 212,49. Su signo positivo y su correlación moderada-alta (ρ = +0,60) obedecen a que las horas de mayor temperatura, en general, coinciden con las de mayor irradiancia solar. Si bien es conocido que temperaturas elevadas de celda reducen la eficiencia de conversión fotovoltaica, este efecto térmico negativo queda enmascarado por la covariación temporal entre temperatura e irradiancia: los momentos más cálidos del día son también los de mayor generación.

Como ocurre con el SSFV CEDENAR, en el SSFV UDENAR también se presentan β < 0 con ρ > 0, como en GTI (ρ = +0,85, β = −0,65) y DNI (ρ = +0,49, β = −0,03). Este comportamiento no implica que una mayor irradiancia reduzca la generación, sino que es consecuencia de la multicolinealidad entre las variables de radiación solar. GHI, GTI, DNI y DHI miden aspectos relacionados de la misma magnitud física, por lo que cuando el modelo ya cuenta con GHI como variable principal, las demás componentes actúan como términos correctivos que ajustan la estimación ante condiciones donde la rela-

ción simple con GHI no es suficiente, tales como cielos parcialmente nublados o predominancia de radiación difusa. En particular, GTI presenta el coeficiente estandarizado negativo de mayor magnitud (-201,58), indicando que corrige la sobreestimación que GHI por sí sola produciría cuando la irradiancia en el plano inclinado difiere de la horizontal.

La dirección del viento presenta un β estandarizado = 158,10, situándose como la cuarta variable más influyente en magnitud absoluta. Su correlación negativa (ρ = −0,33) pero coeficiente β positivo (+2,02) sugiere que, en el contexto multivariable, la dirección del viento actúa como un indicador indirecto de las condiciones atmosféricas predominantes: ciertas orientaciones del viento en Pasto se asocian con masas de aire más secas o menor nubosidad, condiciones que favorecen la generación. La nubosidad (β = +5,73, ρ = −0,38) presenta un coeficiente estandarizado de 129,54 con signo positivo a pesar de su correlación negativa con la potencia. Este comportamiento es un efecto de la multicolinealidad: el modelo ya captura el impacto de las nubes a través de la reducción de GHI medida, de modo que la variable de nubosidad actúa como un ajuste residual.

La humedad relativa (β = +2,03, ρ = −0,52) y la DHI (β = +0,20, ρ = +0,61) presentan coeficientes estandarizados pequeños (53,49 y 23,64 respectivamente), lo que indica una contribución marginal al modelo. El signo positivo de la humedad relativa en el coeficiente β, a pesar de su correlación negativa, es otro efecto de la multicolinealidad: una vez que el modelo ha incorporado la GHI, la nubosidad y la temperatura, la humedad relativa aporta información residual sobre las condiciones atmosféricas que puede favorecer ciertos regímenes de irradiancia difusa.

Finalmente, la DNI muestra el menor coeficiente estandarizado en magnitud absoluta (-22,52), indicando que su aporte al modelo es prácticamente nulo una vez que GHI y las demás variables de radiación están presentes. Esto es consistente con las condiciones climáticas, donde la alta nubosidad hace que la componente directa de la radiación sea menos determinante que la global o la difusa para la estimación de la potencia generada.

III-B. Análisis de correlación entre los datos de SOLCAST y las mediciones de la estación meteorológica

Inicialmente se realizó el análisis de las variables ambientales medidas a través de la estación meteorológica y las variables ambientales estimadas por SOLCAST para el SSFV UDENAR. En la Fig.10 se presenta la matriz de correlación de Spearman para las variables ambientales correspondientes:

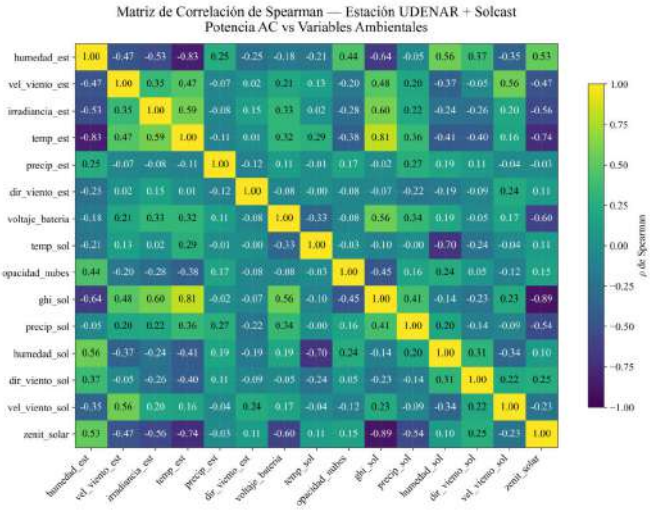


Figura 10. Matriz de correlación de Spearman de las variables ambientales medidas por la estación (sufijo est) y las variables estimadas por SOLCAST (sufijo sol) en el SSFV UDENAR.

Como se puede apreciar en la Fig.10, la correlación entre la irradiancia medida por la estación meteorológica (irradiancia est) y la irradiancia estimada por SOLCAST (ghi sol) es de 0,60, que es la mayor correlación monotónica con las demás variables relacionadas en la matriz. Esto también ocurre con la velocidad del viento medida por la estación meteorológica (vel viento est) y la estimada (vel viento sol), con un coeficiente de correlación de 0,56. Sin embargo, no se evidencia una correlación significativa de las variables humedad relativa, temperatura y dirección del viento. Tampoco se presenta la mayor magnitud de correlación en todas las variables cruzadas de ambas fuentes; por ejemplo, la temperatura medida (temp est) tiene una correlación más significativa con la humedad medida (ρ = −0,83) que con la temperatura estimada por SOLCAST (temp sol).

Aunque la precipitación medida no tiene ninguna correlación significativa con las demás variables, sí presenta su mayor valor con la precipitación estimada por SOLCAST, con ρ = 0,27. Por otra parte, la mayor correlación con la tensión en la batería corresponde al ángulo cenital del sol, lo que es coherente con la ausencia de datos en horas de la madrugada, pues en dichas horas la batería ya no tiene carga (tensión baja) y el ángulo cenital toma valores bajos.

IV. DISCUSIÓN

Luego de evaluar el desempeño de los modelos lineales para la estimación de la generación en los SSFV CEDENAR y UDENAR, se obtuvieron las métricas de desempeño mostradas en la Tabla III y la Tabla IV.

TABLA III
MÉTRICAS DE DESEMPEÑO DE LOS MODELOS PROPUESTOS PARA EL SSFV CEDENAR

MÉTRICA	MODELO SIMPLE	MODELO MULTIVARIABLE
R^2	0,6799	**0,6951**
RMSE [W]	1763,3	1720,8
MAE [W]	1252,5	1215,1
MBE [W]	56,7	-27,5

TABLA IV
MÉTRICAS DE DESEMPEÑO DE LOS MODELOS PROPUESTOS PARA EL SSFV UDENAR

MÉTRICA	MODELO SIMPLE	MODELO MULTIVARIABLE
R^2	0,6761	**0,6967**
RMSE [W]	2209,1	2137,5
MAE [W]	1569,3	1521,5
MBE [W]	142,8	57

Las métricas de desempeño muestran que el modelo multivariable logra una mejora marginal respecto al modelo simple en ambos sistemas. En el SSFV CEDENAR, el R^2 pasa de 0,6799 a 0,6951, representando un incremento de 1,52 %, con una reducción del RMSE del 2,4 % (de 1763,3 W a 1720,8 W) y del MAE del 3,0 % (de 1252,5 W a 1215,1 W). En el SSFV UDENAR se observa una tendencia similar, con un incremento en R^2 de 2,07 % (de 0,6761 a 0,6967) y una reducción del RMSE del 3,2 % (de 2209,1 W a 2137,5 W). En cuanto al sesgo sistemático, en ambos sistemas el modelo multivariable reduce la magnitud del MBE: en CEDENAR pasa de +56,7 W a -27,5 W y en UDENAR de +142,8 W a +57,0 W, indicando que el modelo multivariable corrige parcialmente la subestimación del modelo simple. No obstante, los valores de R^2 inferiores a 0,70 en todos los casos reflejan que ninguno de los dos modelos captura plenamente la variabilidad intradía de la potencia generada, lo cual se atribuye a la alta variabilidad meteorológica de la región andina ecuatorial. La consistencia de las mejoras marginales entre ambos sistemas, con tecnologías de módulos e inversores diferentes y períodos de análisis distintos, confirma que la GHI es la variable dominante en la estimación de la generación fotovoltaica en esta región y que las variables ambientales adicionales aportan únicamente correcciones residuales. No obstante, tanto el modelo simple como el modelo multivariable representan una herramienta práctica y útil para estimar la energía generada por un SSFV en periodos prolongados, como se muestra en las estimaciones de la Tabla V y la Tabla VI.

TABLA V
COMPARACIÓN DE ENERGÍA MENSUAL GENERADA POR EL SSFV CEDENAR CON LAS VARIABLES DISPONIBLES EN SOLCAST

Mes	Energía generada (kWh/mes)	Estimación GHI η_{opt} (kWh/mes)	Estimación multivariable (kWh/mes)	Error modelo GHI	Error modelo multivariable
Febrero	1169,4	1237,1	1219,6	5,80%	4,30%
Marzo	1662	1596,1	1586,1	-4,00%	-4,60%
Abril	1433,5	1402,1	1447,9	-2,20%	1,00%
Mayo	1456,7	1451,8	1494,9	-0,30%	2,60%
Junio	1473,3	1392,1	1434,6	-5,50%	-2,60%
Julio	1297	1254,5	1304,2	-3,30%	0,60%
Agosto	383,5	397,8	411,7	3,70%	7,40%
TOTAL	**8875,4**	**8731,5**	**8899,2**	**-1,60%**	**0,30%**

TABLA VI
COMPARACIÓN DE ENERGÍA MENSUAL GENERADA POR EL SSFV UDENAR CON LAS VARIABLES DISPONIBLES EN SOLCAST

Mes	Energía generada (kWh/mes)	Estimación GHI η_{opt} (kWh/mes)	Estimación multivariable (kWh/mes)	Error modelo GHI	Error modelo multivariable
Julio	1830,2	1797,3	1776,3	-1,80%	-2,90%
Agosto	1800,3	1808,1	1812,3	0,40%	0,70%
Septiembre	1893,9	1858,4	1902,5	-1,90%	0,50%
Octubre	1834,7	1847,2	1904,6	0,70%	3,80%
Noviembre	1804,3	1715,7	1780,6	-4,90%	-1,30%
Diciembre	1609,2	1517,9	1587,1	-5,70%	-1,40%
TOTAL	**10772,5**	**10544,6**	**10763,3**	**-2,10%**	**-0,10%**

Como se aprecia en la Tabla V y la Tabla VI, a pesar de que ninguno de los dos modelos captura la variabilidad intradía de forma precisa, ambos logran reproducir la energía total acumulada con errores globales bajos. En el SSFV CEDENAR, el modelo simple subestima la generación total del período en -1,6 % (8731,5 kWh estimados frente a 8875,4 kWh reales), mientras que el modelo multivariable la sobreestima en apenas +0,3 % (8899,2 kWh). A nivel mensual, el modelo simple presenta sus mayores desviaciones en junio (-5,5 %) y marzo (-4,0 %). El modelo multivariable reduce estas discrepancias al incorporar variables correctivas como el ángulo cenital y la nubosidad, llevando el error de junio a -2,6 % y el de julio de -3,3 % a +0,6 %. No obstante, el modelo multivariable no supera al modelo simple en todos los meses: en febrero, marzo y agosto presenta errores comparables o ligeramente superiores, lo que refuerza la observación de que la incorporación de variables adicionales no garantiza una mejora uniforme.

En el SSFV UDENAR se observa un comportamiento análogo. El modelo simple subestima la energía total en -2,1 % (10544,6 kWh), mientras que el modelo multivariable reduce este error a -0,1 % (10763,3 kWh). Las mayores discrepancias del modelo simple se concentran en los últimos meses del período, con errores de -4,9 %

en noviembre y -5,7 % en diciembre, sugiriendo una pérdida de capacidad predictiva durante períodos con mayor variabilidad climática. El modelo multivariable corrige significativamente estas desviaciones, reduciendo el error de noviembre a -1,3 % y el de diciembre a -1,4 %, aunque introduce una ligera sobreestimación en octubre (+3,8 %). Septiembre constituye el mes de mayor generación registrada (1893,9 kWh), coincidiendo con el equinoccio donde los ángulos de incidencia solar son más perpendiculares sobre los módulos en la latitud de Pasto; en este mes el modelo multivariable logra un error de apenas +0,5 % frente al -1,9 % del modelo simple. En conjunto, ambos modelos proporcionan estimaciones razonables de la energía mensual, pero el modelo multivariable ofrece un desempeño levemente superior en la reducción del sesgo acumulado y la dispersión mensual de los errores, con la ventaja práctica del modelo simple de requerir únicamente datos de irradiancia horizontal y prescindir de coeficientes de ponderación contraintuitivos con el comportamiento físico realista de la influencia de variables ambientales adicionales sobre la potencia generada.

V. CONCLUSIONES

Los sistemas SSFV CEDENAR (12,3 kWp) y SSFV UDENAR (16,5 kWp) presentaron eficiencias globales óptimas de η = 76,81 % y η = 73,04 % respectivamente, valores que permiten reproducir la energía total acumulada con errores inferiores al 2,1 %. La diferencia de 3,77 puntos porcentuales entre ambas eficiencias se atribuye a factores como la orientación e inclinación de los módulos, las condiciones de sombreado, el estado de limpieza de los paneles y las diferencias en la eficiencia de conversión entre los inversores Huawei y Fronius. El promedio de ambos valores (aproximadamente 75 %) constituye un parámetro de referencia para la estimación de generación fotovoltaica en la ciudad de Pasto a partir de datos de GHI de SOLCAST.

El modelo lineal multivariable aportó mejoras marginales en R^2 de +1,52 y +2,07 puntos porcentuales para CEDENAR y UDENAR respectivamente, con reducciones del RMSE del 2,4 % y 3,2 %. La consistencia de estas mejoras moderadas en ambos sistemas, con tecnologías diferentes y períodos de análisis distintos, confirma que la irradiancia global horizontal es el principal determinante de la producción fotovoltaica en esta región y que las variables ambientales adicionales aportan correcciones residuales al modelo base.

La comparación entre las variables ambientales estimadas por SOLCAST y las mediciones de la estación meteorológica del SSFV UDENAR reveló que únicamente la irradiancia (ρ = 0,60) y la velocidad del viento (ρ = 0,56) presentan correlaciones significativas entre ambas fuentes. Las variables de temperatura, humedad relativa y dirección del viento estimadas por SOLCAST difieren sustancialmente de las mediciones locales, lo que limita la interpretación causal de los coeficientes del modelo multivariable y cuestiona la validez física de sus parámetros cuando se construye exclusivamente con datos satelitales.

Para estimaciones de energía mensual y acumulada a largo plazo, el modelo simple basado en GHI con la eficiencia global óptima constituye la herramienta más práctica y robusta, al requerir únicamente datos de irradiancia horizontal y proporcionar errores totales acumulados inferiores al 2,1 %, sin depender de variables ambientales cuya estimación satelital presenta discrepancias significativas con las condiciones locales reales y sin introducir coeficientes de ponderación contraintuitivos con el comportamiento físico de las variables.

Se recomienda para futuros trabajos ampliar el período de monitoreo a un ciclo anual completo, implementar mediciones de irradiancia en el plano inclinado y temperatura de celda para un desglose detallado de las pérdidas, y evaluar procedimientos de adaptación al sitio que permitan corregir las estimaciones satelitales de SOLCAST a partir de las mediciones de la estación meteorológica local.

Como se mostró en la Fig.10, las variables ambientales estimadas por SOLCAST diferentes a la irradiancia no tienen una correlación significativa con las variables reales medidas por la estación meteorológica en el punto geográfico correspondiente al bloque 1 de la Universidad de Nariño, sede Torobajo; por lo tanto, considerar un modelo lineal multivariable no aportaría una mejora significativa si no se aplican ajustes o correcciones en las variables ambientales distintas a la irradiancia estimada por SOLCAST.

Considerando los resultados obtenidos con el análisis de los modelos evaluados, SOLCAST representa una herramienta útil para la estimación de la energía fotovoltaica generada en la ciudad de Pasto, donde resulta conveniente emplear el modelo lineal simple con una eficiencia global η de 0,73 para estimaciones en períodos prolongados (meses), tal como se aprecia en la Tabla V y la Tabla VI, donde la precisión de la estimación alcanza al menos el 94,3 % para el SSFV UDENAR y el 94,2 % para el SSFV CEDENAR.

VI. AGRADECIMIENTOS

Un especial agradecimiento a Centrales Eléctricas de Nariño – CEDENAR S.A. E.S.P. y a la Universidad de Nariño por la financiación y el apoyo brindado en el marco del Convenio de Cooperación No. DEL-209394. Asimismo, se agradece la información suministrada en el marco de los proyectos “Campus Verde Udenar”, identificado con código BPIN 2020000100041, y “Modelos

Transaccionales de Energía – MTE", identificado con código BPIN 2021000100499, financiados por el Fondo de Ciencia, Tecnología e Innovación (CTeI) del SGR y Minciencias, Colombia.

REFERENCIAS

[1].R. M. Ismail, A. N. Makarov, K. B. Korneev, A. Y. Rusin, D. V. Chernyshov, and D. K. Ivanov, "Assessment of solar energy potential in egypt," in 2026 8th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE). IEEE, 2026, pp. 1–8.

[2].International Renewable Energy Agency (IRENA), "Renewable capacity statistics 2024," International Renewable Energy Agency, Abu Dhabi, Tech. Rep., Mar. 2024. [Online]. Available: https://www.irena.org/Publications/2024/Mar/Renewable-capacitystatistics-2024

[3].Congreso de Colombia, "Ley 1715 de 2014: Por medio de la cual se regula la integración de las energías renovables no convencionales al sistema energético nacional," Diario Oficial No. 49.150, May 2014, 13 de mayo de 2014. [Online]. Available: https://www.suinjuriscol.gov.co/viewDocument.asp?ruta=Leyes/1687143

[4].J. G. C. Burgos, F. J. F. Vargas, A. R. D. Revilla, C. H. L. Cabrera, and L. A. S. Almeida, "Una revisión del desarrollo de los modelos matemáticos para la estimación de la potencia eléctrica generada en un sistema fotovoltaico," Ciencia Latina Revista Científica Multidisciplinar, vol. 7, no. 3, pp. 8002–8013, 2023.

[5].A. Bandura, Y. Batsala, and P. Kurliak, "Use of regression models for estimation of electric power generation by photovoltaic power plants," Environmental Research, Engineering and Management, vol. 81, no. 2, pp. 64–77, 2025.

[6].A. E. Mouatasim and Y. Darmane, "Regression analysis of a photovoltaic (pv) system in fpo," in AIP Conference Proceedings, vol. 2056, no. 1. AIP Publishing LLC, 2018, p. 020008.

[7].G. Salazar, D. R. de Miranda, J. V. Furtado, O. C. Vilela, and N. Fraidenraich, "Simulación de generación de módulo fv en la ciudad de salta utilizando valores estadísticamente representativos de radiación solar y temperatura adaptados de bases de datos satelitales," Avances en Energías Renovables y Medio Ambiente-AVERMA, vol. 26, pp. 335–343, 2022.

[8].B. U. D. Abdullah, S. A. Khanday, N. U. Islam, S. Lata, H. Fatima, and S. H. Nengroo, "Comparative analysis using multiple regression models for forecasting photovoltaic power generation," Energies, vol. 17, no. 7, p. 1564, 2024.

[9].A. A. M. Rafi, N. Zaini, and M. F. A. Latip, "Estimation of pv power generation from predicted ghi using satellite imagery and deep learning (cnn and lstm)," in 2025 12th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). IEEE, 2025, pp. 449–456.

[10].N. S. Sitohang, I. D. Budi, T. J. T. Hashim, H. S. Saragi, M. Risal, and G. R. Afdholy, "Comparative evaluation of satellite-derived ghi in tropical regions using xgboost for sustainable solar energy planning," in 2025 International Conference on Decision Aid Sciences and Applications (DASA). IEEE, 2025, pp. 1778–1783.

[11].R. Ledesma and G. A. Salazar, "Machine learning for site adaptation of satellite-derived solar irradiance in northwestern argentina," Journal of Computer Science & Technology, vol. 25, 2025.

[12].J. M. Bright, "Solcast: Validation of a satellite-derived solar irradiance dataset," Solar energy, vol. 189, pp. 435–449, 2019.

[13].G. A. Salazar, R. Alonso-Suarez, A. L. Cirigliano, and R. D. Ledesma, "Evaluación del proceso de adaptación al sitio aplicado a la irradiancia solar global medida en la ciudad de salta, argentina," Avances en Energías Renovables y Medio Ambiente-AVERMA, vol. 25, pp. 353–362, 2021.

[14].J. AlFaraj, E. Popovici, and P. Leahy, "Solar irradiance database comparison for pv system design: A case study," Sustainability, vol. 16, no. 15, p. 6436, 2024.

[15].L. S. Hoyos Gómez, "Forecasting global solar radiation in nariño," Tesis doctoral (Doctora en Ingeniería), Universidad Nacional de Colombia, Manizales, Colombia, 2021.

[16].E. D. Rosas Santacruz, "Modelo de estimación de la potencia generada por un sistema fotovoltaico a partir de variables provenientes de bases de datos de recurso energético solar," Tesis de maestría (Maestría en Gestión de Tecnología de Información), Universidad Nacional Abierta y a Distancia (UNAD), Bogotá D.C., Colombia, 2025.

REVISTA DE INGENIERÍA
https://doi.org/

Facultad de Ingeniería
Centro de Investigación e Interacción Social

Universidad de Nariño
FUNDADA EN 1904

Simulación de un brazo robótico bioinspirado controlado mediante técnicas de Deep Learning

Simulation of a Bioinspired Robotic Arm Controlled Using Deep Learning Techniques

Sara G. Eraso-Montenegro [1]*, Pablo J. Jiménez-Torres [2], Wilson O. Achicanoy-Martínez [3],

[1] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. sara.eraso@udenar.edu.co
[2] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. pablojjto6@udenar.edu.co
[3] Departamento de Electrónica, Universidad de Nariño, Pasto, Nariño, Colombia. wilachic@udenar.edu.co

* Autor de correspondencia: sara.eraso@udenar.edu.co

Citación:. G. Eraso-Montenegro, P. J. Jiménez-Torres y W. O. Achicanoy-Martínez, "Simulación de un brazo robótico bioinspirado controlado mediante técnicas de Deep Learning," Revista de Ingeniería, vol. 4, no. 1, pp. 88-107, Nov. 2026.

Resumen—Este estudio aborda el control de un brazo robótico mediante señales electromiográficas (EMG) con el objetivo de mejorar la interpretación de dichas señales y permitir una integración más natural entre humano y robot. La metodología propuesta abarca la adquisición y procesamiento de señales EMG, la creación de un conjunto de datos y la implementación de cuatro modelos de aprendizaje profundo correspondientes a movimientos específicos de las articulaciones del hombro y el codo. Estos modelos traducen las señales EMG procesadas en ángulos articulares continuos para replicar los movimientos del usuario mediante una estrategia de control en lazo abierto. Para evaluar el rendimiento del sistema, se desarrolló una plataforma de validación basada en visión artificial que compara directamente los ángulos articulares reales del usuario con los ejecutados por un brazo robótico Kinova Gen3 Lite. Los resultados experimentales demuestran una exactitud (accuracy) de 0,70 y un error absoluto medio (MAE) de 5,79°, confirmando la viabilidad del control continuo en lazo abierto mediante señales biológicas. En conclusión, la integración de señales EMG con modelos de aprendizaje profundo constituye una estrategia eficaz para mejorar la naturalidad y la precisión del control de brazos robóticos, presentando un potencial significativo para aplicaciones protésicas y de asistencia avanzadas.

Palabras claves—Ángulos articulares, Brazo robótico, Cinemática directa, EMG, NARX, Visión por computadora.

Abstract—This study addresses the control of a robotic arm through electromyographic (EMG) signals with the aim of improving signal interpretation and enabling a more natural human-robot integration. The proposed methodology encompasses EMG signal acquisition and processing, dataset creation, and the implementation of four deep learning models corresponding to specific shoulder and elbow joint movements. These models translate processed EMG signals into continuous joint angles to replicate user movements using an open-loop control strategy. To evaluate system performance, a computer vision-based validation platform was developed to directly compare the user's actual joint angles with those executed by a Kinova Gen3 Lite robotic arm. Experimental results demonstrate an accuracy of 0.70 and a mean absolute error (MAE) of 5.79°, confirming the feasibility of continuous open-loop control using biological signals. In conclusion, integrating EMG signals with deep learning models constitutes an effective strategy to enhance the naturalness and accuracy of robotic arm control, offering significant potential for advanced prosthetic and assistive applications.

Keywords—Computer vision, Direct kinematics, EMG, Joint angles, NARX, Robotic arm.

I. INTRODUCCIÓN

El movimiento de las extremidades superiores desempeña un papel fundamental en la interacción del ser humano con su entorno, ya que permite realizar actividades cotidianas que requieren precisión y coordinación. Actualmente, alrededor de 57,7 millones de personas a nivel mundial viven con alguna amputación; de estas, aproximadamente 22,3 millones corresponden a miembros superiores. Dentro de este grupo, el 19,6 % presenta una amputación unilateral y el 19,1 % una amputación bilateral [1]. La pérdida de una extremidad puede originarse por diversas causas, principalmente de carácter traumático, congénito o patológico.

En Colombia, a estas causas se suman factores agravantes derivados del contexto sociopolítico del país: el conflicto armado y la violencia. Específicamente en el departamento de Nariño, se han registrado incidentes vinculados a más de 30 minas antipersonal y 42 artefactos explosivos [2], los cuales ocasionan traumatismos severos que frecuentemente derivan en amputaciones. De acuerdo con datos oficiales del DANE, más de 27.000 personas en el departamento presentan alteraciones en la función motora de sus extremidades. A nivel local, en la ciudad de San Juan de Pasto, esta cifra supera los 9.000 habitantes con dificultades para la movilidad y manipulación de objetos [3]. Ante esta problemática regional, resulta imperativo el desarrollo de tecnologías orientadas a asistir, restaurar o complementar el movimiento. En este escenario, la robótica médica cobra especial relevancia, particularmente en el diseño de dispositivos capaces de emular la cinemática de la extremidad humana y adaptarse a las necesidades específicas de cada usuario.

Los brazos robóticos constituyen una de las principales alternativas tecnológicas para abordar estas necesidades, gracias a su capacidad para ejecutar movimientos precisos y repetibles. En el estado del arte de las prótesis de miembro superior se distinguen sistemas accionados por el cuerpo (body-powered), así como alternativas pasivas, mecánicas y neumáticas, las cuales presentan una mayor tasa de abandono o rechazo en comparación con las prótesis mioeléctricas [4], [5]. Estas últimas operan a partir de señales biológicas, específicamente las electromiográficas (EMG), y su desempeño ha mejorado significativamente al ser integradas con técnicas de aprendizaje profundo (deep learning) [6]. No obstante, la utilidad de las señales EMG no se restringe a los dispositivos protésicos, sino que se extiende al telecuerpo y control de brazos robóticos, enfoque adoptado en la presente investigación. En consecuencia, surge la necesidad de desarrollar esquemas de control avanzados que permitan establecer un mapeo directo y continuo entre la actividad muscular del usuario y la cinemática del manipulador robótico.

El uso de señales EMG para el control de prótesis y manipuladores robóticos ha evolucionado desde métodos clásicos, basados en umbrales de activación, esquemas on-off y control proporcional, hacia el reconocimiento de patrones mediante aprendizaje automático (machine learning). Los primeros avances emplearon clasificadores tradicionales como el análisis discriminante lineal (LDA) y k-nearest neighbors (KNN) [7], [8]. Sin embargo, para capturar la compleja dinámica no lineal entre la actividad muscular y la estimación continua de ángulos articulares, el enfoque ha transicionado hacia el aprendizaje profundo (deep learning). Actualmente, arquitecturas avanzadas como las redes convolucionales (CNN), recurrentes (RNN) y estructuras híbridas como LSTM y NARX (Nonlinear Autoregressive with Exogenous Inputs) han demostrado una notable superioridad en la interpretación de estas bioseñales, permitiendo un control más fluido y natural [9]–[11].

Teniendo en cuenta estos avances, se propone en esta investigación una decodificación de señales EMG de diferentes movimientos de las extremidades superiores con modelos de aprendizaje profundo, particularmente una NARX, para estimar los ángulos de estos movimientos que permitan su replicación en un brazo robótico real Kinova Gen3 Lite. Experimentalmente, se ha logrado una exactitud (accuracy) de 0,70 y un error absoluto medio (MAE) de 5,79° en la estimación de dichos ángulos, confirmando que es viable realizar un control continuo en lazo abierto a partir de las señales EMG. De esta manera, se concluye que la integración de señales EMG a partir de modelos de deep learning es una estrategia viable para la evaluación de brazos robóticos bioinspirados con mayor precisión, adaptabilidad e interacción natural con el usuario. A través de esta investigación, se brinda la oportunidad de integrar conocimientos de la ingeniería electrónica, procesamiento de señales, robótica e inteligencia artificial.

Este artículo se estructura de la siguiente manera: en la Sección II se presenta la metodología seguida; en la Sección XX se detalla el XX. Finalmente, se presentan las conclusiones de la investigación.

II. METODOLOGÍA

La metodología propuesta, esquematizada en la Fig. 1, inicia con la adquisición de señales electromiográficas (EMG) mediante dos canales de electrodos de superficie ubicados estratégicamente en el miembro superior. Estos registros conforman el conjunto de datos que alimenta el núcleo computacional del sistema: una red neuronal de tipo NARX. Esta arquitectura procesa la información mioeléctrica para estimar de manera continua los ángulos articulares del usuario. De forma

paralela, se formuló un modelo cinemático basado en el análisis anatómico de la extremidad, el cual proporciona la base matemática para mapear espacialmente dichas estimaciones. Los ángulos predichos por la red construyen un vector de configuración articular que se utiliza de manera dual: por un lado, gobierna una simulación computacional donde se recrea el movimiento de un brazo virtual; por otro, se transmite como comando de posición al brazo robótico Kinova Gen3 Lite, replicando físicamente el desplazamiento mecánico dictado por la intención motora del usuario.

Para la validación del sistema, se implementó un módulo de visión por computadora encargado de registrar el movimiento del usuario en tiempo real. Mediante algoritmos de estimación de pose, este sistema identifica los puntos articulares clave de la extremidad, lo que permite calcular los ángulos reales durante la ejecución del gesto motor. De manera homóloga, una interfaz de captura óptica monitorea la trayectoria física ejecutada por el manipulador Kinova Gen3 Lite. Así, las mediciones angulares extraídas por la cámara actúan como un sistema de referencia independiente que permite cuantificar el error entre la estimación continua de la red NARX, la cinemática real del usuario y el posicionamiento mecánico del robot. En las siguientes subsecciones se detalla cada etapa metodológica.

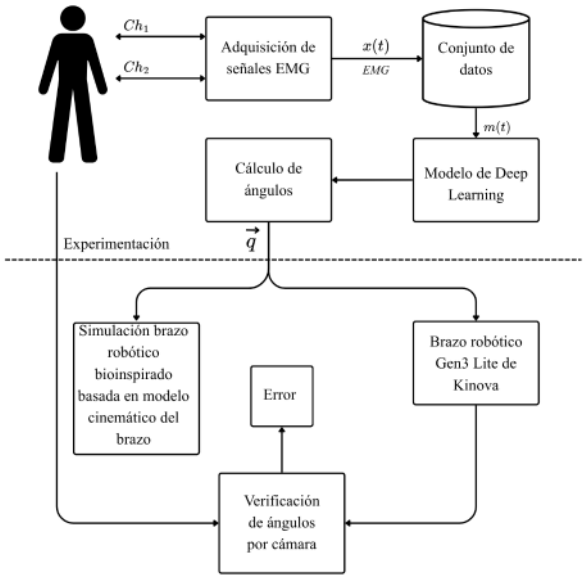


Figura 1: Diagrama de bloques de la metodología del desarrollo de la investigación. Fuente: elaboración propia.

II-A. Cinemática de un brazo antropomórfico

La cinemática permite describir el movimiento de la extremidad superior y del sistema robótico asociado respecto a un marco de referencia. En este trabajo se emplea la cinemática directa como enfoque principal, ya que el objetivo de la investigación consiste en la estimación de ángulos articulares a partir de señales EMG. Puesto que la actividad electromiográfica refleja la activación muscular responsable del movimiento, dichas señales contienen información directa y relevante para predecir la configuración articular [12]. El esquema cinemático propuesto corresponde a un manipulador de arquitectura antropomórfica, como se ilustra en la Fig. 2, concebido para emular el comportamiento del brazo humano. Este sistema se representa como una cadena cinemática abierta formada por eslabones rígidos unidos mediante articulaciones, donde un extremo equivale a la base (el hombro) y el otro al efector final (la mano) [13].

En la Fig. 2 se representan los eslabones del brazo mediante líneas de color azul y rojo para la cadena cinemática principal, y líneas de color verde, cian y rosa para el sistema asociado al efector final (el cual queda fuera del alcance de este análisis). Las articulaciones rotacionales se modelan por convención mediante cilindros. Con respecto a los ejes del marco de referencia global, se adoptan los planos anatómicos del cuerpo humano: el eje craneocaudal corresponde al eje Z, el eje dorsoventral al eje Y y el eje laterolateral al eje X. La combinación de los ejes XYZ define los planos anatómicos. Así, el plano definido por los ejes YZ corresponde al plano sagital; el plano XZ, al plano frontal o coronal; y el plano XY, al plano transversal o axial.

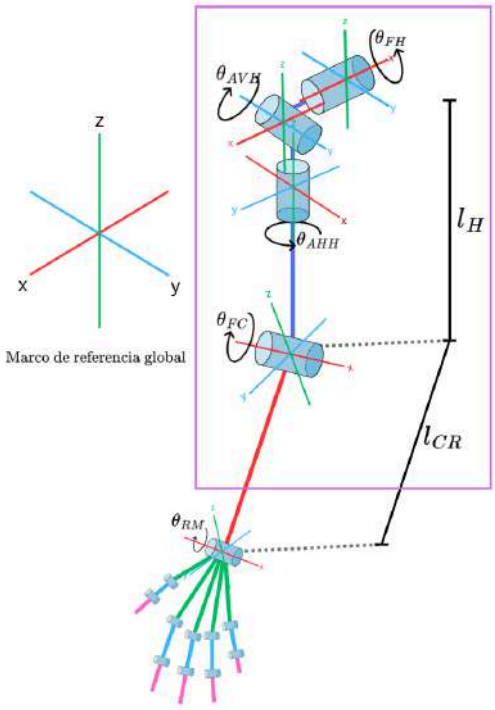


Figura 2: Esquema cinemático de la investigación, el cual ilustra los ejes de rotación y las longitudes de los eslabones en la cadena cinemática principal (resaltada en el recuadro lila). En el costado izquierdo se detalla el marco de referencia global. Fuente: elaboración propia.

En esta misma figura se detallan los marcos de referencia de cada grado de libertad de la cadena cinemática principal, donde las longitudes de los eslabones lH y lCR corresponden a los huesos del húmero y del cúbito/radio, respectivamente. El primer cilindro, θFH, está asociado al movimiento de flexión y extensión del hombro, realizado en el plano sagital; el segundo cilindro, θAVH, describe la aducción y abducción vertical del hombro en el plano frontal; el tercer cilindro, θAHH, se asocia a la aducción y abducción horizontal del hombro, ocurriendo en el plano transversal; finalmente, el cuarto cilindro, θFC, corresponde al ángulo de flexión y extensión del codo, asignado al último grado de libertad del sistema, el cual es paralelo a θFH dado que ambos movimientos ocurren en el mismo plano sagital. El último eslabón, θRM, correspondiente a la rotación de la muñeca (no incluido en esta investigación), se ubica fuera de la cadena cinemática principal y se ejecuta en el plano sagital.

Cabe destacar que la flexión y extensión de todas las articulaciones del brazo se producen en el plano sagital, por lo que el parámetro de torsión es αi = 0 en las matrices de DH asociadas a dichos grados de libertad. En contraste, el movimiento de aducción y abducción horizontal transiciona al plano transversal, implicando que αi = 90° en su respectiva matriz. De forma análoga, la aducción y abducción vertical del hombro se desarrolla en el plano frontal, lo que también exige un parámetro αi = 90°.

Establecidas estas variables, la cinemática de la cadena principal se modela mediante la convención de Denavit-Hartenberg (DH) [13]. En esta formulación, el marco de referencia global, denotado por {0}, se fija en la base del manipulador. Posteriormente, se asigna un marco de referencia local {i} a cada articulación de la cadena cinemática, con i = 1, 2, ..., n, donde n es el número total de grados de libertad. Cada matriz homogénea Ai representa la transformación relativa del marco {i} respecto al marco inmediatamente anterior {i−1}; es decir, para cada articulación, la transformación se expresa como:

$$A_i = i^{-1}T_i, \tag{1}$$

donde el superíndice indica el marco de referencia base de la transformación y el subíndice señala el marco local del eslabón actual. El modelo cinemático completo se obtiene mediante la multiplicación sucesiva de estas matrices de transformación homogénea, logrando así una descripción matemática rigurosa de toda la cadena. La matriz general del modelo DH estándar se define como [13]:

$$A_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

La ecuación (2) describe simultáneamente la orientación y la posición del marco local {i} con respecto al marco {i − 1}. Los componentes en las tres primeras filas y columnas describen la rotación de la articulación, mientras que los tres primeros elementos de la cuarta columna representan el vector de traslación. Esta matriz se construye mediante el producto secuencial de una rotación y traslación en el eje Z, seguidas de una traslación y rotación en el eje X. Así, ai es la longitud del eslabón (traslación en X), di es el desplazamiento lineal de la articulación (traslación en Z), αi es el ángulo de torsión del eslabón (rotación en X) y θi es el ángulo articular (rotación en Z).

Para evitar ambigüedades en la asignación geométrica, el marco base {0} se fija rígidamente en el hombro del brazo robótico. Los marcos locales {1}, {2}, {3} y {4} se asignan de manera secuencial a las articulaciones asociadas con la flexión/extensión del hombro, aducción/abducción vertical del hombro, aducción/abducción horizontal del hombro y flexión/extensión del codo, respectivamente. De esta forma, el índice i ∈ {1, 2, 3, 4} identifica unívocamente cada grado de libertad dentro de la cadena cinemática principal.

II-B. Adquisición de la señal EMG

La Tabla I presenta las principales características de los distintos métodos de adquisición de señales EMG, incluyendo el tipo de electrodo, el rango de frecuencias de operación y el rango dinámico, definido este último como el intervalo de amplitudes de voltaje medibles de manera útil. Para el desarrollo de esta investigación, se seleccionó la electromiografía de superficie (sEMG), la cual permite capturar la actividad muscular de forma no invasiva mediante electrodos adheridos a la piel.

TABLA I
CARACTERÍSTICAS DE ADQUISICIÓN DE SEÑALES EMG [14]

Tipo	Método	Frecuencia	Rango dinámico y observación
EMG	Aguja	100 Hz–1 kHz	5 µV–1 mV; potencial de rama nerviosa
SFEMG	Aguja	500 Hz–10 kHz	1–10 µV; fibras individuales
MUAP	Aguja	5 Hz–10 kHz	50 µV–5 mV; potencial de unidad motora
SEMG	Superficie	20–500 Hz	50 µV–5 mV; señal superficial

II-B1. Diseño de la placa de adquisición de la señal EMG: Con base en la literatura revisada sobre las señales EMG de las extremidades superiores, se determinó que el ancho de banda del circuito de adquisición debe operar en el rango de 20 Hz a 500 Hz [14], utilizando una frecuencia de muestreo de 1 kHz para cumplir satisfactoriamente con el teorema de Nyquist. El diseño electrónico para cada canal de adquisición (Ch1 y Ch2) —compuesto por dos electrodos de medición y uno de referencia— se estructuró en las siguientes etapas consecutivas: primero, una etapa de amplificación con ganancia fija de G = 100, implementada mediante un amplificador de instrumentación AD621, el cual procesa la diferencia de potencial entre los electrodos y rechaza el ruido en modo común (como la interferencia de 60 Hz de la red eléctrica). Seguidamente, se incorporó una etapa de filtrado pasa-banda de respuesta plana y ganancia unitaria, conformada por un filtro Butterworth pasa-altas y un pasa-bajas, ambos de tercer orden, con frecuencias de corte ajustadas a 20 Hz y 500 Hz, respectivamente. Finalmente, la señal pasa por un rectificador de precisión de onda completa, diseñado con resistencias de valores idénticos, amplificadores operacionales y diodos de uso general. El diagrama esquemático que detalla esta arquitectura electrónica para un solo canal se presenta en la Fig. 3.

II-B2. Construcción de la placa de adquisición de la señal EMG: En todas las etapas del circuito, a excepción de la amplificación de instrumentación inicial, se emplearon amplificadores operacionales comerciales de uso extendido, como el µA741. Debido al requerimiento de fuentes de alimentación duales, las pruebas iniciales se realizaron con los equipos del Laboratorio de Electrónica de la Universidad de Nariño. Tras validar la simulación del circuito (ver Fig. 3), se procedió al montaje en protoboard, donde las ganancias y frecuencias de corte fueron rigurosamente ajustadas y verificadas mediante un generador de señales y un osciloscopio. Una vez comprobado el correcto funcionamiento, se diseñaron y fabricaron las placas de circuito impreso (PCB) —una por cada canal— siguiendo los estándares internacionales de la IPC (Association Connecting Electronics Industries), específicamente la norma IPC-2221B. A este diseño propietario de tarjeta de adquisición se le denominó Freya EMG Sensor. La Fig. 4 ilustra la tarjeta final con todos los componentes soldados.

II-B3. Pruebas de la PCB Freya EMG Sensor en laboratorio: Para validar el desempeño de las PCB en un escenario de aplicación real, se llevaron a cabo pruebas de adquisición en el Laboratorio de Electrónica de la Universidad de Nariño. Se colocaron electrodos de superficie en músculos antagónicos de la extremidad superior para registrar simultáneamente los canales Ch1 y Ch2, permitiendo observar la actividad coordinada durante la ejecución de un gesto motor. La Fig. 5 muestra la captura oscilográfica del par de señales EMG obtenidas: la traza azul corresponde a la actividad del músculo braquiorradial (Ch2), mientras que la traza amarilla representa al músculo bíceps braquial (Ch1). Estos registros se capturaron mientras un sujeto realizaba una flexión sostenida del codo a un ángulo constante de 90° durante un período de 20 s.

Finalmente, se comprobó que los registros temporales pudieran ser exportados y almacenados correctamente en una estación de trabajo, donde se procesaron computacionalmente para graficar su respuesta temporal y calcular su densidad espectral de potencia mediante la Transformada Rápida de Fourier (FFT). Este análisis en frecuencia confirmó empíricamente que la energía principal de las señales se concentra dentro del ancho de banda operativo diseñado, comprendido entre los 20 Hz y 500 Hz.

II-C. Conjunto de datos

Para la construcción de un conjunto de datos propio, la adquisición de las señales EMG se llevó a cabo en las instalaciones del Laboratorio de Electrónica de la Universidad de Nariño, en un entorno acondicionado específicamente para este propósito. La Fig. 6 ilustra el esquema del montaje experimental, el cual incluye un panel transparente ubicado frente al participante. Sobre este panel se trazó un círculo trigonométrico empleado como guía visual para asegurar que los movimientos de la extremidad superior alcanzaran la amplitud angular deseada. Durante las pruebas, un investigador se posicionó frente al panel para supervisar la correcta ejecución biomecánica del gesto motor, mientras que un sistema de captura óptica (cámara) se instaló frontalmente para registrar y validar los valores angulares alcanzados. Con el fin de evitar interferencias electromagnéticas y visuales, el resto del instrumental de adquisición se dispuso detrás del panel de pruebas.

II-C1. Ubicación de los electrodos en la adquisición: Los electrodos de superficie se dispusieron sobre la musculatura del miembro superior, dirigidos a los pares antagónicos (agonista-antagonista) responsables de la articulación bajo estudio. La Fig. 7 detalla la topología de colocación para el hombro y el codo, involucrando los músculos bíceps braquial, deltoides (porciones anterior, media/externa y posterior) y el supinador largo del antebrazo. Para cada gesto articular, la instrumentación se configuró de la siguiente manera:

- **Flexión y extensión del codo:** Se instrumentaron el músculo bíceps braquial (ver Fig. 7a) y el supinador largo del antebrazo (ver Fig. 7e). El músculo antagonista primario (tríceps braquial) fue excluido de la captura por dos motivos principales:

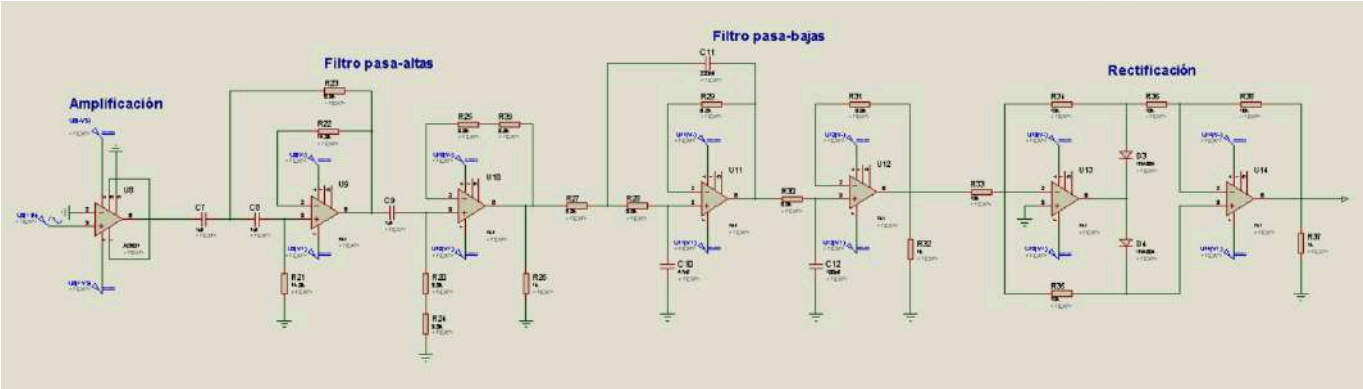


Figura 3: Diagrama esquemático de la electrónica de adquisición diseñada para un canal de la señal EMG utilzando electrodos de superficie. Fuente: elaboración propia.

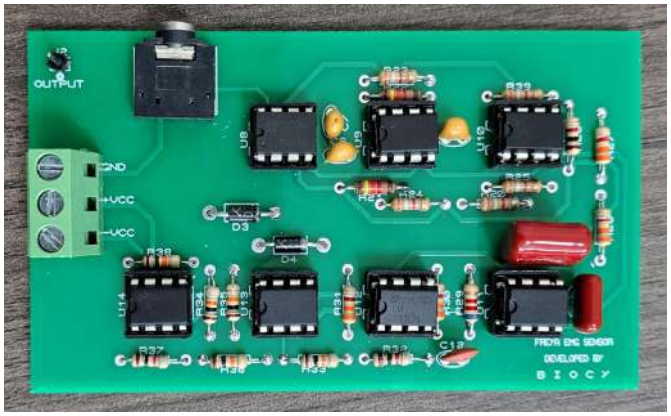


Figura 4: PCB final de la tarjeta Freya EMG Sensor. Fuente: elaboración propia.

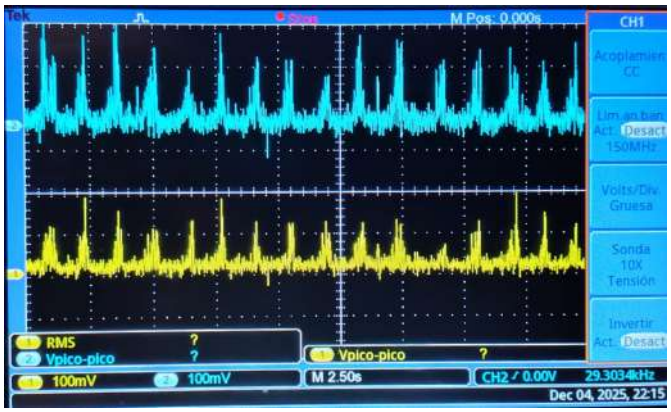


Figura 5: Captura oscilográfica de señales EMG reales en el dominio del tiempo. La señal Ch1 corresponde al supinador del antebrazo y Ch2 al bíceps braquiorradial, ambas adquiridas con la tarjeta Freya EMG Sensor. Fuente: elaboración propia.

En primer lugar, la literatura respalda que la actividad del complejo bíceps-braquiorradial presenta una fuerte correlación directa con el ángulo de flexión del codo [16]; en segundo lugar, la inclusión del supinador largo no solo complementa la lectura del bíceps, sino que enriquece el conjunto de datos con información cinemática de la muñeca y la mano, lo cual resulta de alto valor para futuras expansiones del proyecto enfocadas en el control

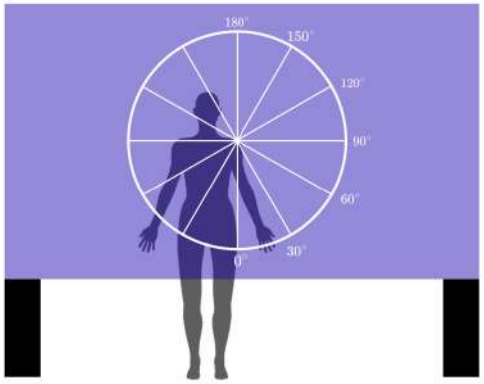


Figura 6: Esquema del montaje experimental y ubicación del participante en el laboratorio. Fuente: elaboración propia.

- **Flexión y extensión del hombro (plano sagital):** Los electrodos se fijaron en el deltoides anterior (ver Fig. 7c) y en el deltoides posterior (ver Fig. 7d).
- **Aducción y abducción vertical del hombro (plano frontal):** La instrumentación se centró en el deltoides anterior (ver Fig. 7c) y el deltoides medio/externo (ver Fig. 7b).
- **Aducción y abducción horizontal del hombro (plano transversal):** El registro mioeléctrico se tomó sobre el deltoides medio/externo (ver Fig. 7b) y el deltoides posterior (ver Fig. 7d).

II-C2. Protocolo de adquisición: Para estandarizar la captura de datos, se definió un protocolo de adquisición que delimita los gestos motores abordados en esta investigación. La Tabla II detalla los cuatro movimientos articulares seleccionados: flexión y extensión del codo, flexión y extensión del hombro, aducción y abducción vertical del hombro, y aducción y abducción horizontal del hombro. Con el fin de ilustrar el procedimiento de manera integral, la tabla incluye representaciones visuales de cada acción, la secuencia angular objetivo, los tiempos de ejecución fijados por serie y la respectiva configuración espacial de los electrodos.

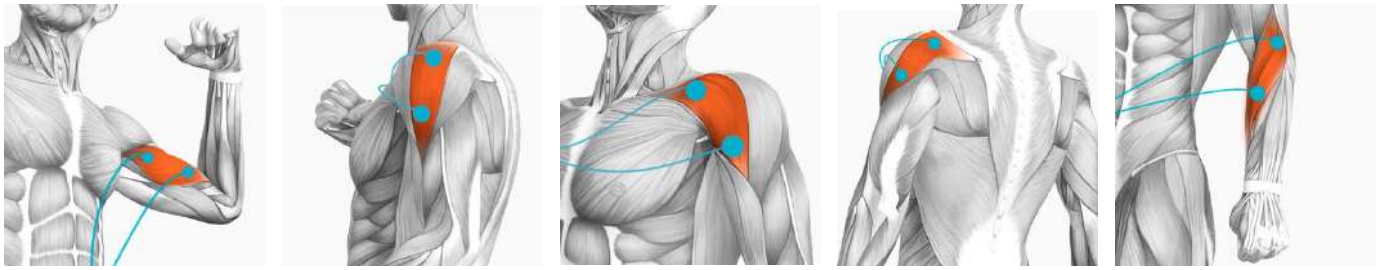


Figura 7: Puntos de instrumentación para los electrodos de superficie en la musculatura del miembro superior. Fuente: [15].

TABLA II: PROTOCOLO DE EJERCICIOS, SECUENCIA TEMPORAL Y UBICACIÓN DE ELECTRODOS DE SUPERFICIE PARA LA ADQUISICIÓN DE SEÑALES EMG EN HOMBRO Y CODO. FUENTE: ELABORACIÓN PROPIA A PARTIR DE [18].

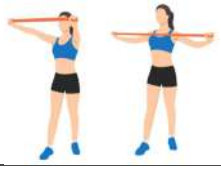
Movimiento articular	Imagen	Secuencia	Tiempo	Ubicación de electrodos
Flexión y extensión de codo (supinación) (M1)		Flexión y extensión del codo con antebrazo en supinación, en incrementos angulares regulares de 30°, 60°, 90°, 120° y 150°.	3 s por incremento (≈ 15 s en total).	Biceps braquial durante la flexión (Fig. 7a) y músculo supinador del antebrazo durante la extensión en supinación (Fig. 7e).
Flexión y extensión del hombro (M2)		Flexión del hombro en incrementos de 30°, 60°, 90°, 120° y 150°, y extensión en la misma secuencia.	3 s por incremento (≈ 15 s en total).	Deltoides anterior (Fig. 7c) y deltoides posterior (Fig. 7d).
Aducción y abducción vertical del hombro (M3)		Elevación del hombro en incrementos de 30°, 60°, 90°, 120° y 150°.	3 s por incremento (≈ 15 s en total).	Deltoides anterior (Fig. 7c) y deltoides medio/ externo (Fig. 7b).
Aducción y abducción horizontal del hombro (M4)		Aducción del hombro en 18°, 36°, 54°, 72° y 90°.	3 s por incremento (≈ 15 s en total).	Deltoides medio/externo (Fig. 7b) y deltoides posterior (Fig. 7d).

TABLA III: RANGOS DE MOVIMIENTO DE LAS ARTICULACIONES DEL MIEMBRO SUPERIOR. FUENTE: ELABORACIÓN PROPIA.

Articulación	GDL	Movimiento	Rango (°)
Hombro	GDL1	Aducción horizontal	0°–130°
		Abducción horizontal	0°–45°
	GDL2	Aducción vertical	0°–45°
		Abducción vertical	0°–180°
	GDL3	Flexión	0°–180°
		Extensión	0°–60°
Codo	GDL4	Flexión	0°–150°
		Extensión	0°

Es importante señalar que los rangos de movilidad están intrínsecamente restringidos por límites biomecánicos naturales. Por ello, la Tabla III resume las articulaciones evaluadas, su asignación como grados de libertad (GDL) y sus correspondientes límites anatómicos de movimiento.

II-C3. Participantes: Para la adquisición de las señales se conformó una muestra de diez participantes (6 hombres y 4 mujeres) con el propósito de garantizar la variabilidad inter-sujeto en los datos registrados, proyectando la expansión de este corpus en trabajos futuros. El protocolo requirió que cada participante ejecutara un total de 40 series (10 repeticiones por cada uno de los cuatro movimientos analizados), logrando así un acumulado de 100 series por tipo de movimiento articular. Cada serie consistió en cinco activaciones musculares

estáticas (pulsos) consecutivas, correspondientes a los cinco incrementos angulares definidos en el protocolo experimental.

II-C4. Descripción de los archivos del conjunto de datos: La Tabla IV detalla la nomenclatura adoptada para el etiquetado y almacenamiento sistemático de las señales adquiridas. Esta codificación se fundamenta en los grupos musculares instrumentados: el complejo bíceps-braquiorradial, el supinador largo del antebrazo y las porciones anterior, media (externa) y posterior del músculo deltoides. La base de datos incluye tanto las series temporales originales como los archivos correspondientes a su densidad espectral, calculada mediante la Transformada Rápida de Fourier (FFT). Por su parte, la Tabla V consolida el volumen total de señales que conforman la primera versión de este conjunto de datos, categorizándolas según el gesto articular y el plano anatómico de ejecución (frontal, sagital o transversal). Es pertinente aclarar que la totalidad de los registros almacenados son señales crudas, cuyo único acondicionamiento previo proviene del procesamiento analógico (amplificación y filtrado) provisto por el hardware de la tarjeta Freya EMG Sensor.

TABLA IV: MÚSCULOS INSTRUMENTADOS Y NOMENCLATURA DE LAS SEÑALES EMG. FUENTE: ELABORACIÓN PROPIA.

Músculo	Nomenclatura
Bíceps braquial	B, B_fft
Supinador largo del antebrazo	A, A_fft
Deltoides anterior (hombro plano sagital)	DAHS, DAHS_fft
Deltoides posterior (hombro plano sagital)	DPHS, DPHS_fft
Deltoides anterior (hombro plano frontal)	DAHF, DAHF_fft
Deltoides medio/externo (hombro plano frontal)	DEHF, DEHF_fft
Deltoides medio/externo (hombro plano transv.)	DEHT, DEHT_fft
Deltoides posterior (hombro plano transv.)	DPHT, DPHT_fft

II-D. Modelo de estimación del ángulo articular

En esta investigación se implementó un modelo de aprendizaje profundo (deep learning) basado en una arquitectura de red neuronal recurrente tipo NARX (Nonlinear Autoregressive with eXogenous inputs). Este tipo de redes destaca por su robustez en el pronóstico y modelado de series temporales no lineales. En la literatura especializada, su uso se ha reportado como altamente eficaz para modelar la relación dinámica entre las bioseñales electromiográficas y la cinemática articular [19]–[21].

Desde una perspectiva biomecánica, el movimiento de las extremidades no está compuesto por eventos estáticos o aislados, sino que constituye un proceso continuo. En consecuencia, el ángulo de una articulación (como el codo o el hombro) en un instante de tiempo t exhibe una fuerte dependencia de sus estados angulares previos, producto de la inercia del segmento corporal y la fluidez natural del movimiento. Dada su naturaleza recurrente, la arquitectura NARX emplea una retroalimentación de sus propias salidas pasadas (el histórico de ángulos estimados) para predecir el estado actual. De este modo, la red logra mapear las señales mioeléctricas adquiridas para estimar de forma continua el ángulo articular objetivo durante cada gesto motor.

Por otro lado, la correspondencia entre la activación eléctrica muscular (bíceps, antebrazo, deltoides) y el desplazamiento articular resultante es intrínsecamente no lineal. En esta dinámica intervienen múltiples factores biomecánicos complejos, tales como la relación longitud-tensión de las fibras musculares, la curva fuerza-velocidad, las variaciones dinámicas del brazo de palanca y la fricción articular intrínseca. Las redes NARX están diseñadas precisamente para capturar estas dinámicas no lineales sin la necesidad de formular o derivar analíticamente ecuaciones diferenciales anatómicas complejas. Esto representa un aporte metodológico significativo de la presente investigación en la simplificación del análisis del movimiento humano.

En este contexto, la selección de las señales EMG como entradas exógenas del modelo resulta idónea, dado que la predicción angular depende tanto del historial del propio sistema como de estímulos externos independientes (se utilizaron al menos dos canales EMG correspondientes a distintos músculos por cada movimiento). Fisiológicamente, el impulso eléctrico registrado en la miografía antecede a la contracción mecánica y actúa como la fuerza motriz generadora del movimiento. En consecuencia, el uso de bioseñales EMG como entradas exógenas respeta de manera natural el principio de causalidad biológica del sistema neuromuscular, garantizando una interpretación físicamente coherente del fenómeno.

Matemáticamente, cuando esta arquitectura se emplea para tareas de regresión continua, la salida y(t) (el ángulo articular estimado) se expresa como una función de los retardos temporales tanto de las variables de entrada como de las salidas previas. Esta formulación dota al modelo de una fuerte dependencia temporal y de la capacidad de integrar la dinámica interna del sistema [22], definida por la ecuación:

$$y(t) = F(u(t-1:t-n_u), y(t-1:t-n_y)), \quad (3)$$

donde u representa el vector de las entradas exógenas pasadas, nu es el tamaño de la ventana de retardo para dichas entradas, e y (dentro del argumento de la función

TABLA V: RESUMEN DE MOVIMIENTOS, SECUENCIAS ANGULARES, NOMENCLATURA Y RECUENTO TOTAL DE SEÑALES EMG ADQUIRIDAS. FUENTE: ELABORACIÓN PROPIA.

Movimiento	Secuencia angular (°)	Nomenclatura	Señales crudas	Señales FFT
Flexión/Extensión de codo (Plano sagital)	30, 60, 90, 120, 150	B, A	100 B; 100 A	100 B_fft; 100 A_fft
Aducción/Abducción hombro (Plano frontal)	30, 60, 90, 120, 150	DAHF, DEHF	100 DAHF; 100 DEHF	100 DAHF_fft; 100 DEHF_fft
Flexión/Extensión hombro (Plano sagital)	30, 60, 90, 120, 150	DAHS, DPHS	100 DAHS; 100 DPHS	100 DAHS_fft; 100 DPHS_fft
Aducción/Abducción hombro (Plano transv.)	18, 36, 54, 72, 90	DEHT, DPHT	100 DEHT; 100 DPHT	100 DEHT_fft; 100 DPHT_fft
Total general			**800 archivos**	**800 archivos**

no lineal F(·)) denota el historial de las salidas previas, siendo ny el tamaño de la respectiva ventana de retroalimentación.

El objetivo central del modelo NARX propuesto es estimar el ángulo articular continuo a partir del conjunto de datos de señales electromiográficas de superficie (sEMG). Posteriormente, esta predicción angular es transmitida como consigna de posición al brazo robótico bioinspirado para su ejecución mediante una estrategia de control en lazo abierto. Para concretar el desarrollo de este modelo, el flujo de trabajo computacional se estructuró en las siguientes cinco etapas operativas: (i) preprocesamiento de la señal EMG, (ii) segmentación temporal y extracción de características, (iii) construcción de los vectores de entrada para la red NARX, (iv) entrenamiento del modelo neuronal y (v) evaluación de desempeño en inferencia. A continuación se detalla cada etapa.

II-D1. Preprocesamiento de la señal EMG: Las señales EMG se adquieren mediante dos canales independientes (Ch1 y Ch2) que registran la actividad muscular simultánea durante la ejecución del gesto motor. Cada registro contiene dos variables: el tiempo t en segundos (adquirido a una frecuencia de muestreo de 1 kHz) y la amplitud mioeléctrica x(t) en voltios. Con el fin de mitigar la variabilidad inherente de las bioseñales entre diferentes sesiones y participantes, se aplicó un proceso de normalización robusta basada en el puntaje Z (z-score robusto), definida como $z = (x - \text{med}(x)) / (1{,}4826 \cdot \text{MAD}(x))$, donde med(x) corresponde a la mediana de la señal y MAD(x) representa la desviación absoluta de la mediana. Este método es especialmente adecuado para señales estocásticas biológicas, ya que reduce drásticamente la sensibilidad a valores atípicos en comparación con las normalizaciones clásicas basadas en la media y la desviación estándar [23]. Posteriormente, las señales de ambos canales se sincronizan y se truncan a la misma longitud, garantizando que cada ventana temporal de análisis analice exactamente el mismo instante de la dinámica muscular.

II-D2. Segmentación temporal y extracción de características: Dada la naturaleza estocástica y no estacionaria de las señales EMG, el análisis temporal se ejecutó utilizando ventanas deslizantes superpuestas. Cada señal se segmentó en tramos de longitud fija W, con un desplazamiento de avance Δ. Tras un proceso de sintonización empírica destinado a maximizar la precisión en la estimación angular, los parámetros óptimos se establecieron en W = 350 ms y Δ = 50 ms. Esta alta superposición garantiza una captura continua de la dinámica muscular con suficiente resolución temporal, permitiendo que el modelo extraiga los patrones del movimiento en curso sin perder el contexto de las transiciones articulares previas.

A partir de cada segmento W, se extrajeron características representativas en el dominio del tiempo y de la frecuencia, ampliamente validadas en la literatura de electromiografía [24]. Las métricas temporales calculadas fueron:

- **RMS (Root Mean Square):** Cuantifica la energía efectiva de la señal EMG.
- **MAV (Mean Absolute Value):** Calcula el promedio del valor absoluto de la amplitud, reflejando el nivel de contracción.
- **WL (Waveform Length):** Mide la longitud acumulada de la forma de onda mediante la suma de las variaciones absolutas consecutivas, lo cual es un indicador directo de la complejidad de la señal.
- **ZC (Zero Crossings):** Contabiliza el número de veces que la señal cruza el eje cero, proporcionando una aproximación rápida del contenido frecuencial.
- **SSC (Slope Sign Changes):** Registra los cambios de signo en la pendiente de la señal, capturando fluctuaciones y variaciones rápidas.

Complementariamente, se integraron características espectrales calculadas a través de la Transformada Rápida de Fourier (FFT), específicamente la potencia

espectral en bandas predefinidas y la frecuencia mediana del espectro. La concatenación de estas métricas conforma el vector de características ut, el cual actúa como la entrada exógena de la red NARX en el instante t, encapsulando el estado electromuscular completo de esa ventana temporal.

II-D3. Construcción de los vectores de entrada para la red NARX: El modelo NARX modela la salida continua y(t) como una función no lineal dependiente tanto de la historia reciente de los estímulos externos (entradas exógenas) como de la retroalimentación de sus propios estados pasados. En este esquema, los vectores de características extraídos de los canales Ch1 y Ch2 operan como las variables exógenas del sistema. Matemáticamente, la predicción en el instante t se formula como:

$$y_t = F(u^{(1)}{}_{t-1}, \dots, u^{(1)}{}_{t-n_u}, u^{(2)}{}_{t-1}, \dots, u^{(2)}{}_{t-n_u}, y_{t-1}, \dots, y_{t-n_y}), \quad (4)$$

donde $u_t^{(1)}$ y $u_t^{(2)}$ corresponden a los vectores de características de ambos músculos, n_u define el número de retardos temporales de entrada, n_γ establece el tamaño de la ventana de retroalimentación de la salida, y F(·) representa la función de mapeo no lineal aproximada por la arquitectura neuronal. A partir de estos parámetros de retardo se concatena el vector de estado x_t, el cual condensa la información espacio-temporal requerida para estimar el ángulo articular.

La función F(·) fue implementada a través de un Perceptrón Multicapa (MLP, por sus siglas en inglés). La red recibe el vector x_t y genera una predicción angular escalar continua (en grados). La arquitectura interna está compuesta por capas densas activadas mediante funciones ReLU, optimizadas con el algoritmo Adam minimizando el Error Cuadrático Medio (MSE) respecto a los ángulos reales.

Para garantizar la capacidad de generalización del modelo, la partición de los registros rotulados (archivos con ángulos y movimientos conocidos) se ejecutó mediante una asignación aleatoria, estableciendo la siguiente distribución:

- **Entrenamiento (70 %):** Datos empleados para el ajuste iterativo de los pesos neuronales.
- **Validación (10 %):** Subconjunto extraído del entrenamiento inicial, utilizado estrictamente para monitorizar el criterio de parada temprana (Early Stopping) y prevenir el sobreajuste.
- **Prueba (20 %):** Datos completamente aislados (no vistos) durante las fases de entrenamiento y validación, destinados a evaluar el desempeño final en inferencia.

El rendimiento en la etapa de prueba se cuantificó utilizando el Error Absoluto Medio (MAE). Adicionalmente, como aporte analítico de este trabajo y considerando que el objetivo práctico del control es alcanzar estados articulares específicos, la precisión de la regresión continua se evaluó también mediante métricas de exactitud (accuracy) acotadas a bandas de tolerancia de error predefinidas.

II-D4. Entrenamiento del modelo neuronal: Acorde al protocolo experimental, se entrenaron cuatro configuraciones de redes neuronales, una para cada plano de movimiento articular: flexión/extensión del codo (M1), flexión/extensión del hombro (M2) y aducción/abducción vertical del hombro (M3), mapeados para los ángulos estáticos de 30°, 60°, 90°, 120° y 150°; y un modelo para la aducción/abducción horizontal del hombro (M4) para los ángulos de 18°, 36°, 54°, 72° y 90°. Dado que la naturaleza estocástica del potencial de acción muscular es invariable independientemente del grupo muscular, se mantuvo una arquitectura neuronal unificada para los cuatro modelos. Los parámetros n_u, n_γ y los hiperparámetros de ventana se ajustaron empíricamente, incrementando progresivamente la profundidad temporal hasta lograr el balance óptimo entre precisión predictiva y costo computacional. La implementación de los modelos se realizó en lenguaje Python apoyada en el módulo de redes neuronales de scikit-learn. Los hiperparámetros definitivos empleados se resumen en la Tabla VI.

TABLA VI: HIPERPARÁMETROS DE ARQUITECTURA Y OPTIMIZACIÓN PARA LOS MODELOS NARX-MLP. FUENTE: ELABORACIÓN PROPIA.

Hiperparámetro	Valor asignado (Scikit-learn)
Capas ocultas (neuronas)	(128, 64, 32)
Función de activación	ReLU
Optimizador de pesos	Adam
Tasa de aprendizaje	10^{-3}
Regularización L2	$\alpha = 10^{-3}$
Algoritmo Dropout	No implementado
Tamaño de lote (batch)	Automático
Épocas máximas	600
Parada temprana (Early stop)	Activada
Tasa de validación interna	10%
Paciencia de iteración	15 épocas

Las curvas de aprendizaje resultantes se presentan en la Fig. 8, donde se evidencia que los cuatro modelos (M1 a M4) alcanzaron una convergencia estable en menos de 70 épocas, minimizando rápidamente el error cuadrático medio sin presentar signos de divergencia entre los conjuntos de entrenamiento y validación.

Para complementar la evaluación continua, se generaron matrices de confusión utilizando una tolerancia de umbral de ±5° respecto a los ángulos discretos objetivo (Fig. 9). Este enfoque clasificador permite identificar los patrones de error estacionario de la red NARX. De su análisis se infiere que las predicciones son sustancial-

mente más precisas en los rangos articulares inferiores a 90°. Por el contrario, la mayor tasa de confusión cruzada se presenta en la transición de posiciones extendidas, siendo particularmente evidente en todos los modelos la dificultad para discriminar entre los ángulos de 120° y 150°.

Este comportamiento se cuantifica de manera explícita en los diagramas de barras de exactitud de la Fig. 10. Los modelos M3 y M4 (correspondientes a los planos frontal y transversal del hombro) exhiben el mayor índice de confiabilidad general. El modelo M1 (codo) presenta el mayor margen de error en sus estimaciones. No obstante, es imperativo destacar que el rendimiento global es altamente satisfactorio, logrando aciertos por encima del 67 % en la peor de las posiciones articulares extremas, y alcanzando picos de exactitud de hasta el 95 % en las posiciones intermedias de 40° a 60°.

II-D5. Evaluación de desempeño en inferencia: Dado que el objetivo primario para el control robótico consiste en alcanzar las consignas discretas del movimiento sin comprometer la estabilidad en el transitorio continuo, los resultados de la inferencia final sobre el conjunto de prueba se caracterizaron mediante bandas de tolerancia (ACC). La Tabla VII compila esta evaluación autorregresiva, reportando tanto el Error Absoluto Medio (MAE) general como la tasa de acierto dentro de umbrales angulares incrementales de error permitido.

TABLA VII: DESEMPEÑO INFERENCIAL DE LOS MODELOS NARX EVALUADO MEDIANTE MAE GLOBAL Y UMBRALES DE EXACTITUD CRUZADA. FUENTE: ELABORACIÓN PROPIA.

Modelo	MAE (°)	ACC ±5°	ACC ±6°	ACC ±7°	ACC ±10°
M1	11,66	0,55	0,62	0,68	0,79
M2	8,64	0,53	0,62	0,7	0,83
M3	6,59	0,66	0,72	0,79	0,89
M4	5,79	0,7	0,75	0,81	0,88

El análisis de estos indicadores confirma que la articulación del codo (M1) es biomecánicamente la más susceptible al error de mapeo, registrando el MAE más elevado (11,66°) y las menores tasas de exactitud. Simultáneamente, se demuestra una correlación directa entre la amplitud de la banda de tolerancia y la robustez de la predicción; al relajar el umbral de aceptación a ±10°, la capacidad predictiva de la red oscila entre el 79 % y el 89 %. Esta flexibilidad en la evaluación demuestra que la arquitectura propuesta es plenamente capaz de reconstruir la intención de movimiento de la extremidad humana y generar comandos angulares estables para la actuación electromecánica de la plataforma robótica y su entorno simulado.

II-E. Simulación del brazo robótico bioinspirado

Para el desarrollo de la simulación computacional, se implementó el modelo cinemático descrito en la Sección II-A, integrando los parámetros geométricos y articulares específicos del brazo humano. La animación en el entorno virtual se generó evaluando progresivamente los valores articulares finales mediante una estrategia de interpolación lineal, lo cual garantiza una transición de movimiento suave y continua entre las distintas configuraciones espaciales. El vector de ángulos articulares finales, que actúa como consigna de la simulación, se define como:

$$\mathbf{q}_{\text{brazo, final}} = \{\theta_{AHH}, \theta_{AVH}, \theta_{FH}, \theta_{FC}, \theta_{RM}\}, \quad (5)$$

mientras que las longitudes de los eslabones que conforman la cadena cinemática, correspondientes a los segmentos óseos del húmero (a_h) y del conjunto radio/cúbito (a_CR), se establecen como constantes fijas expresadas en el vector.

$$\mathbf{L}_{\text{brazo}} = \{a_H, a_{CR}\}. \quad (6)$$

Para ilustrar el procedimiento cinemático, a continuación se detallan las condiciones iniciales del modelo de Denavit-Hartenberg (DH) configuradas para ejecutar el movimiento de aducción y abducción horizontal del hombro. Al definir los parámetros $a_1 = 0$, $\alpha_1 = 90°$, $d_1 = 0$ y el ángulo variable $\theta_AHH = q_{brazo(1)}$, la matriz de transformación homogénea local resultante se formula como:

$$A_1 = \begin{bmatrix} \cos(\theta_{AHH}) & 0 & \sin(\theta_{AHH}) & 0 \\ \sin(\theta_{AHH}) & 0 & -\cos(\theta_{AHH}) & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7)$$

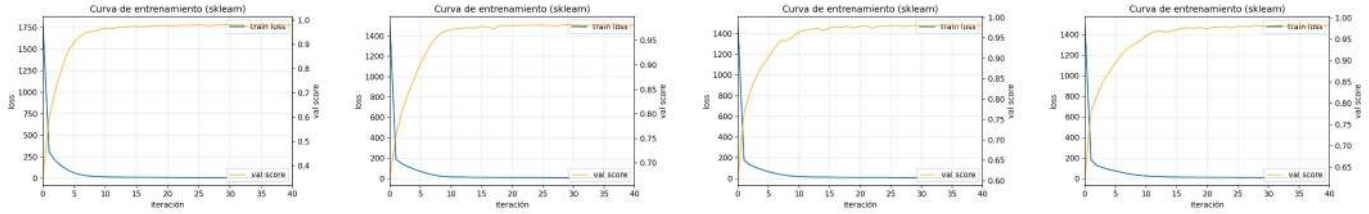
En consecuencia, la matriz de transformación compuesta global equivale a:

$$MA = A_1, \quad (8)$$

lo que permite extraer las coordenadas cartesianas tridimensionales del punto extremo del eslabón mediante la cuarta columna de la matriz de transformación, dada por:

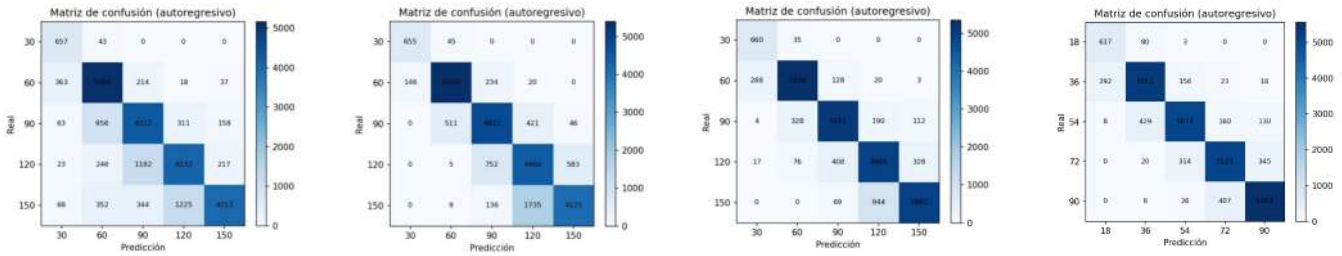
$$P_1 = MA(1:3, 4). \quad (9)$$

La arquitectura lógica que rige esta simulación se resume en el Algoritmo 1, el cual describe la rutina implementada en el entorno MATLAB® para comandar



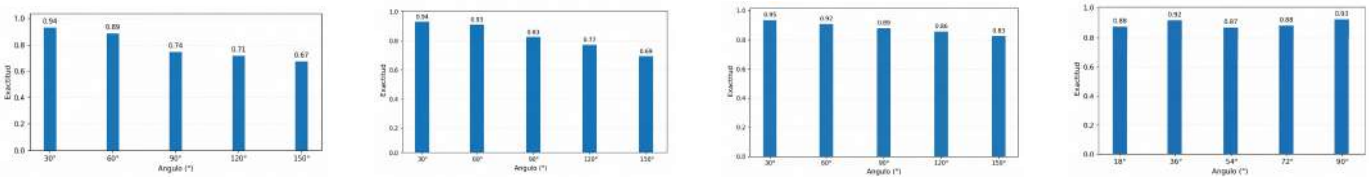
(a) Modelo M1 (b) Modelo M2 (c) Modelo M3 (d) Modelo M4

Figura 8: Curvas de aprendizaje (pérdida del MSE) durante las fases de entrenamiento y validación para los cuatro modelos. Fuente: elaboración propia.



(a) Modelo M1 (b) Modelo M2 (c) Modelo M3 (d) Modelo M4

Figura 9: Matrices de confusión para la estimación de ángulos estacionarios con banda de tolerancia cruzada de ±5°. Fuente: elaboración propia.



(a) Modelo M1 (b) Modelo M2 (c) Modelo M3 (d) Modelo M4

Figura 10: Porcentaje de exactitud en la predicción angular agrupado por posturas objetivo para los modelos M1 a M4. Fuente: elaboración propia.

simultáneamente el gemelo virtual y el brazo robótico Kinova Gen3 Lite. La entrada principal (input) de esta secuencia de control son los ángulos articulares estimados previamente por la red neuronal NARX.

Desglosando el modelo, el primer grado de libertad (q_1) rige la aducción y abducción horizontal del hombro en el plano transversal, denotada como θ_{AHH}. En esta formulación geométrica, una condición inicial de $\theta_{AHH} = 0°$ posiciona el brazo derecho alineado paralelamente a la columna vertebral, mientras que $\theta_{AHH} = -180°$ genera la postura especular para el brazo izquierdo. Al establecer $\theta_{AHH} = -90°$, el brazo adquiere una extensión frontal ortogonal al cuerpo; desde esta postura, la elevación y descenso de la extremidad son gobernados por el ángulo de flexión-extensión del hombro (θ_FH). En consecuencia, cualquier variación de θ_FH a partir de dicha orientación rige el movimiento estrictamente en el plano sagital, independientemente de la lateralidad (derecha o izquierda) de la extremidad.

El segundo grado de libertad (q_2) corresponde a la aducción y abducción vertical del hombro (θ_AVH), cutada sobre el plano frontal. La matriz de rotación de θ_AVH comparte la misma estructura matemática que la de θ_AHH, diferenciándose únicamente por el cambio de plano aplicado mediante el parámetro de torsión α. Sin embargo, cuando θ_AVH asume un valor no nulo, se genera un desfase o acoplamiento espacial entre el ángulo de flexión/extensión anatómico puro y el ángulo mecánico resultante en el modelo. Por ejemplo, si la articulación presenta θ_AVH = 150° y se comanda una flexión de hombro de 10°, la orientación geométrica final respecto al eje vertical será de 160°. De igual manera, una configuración concurrente de θ_AVH = 90° y θ_FH = 90° colapsa en una orientación efectiva de 180°. Este fenómeno evidencia que el parámetro θ_FH se evalúa referenciado a un sistema de coordenadas que previamente ya ha sido rotado por el estado de q_2.

Con el objetivo de desacoplar estos movimientos y garantizar que θ_FH se comporte como un ángulo absoluto de flexión y extensión referenciado directamente a la postura anatómica neutral, fue necesario aplicar una compensación algebraica a la variable θ_1 en su respectiva matriz DH.

Algoritmo 1 Integración de la simulación virtual y control físico del brazo robótico Kinova Gen3 Lite

Fase de Inicialización

Establecer conexión de red UDP y establecer enlace con la API del robot Kinova.
Definir configuración articular base qbase, índice de articulación objetivo j y factor de suavizado dinámico $\alpha \in (0,1]$.
Cargar longitudes antropométricas de los eslabones del brazo {lk} y extremidad distal {dk}.
Inicializar variables angulares: $\theta robot \leftarrow 0$, $\theta sim \leftarrow 0$.

Bucle de Control Principal

while el sistema se encuentre activo **do**

— Recepción y procesamiento de comandos —

if existen paquetes de datos disponibles en el buffer UDP **then**
Extraer el ángulo estimado θemg del datagrama.
if θemg está dentro de los límites anatómicos válidos **then**
$\theta robot \leftarrow \theta emg$
Construir vector de comando articular $q \leftarrow qbase$ asignando $q[j] \leftarrow \theta robot$
Enviar comando de posición q a los servomotores del robot Kinova.
end if
end if

— Actualización de la cinemática del simulador —

Calcular interpolación angular: $\theta sim \leftarrow \theta sim + \alpha \square (\theta robot - \theta sim)$.
Interpolar estado angular del brazo y efector final para el frame actual.
Actualizar vector de simulación $qsim[j] \leftarrow \theta sim$.

— Resolución de Cinemática Directa (DH) —

Computar matrices de transformación Ti para hombro, codo y muñeca.
Extraer posiciones espaciales (3D) {pi} de cada junta desde las matrices {Ti}.

— Motor de Renderizado —

Proyectar eslabones del brazo geométricamente (cilindros y esferas articulares).
for cada segmento distal (dedo) d **do**
Calcular transformaciones de las falanges (base, proximal y distal) de d.
Renderizar la geometría del segmento d.
end for
Refrescar buffer de video y actualizar ventana de simulación gráfica.
Introducir retardo Δt ▷ Garantizar sincronismo en la frecuencia de control
end while

Cuando el eslabón se encuentra en su posición nominal ($q_{brazo}(3) = 0$), se conserva la relación directa.

$$\theta_1 = q_{\text{brazo}}(3) + 90^\circ. \tag{10}$$

Por el contrario, cuando existe un desplazamiento previo ($q_{brazo}(3) \neq 0$), la matriz se alimenta con la expresión compensada.

$$\theta_1 = q_{\text{brazo}}(3) - q_{\text{brazo}}(2) + 90^\circ. \tag{11}$$

Esta sustracción anula la perturbación posicional introducida por q2, garantizando que el movimiento sagital de flexión y extensión permanezca referenciado al sistema absoluto. Gracias a este artificio matemático, un incremento unitario en θ_{FH} producirá siempre la misma magnitud de desplazamiento angular sobre el brazo físico, sin importar la orientación transversal impuesta por θ_{AVH}.

En conjunto, la selección cuidadosa de estos parámetros articulares y la inyección de offsets compensatorios aseguran que la cinemática del modelo computacional mantenga una fidelidad estricta anatómica bioinspirada. Simultáneamente, esta formulación permite mapear los comandos hacia los actuadores del brazo robótico Kinova Gen3 Lite, cerrando la brecha entre el modelo matemático, la simulación virtual (ver Fig. 11) y la actuación electromecánica real.

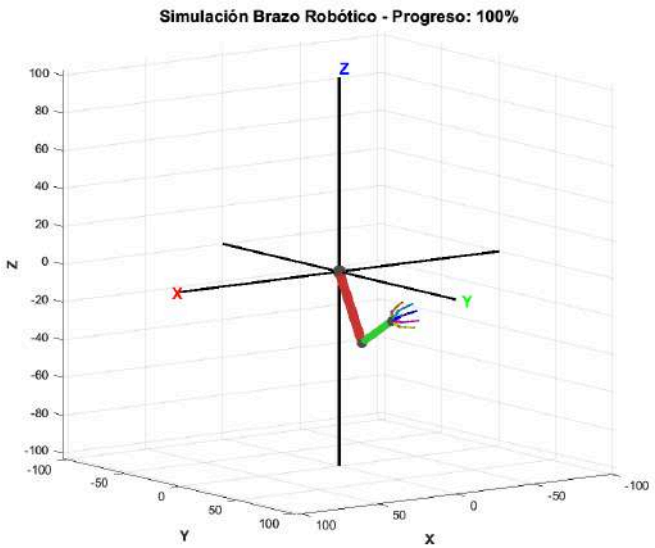


Figura 11: Captura del entorno MATLAB® renderizando el 100 % de la interpolación cinemática para un estado final $q_{brazo\ final}$ = [−90°, 150°, 150°, −90°, 5°]. Fuente: elaboración propia.

II-F. Visión por computadora

Como estrategia complementaria para la validación del análisis biomecánico y de las señales EMG, se implementó un módulo de visión por computadora fundamentado en el entorno de estimación de pose MediaPipe [25]. Este sistema de captura óptica permite cuantificar en tiempo real los ángulos articulares ejecutados por el participante. La información geométrica extraída opera como un sistema de referencia independiente (ground truth) para validar las estimaciones angulares inferidas por la red neuronal NARX.

El modelo computacional detecta y rastrea los vértices articulares clave del hombro y del codo. A partir de la extracción de las coordenadas bidimensionales en el plano de la imagen, el sistema calcula los vectores direccionales que representan la orientación espacial de los segmentos del brazo y antebrazo. La Fig. 12 ilustra la superposición de este esqueleto topológico y la respectiva lectura angular sobre el participante en una postura de calibración.

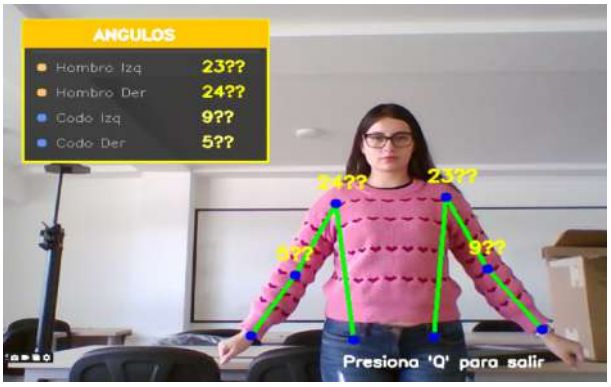


Figura 12: Visualización por computadora de la estimación de pose del participante. Fuente: elaboración propia.

El cálculo cinemático de los ángulos se efectúa evaluando tres puntos de referencia detectados por MediaPipe en torno a la articulación de interés. Matemáticamente, el ángulo se determina mediante el producto punto de los vectores convergentes. En términos generales, si P1, P2 y P3 representan tres coordenadas nodales consecutivas, donde P2 corresponde al vértice o pivote articular, el ángulo geométrico interno θ_geom se define como:

$$\theta_{\text{geom}} = \arccos\left(\frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\|\|\vec{v}_2\|}\right), \tag{12}$$

donde los vectores direccionales de los segmentos adyacentes son $\vec{v1}$ = P1 − P2 y $\vec{v2}$ = P3 − P2.

Para el análisis específico del hombro, se seleccionaron como nodos de referencia la cadera, el hombro (vértice P2) y el codo. Consecuentemente, el ángulo se computa a partir de los vectores trazados desde el hombro hacia la cadera y hacia el codo, respectivamente. La magnitud resultante describe el ángulo geométrico formado entre el tronco y el brazo proyectado sobre el plano frontal de la imagen.

Por su parte, para la articulación del codo, el sistema monitorea los nodos del hombro, el codo (vértice P2) y la muñeca. La medición angular se extrae evaluando los vectores direccionales originados en el codo hacia los extremos proximal y distal. Dado que el algoritmo geométrico tiende a reportar valores cercanos a 180° cuando la extremidad se encuentra completamente alineada o extendida, fue necesario aplicar una transformación para mapear esta lectura al marco anatómico clínico estándar de flexión-extensión, en el cual la extensión máxima equivale a 0°. Por lo tanto, el ángulo anatómico rectificado del codo se modela mediante la siguiente transformación lineal:

$$\theta_{\text{codo}} = 180^\circ - \theta_{\text{geom}}. \tag{13}$$

Esta conversión garantiza que la parametrización del movimiento sea estrictamente coherente con la convención anatómica clásica, donde el gradiente angular aumenta de forma monótona a medida que el codo flexiona el antebrazo sobre el brazo.

Con el objetivo de proveer una interfaz visual intuitiva (ver Fig. 13), se desarrolló una aplicación personalizada en OpenCV que superpone la telemetría en tiempo real sobre la transmisión de la cámara. La pantalla del usuario renderiza lo siguiente:

- Coordenadas articulares de cadera, hombro, codo y muñeca.
- Vectores alámbricos que representan los segmentos biomecánicos del tronco, brazo y antebrazo.
- Etiquetas con los valores numéricos correspondientes a los ángulos estimados en tiempo real.

Adicionalmente, el software incorpora un módulo de captura de estados estáticos. Mediante un evento de teclado (tecla G), el investigador registra de forma instantánea el estado angular capturado por la cámara, empaquetando esta información en el vector articular de referencia:

$$\vec{q}_{\text{cámara}} = [0, 0, \theta_{\text{hombro}}, \theta_{\text{codo}}, 0]. \tag{14}$$

Este vector se despliega en la interfaz como un patrón de control o validación cruzada para contrastar la exactitud de los resultados arrojados por el modelo neuronal NARX.

Para extender este marco de validación a la plataforma robótica Kinova Gen3 Lite, se optó por el rastreo

fiduciario de marcadores ArUco. Estos códigos matriciales de alto contraste están optimizados para algoritmos de visión por computadora [26]. La adopción de esta técnica responde a que el núcleo de MediaPipe está entrenado exclusivamente para reconstruir topologías corporales humanas, resultando inviable para la detección morfológica directa de la estructura del robot.

La instrumentación óptica del robot requirió la fijación de tres marcadores en eslabones estratégicos. Para rastrear la cinemática del codo artificial, los códigos se adhirieron en las carcasas de los actuadores 2, 3 y 5. Simultáneamente, el seguimiento del hombro se posibilitó colocando marcadores en la base estacionaria y en los actuadores 2 y 3. La Fig. 13 muestra la detección geométrica de esta configuración en el entorno virtual.

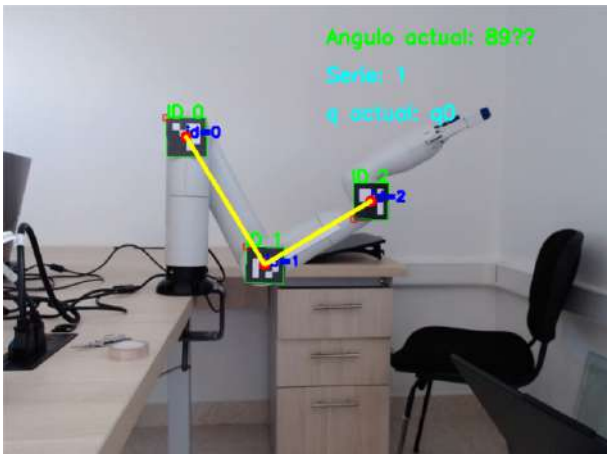


Figura 13: Sistema de captura óptica y rastreo de marcadores ArUco implementado para auditar la cinemática del robot Kinova Gen3 Lite. Fuente: elaboración propia.

El ensamble de este ecosistema de visualización computacional garantiza la trazabilidad cruzada del experimento, permitiendo comparar milimétricamente el gesto biológico original, la reconstrucción mioeléctrica predicha por la EMG y la respuesta mecánica del actuador robótico final. Del mismo modo, el sistema permite evidenciar posturas compensatorias no deseadas, como la basculación involuntaria del hombro o rotaciones axiales parásitas durante la ejecución. En síntesis, este módulo de visión actúa como la herramienta definitiva de auditoría cuantitativa para corroborar el desempeño y la viabilidad del control basado en redes NARX.

II-G. Gen3 Lite de Kinova

El brazo robótico Gen3 Lite de Kinova está compuesto por seis grados de libertad y un efector final. En esta investigación se restringen (no se usan) los grados de libertad correspondientes a la muñeca y al efector final; es decir, los actuadores 4, 5 y 6 y el gripper (ver Fig. 14 para el detalle de todos los actuadores del robot). En consecuencia, se utilizan únicamente los ángulos articulares q_1, q_2 y q_3 asociados a los actuadores 1, 2 y 3, respectivamente. En comparación con el modelo simulado y real de brazo desarrollado en este trabajo, estos grados de libertad corresponden a los ángulos q_{AHH}, q_{FH} y q_{FC}.

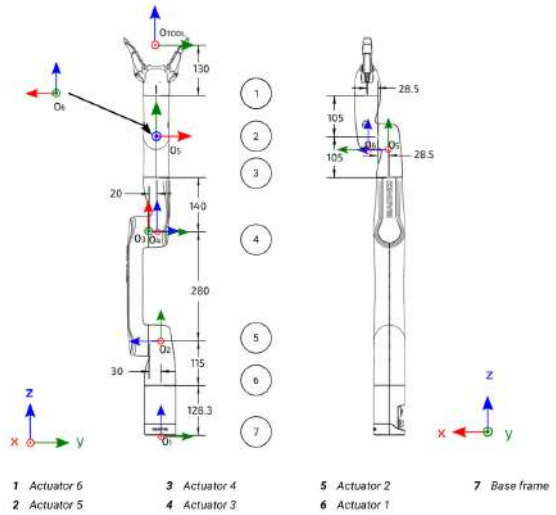


Figura 14: Parámetros DH para el Gen3 Lite y definición de marcos de referencia. Fuente: [27].

La Fig. 14 muestra los actuadores del robot Kinova Gen3 Lite y los marcos de referencia de las articulaciones humanas del brazo. En esta investigación, se controlan exclusivamente los actuadores 1 (base), 2 (hombro) y 3 (codo), los cuales se encuentran enmarcados con los números 6, 5 y 4, respectivamente. La manera de controlar el robot es similar a la usada en la simulación; los movimientos se generan a partir de un vector q que almacena vectores angulares y que se envían desde MATLAB®.

III. RESULTADOS EXPERIMENTALES

Para evaluar el desempeño integral del sistema propuesto, compuesto por el modelo predictivo NARX, el entorno de simulación, el control en lazo abierto y la instrumentación por visión computacional, se diseñó un protocolo experimental en un entorno semicontrolado. Esta aproximación garantizó la supervisión clínica de los movimientos sin imponer restricciones mecánicas externas de movilidad, con el fin de emular las condiciones dinámicas de un escenario real. Las señales mioeléctricas capturadas se vectorizaron y se inyectaron en la red NARX, la cual estimó en tiempo real las coordenadas angulares objetivo, transmitiendo estos comandos simultáneamente al gemelo virtual (simulación) y al manipulador Kinova Gen3 Lite.

En paralelo, el módulo de visión computacional actuó como sistema de auditoría transversal. Dicho sistema registró vectores angulares independientes en cada iteración del experimento, los cuales se contrastaron contra

las estimaciones generadas por la red NARX para cuantificar el Error Absoluto Medio (MAE). Dado que la simulación y el robot físico operan simultánea y homólogamente bajo la misma formulación cinemática, el módulo de captura óptica también validó la exactitud mecánica final ejecutada por los actuadores del robot, aislando así el error de inferencia del error mecánico.

Los hallazgos de esta experimentación demuestran de manera empírica la viabilidad de utilizar señales EMG como entradas exógenas en modelos recurrentes para controlar plataformas robóticas basadas en cinemática directa. Este resultado fundamenta la hipótesis de que es técnica y biológicamente plausible aprovechar los biopotenciales residuales de la musculatura de personas con amputaciones. Mediante arquitecturas de deep learning previamente entrenadas, estas bioseñales podrían decodificarse para replicar de forma parcial o total la funcionalidad mecánica de las articulaciones ausentes.

III-A. Validación cinemática del manipulador robótico

La Fig. 15 ilustra las gráficas de dispersión obtenidas al correlacionar las mediciones ópticas del módulo de visión con los comandos cinemáticos programados en el robot Kinova. Se evaluaron las articulaciones análogas: codo (M1, Fig. 15a), hombro en plano sagital (M2, Fig. 15b) y hombro en plano transversal (M4, Fig. 15c). La estrecha alineación de los datos a la bisectriz confirma una relación lineal altamente consistente (aislando desviaciones atípicas u outliers). Esto evidencia que los actuadores mecánicos reproducen las trayectorias estimadas con un margen de error instrumental inferior a 2,5°. Los valores exactos del MAE para el sistema físico se reportan en la Tabla VIII.

TABLA VIII: ERROR ABSOLUTO MEDIO (MAE) ENTRE LA ACTUACIÓN DEL MANIPULADOR ROBÓTICO Y EL MÓDULO DE VISIÓN COMPUTACIONAL. FUENTE: ELABORACIÓN PROPIA.

Articulación objetivo	MAE (°)
Codo	2,38
Hombro (Sagital)	1,03
Hombro (Transversal)	0,33

III-B. Validación biomecánica y propagación de error humano

Por su parte, la Fig. 16 contrasta la predicción de la red NARX directamente contra el movimiento libre ejecutado por el participante humano para los cuatro modelos articulares. A diferencia del sistema robótico estructuralmente rígido, la dispersión biológica es notablemente superior. Este comportamiento es inherente a la naturaleza de la captura óptica sobre la anatomía humana sin puntos de fijación mecánicos externos, que hacen que el movimiento biológico no se restrinja matemáticamente a un plano cartesiano puro, sino que está supeditado al control postural, oscilaciones estabilizadoras tridimensionales y el juego articular intrínseco de cada participante. Cuantitativamente, como se observa en la Tabla IX, la desviación (MAE) alcanza sus máximos en los planos sagital y frontal del hombro (> 20°), seguidos de la articulación del codo (12,46°).

TABLA IX: ERROR ABSOLUTO MEDIO (MAE) ENTRE LA CINEMÁTICA HUMANA Y LA MEDICIÓN POR CÁMARA. FUENTE: ELABORACIÓN PROPIA.

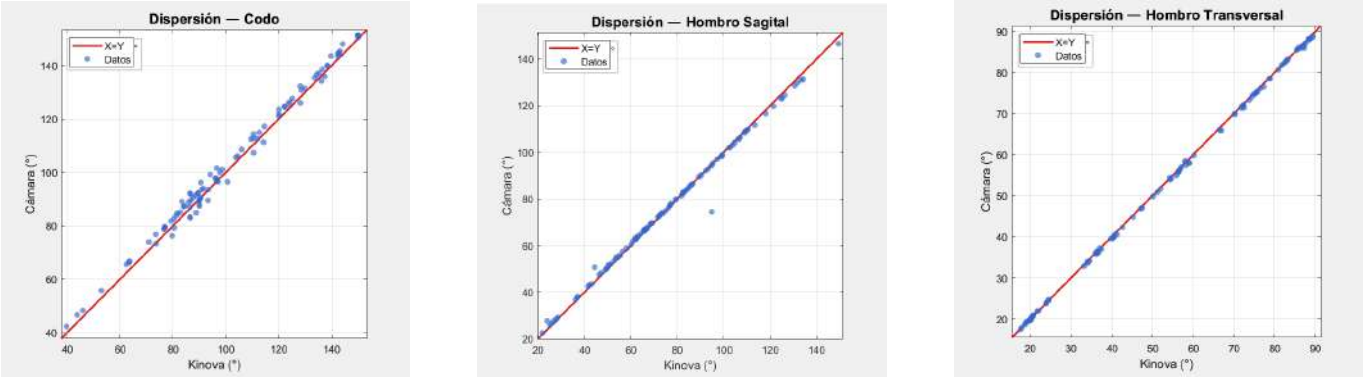
Modelo analizado	MAE (°)
M1 (Codo)	12,46
M2 (Hombro Sagital)	20,14
M3 (Hombro Frontal)	20,47
M4 (Hombro Transversal)	7,21

IV. DISCUSIÓN

El análisis integral de los resultados experimentales revela desafíos biomecánicos clave en la instrumentación mioeléctrica. Particularmente, la tendencia al alza del MAE en el modelo M1 se explica por la naturaleza biarticular y multifuncional del bíceps braquial. Este músculo interviene simultáneamente en la flexión del codo, la supinación del antebrazo y la estabilización glenohumeral, por lo que la magnitud de su activación electromiográfica no está correlacionada de forma unívoca y exclusiva con el ángulo del codo [28], [29]. A esto se suma que la ausencia de soportes isométricos incrementa sustancialmente la demanda estabilizadora y la variabilidad inter-repetición [30]. La combinación de estos factores diluye la correspondencia directa entre el espectro mioeléctrico y la intención angular final, introduciendo un ruido biológico que hace más complejo el proceso de regresión [31], [32].

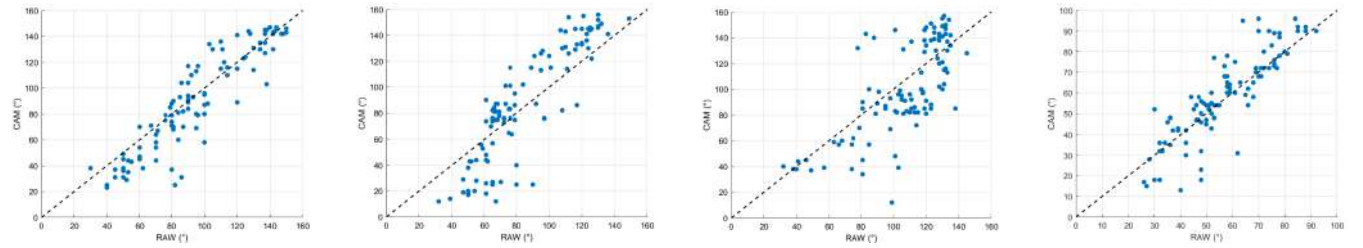
IV-A. Nivel de Madurez Tecnológica (TRL)

Atendiendo a la escala Technology Readiness Level (TRL), la cual clasifica el grado de desarrollo de una innovación técnica [33], este sistema se enmarca en un TRL 3. El proyecto clasifica como una prueba de concepto analítica y experimental probada en laboratorio [34], en virtud de estar fundamentada en una validación con hardware propietario de captura, una muestra demográfica inicial de usuarios y sesiones únicas de adquisición. La Tabla X contextualiza el desempeño de esta arquitectura NARX frente a otras aproximaciones del estado del arte que comparten un nivel de madurez similar (TRL 3 y 4).



(a) Articulación del codo (M1) (b) Hombro — plano sagital (M2) (c) Hombro — plano transversal (M4)

Figura 15: Gráficas de dispersión del control cinemático ejecutado por el robot Kinova Gen3 Lite vs. lectura de visión computacional. Fuente: elaboración propia.



(a) Codo (M1) (b) Hombro Sagital (M2) (c) Hombro Frontal (M3) (d) Hombro Transv. (M4)

Figura 16: Gráficas de dispersión del movimiento humano real (módulo de cámara) vs. estimación EMG (red NARX). Fuente: elaboración propia.

TABLA X: COMPARATIVA DE DESEMPEÑO FRENTE A ESTUDIOS DE LA LITERATURA CON UN TRL EQUIVALENTE. FUENTE: ELABORACIÓN PROPIA.

Referencia	TRL	Arquitectura y Articulación	Métrica de Error
Este trabajo	3	NARX+MLP (Hombro, Codo)	MAE 5,7°–11,6°
Suplino et al. [35]	4	NARX (Codo)	RMSE 7°
Wang et al. [36]	4	GA-Elman (Hombro, Codo)	RMSE 3,4°–4,1°
Liu et al. [37]	3-4	NARX (Hombro, Codo, Muñeca)	VAF > 98 %

Es fundamental interpretar comparativamente estas métricas dadas sus diferencias algorítmicas. Aunque el MAE y el RMSE se expresan en unidades angulares equivalentes, el RMSE penaliza de forma cuadrática las divergencias atípicas, resultando sistemáticamente en valores superiores o iguales al MAE para una misma población de datos [38]. Desde esta óptica matemática, el MAE reportado en este trabajo para la estimación del hombro (4,09°–5,63°) es altamente competitivo y se homologa funcionalmente con los valores RMSE publicados por Wang et al. [36] y Suplino et al. [35] (3,47°–7°).

No obstante, en la articulación del codo, la desviación experimental (11,66°) excede los umbrales de la literatura, evidenciando que el error proviene de una limitación de diseño experimental más que de un artefacto de la métrica (como se discutió respecto a la influencia estabilizadora del bíceps). Por su parte, la métrica VAF (Variance Accounted For) empleada por Liu et al. [37] equivale estadísticamente al coeficiente de determinación (R^2); si bien reportan un éxito superior al 98 %, esta magnitud indica proporción de varianza explicada y no es analíticamente traducible a error angular continuo sin conocer la varianza total poblacional del experimento original [39].

IV-B. Limitaciones del estudio y prospectiva

Las discrepancias y los márgenes de error observados, lógicos dentro de un desarrollo de alcance TRL 3, se atribuyen a dos limitaciones críticas del entorno experimental actual:

- **Instrumentación muscular subactuada.** La topología de sensores para el modelo M1 omitió el registro del tríceps (antagonista principal), priorizando los canales de adquisición sobre el supinador para expandir, a futuro, los modelos cinemáticos de la mano. Sin embargo, esta decisión restringió la capacidad de la red para ponderar la co-contracción

fisiológica natural (agonista-antagonista) durante las flexo-extensiones del brazo humano [40], [41].

- **Muestra poblacional y temporalidad de la captura.** El corpus de entrenamiento consta de 800 perfiles mioeléctricos extraídos de una muestra transversal de 10 participantes en sesiones únicas de corta duración. En contraste con bancos de datos biomédicos longitudinales maduros como Ninapro [42], el volumen actual limita la generalización del modelo frente a la adaptación neuromuscular intra-sujeto (fatiga a largo plazo) y la variabilidad fenotípica inter-sujeto (densidad de tejido adiposo, conductividad epidérmica, volumen muscular).

El desempeño del modelo NARX presentado constituye una validación analítica exitosa del control mioeléctrico sobre cadenas cinemáticas en robótica. Superar los umbrales de exactitud actuales requerirá obligatoriamente migrar hacia niveles TRL superiores, lo cual implica instrumentar arreglos musculares completos (agonistas, antagonistas y sinergistas) y expandir el espectro demográfico de las campañas de adquisición de datos.

V. CONCLUSIONES

En esta investigación se consolidó exitosamente la base metodológica y tecnológica para el control de un brazo robótico bioinspirado, partiendo del desarrollo de un modelo cinemático riguroso de la extremidad superior humana. Para incluir en el modelado los datos fisiológicos reales, se diseñó e implementó un sistema de adquisición en hardware propio denominado Freya EMG Sensor. Mediante la ejecución de un protocolo experimental estandarizado con la participación de diez sujetos, se logró conformar un conjunto de datos compuesto por 800 señales electromiográficas (EMG) crudas, estableciendo un conjunto de datos fundamental para el análisis biomecánico y el entrenamiento computacional.

El núcleo analítico del proyecto se materializó a través del diseño y entrenamiento de una red neuronal recurrente tipo NARX. El procesamiento de las señales EMG mediante esta arquitectura de deep learning permitió consolidar cuatro modelos predictivos independientes (uno para cada movimiento articular evaluado). Estos modelos demostraron una capacidad notable para mapear la actividad muscular no lineal y reconstruirla en trayectorias angulares discretas. De este modo, se alcanzó el objetivo principal de la investigación, que consistió en controlar en lazo abierto de forma fluida la simulación de un brazo robótico bioinspirado, traduciendo directamente la intención motora humana en comandos cinemáticos viables.

Para garantizar la rigurosidad técnica, se incorporó paralelamente un módulo de visión por computadora que actuó como sistema independiente de validación de los ángulos ejecutados. La integración final demostró que la información inferida por la red NARX es plenamente funcional para comandar, de forma simultánea y en lazo abierto, tanto el gemelo virtual como la plataforma robótica física Kinova Gen3 Lite. Estos resultados evidencian el enorme potencial de este enfoque para futuras aplicaciones en control protésico y rehabilitación neuromotora. Con el ánimo de fomentar el desarrollo colaborativo en esta área de estudio, la totalidad de los recursos, códigos y bases de datos asociados al proyecto es de acceso público en el repositorio de GitHub [43].

AGRADECIMIENTOS

Los autores expresan su agradecimiento a la Universidad de Nariño, de manera particular al Departamento de Electrónica y al Grupo de Investigación en Ingeniería Eléctrica y Electrónica (GIIEE), por el respaldo académico y administrativo brindado para el desarrollo de este trabajo. Asimismo, extienden su sincera gratitud a todos los voluntarios, cuya disposición y participación en los experimentos hicieron posible la consolidación del conjunto de datos fundamental para esta investigación.

REFERENCIAS

[1].C. L. McDonald, S. Westcott-McCoy, M. R. Weaver, J. Haagsma, and D. Kartin, “Global prevalence of traumatic non-fatal limb amputation,” Prosthetics and Orthotics International, vol. 45, no. 2, pp. 105–114, 2021.

[2].Departamento Administrativo Nacional de Estadística (DANE), “Por conflicto armado, según departamentos y municipios,” 2010, cuadro 43.

[3].——, “Resultados preliminares de la implementación del registro para la localización y caracterización de las personas con discapacidad departamento de Nariño,” Bogotá, p. 16, Aug. 2007.

[4].S. Salminger, A. Sturma, C. Hofer et al., “Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance?” Disability and Rehabilitation, 2020.

[5].E. A. Biddiss and T. T. Chau, “Upper limb prosthesis use and abandonment: a survey of the last 25 years,” Prosthetics and Orthotics International, vol. 31, no. 3, pp. 236–257, 2007.

[6].A. Álvarez, “Desarrollo de una prótesis mioeléctrica,” Tesis de Ingeniería Mecatrónica, Centro de Enseñanza Técnica Industrial (CETI), Guadalajara, México, 2020, pp. 9–11.

[7].B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” IEEE Transactions on Biomedical Engineering, vol. 40, no. 1, pp. 82–94, 1993.

[8].K. Englehart and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control,” IEEE Tran-

sactions on Biomedical Engineering, vol. 50, no. 7, pp. 848–854, 2003.

[9].E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," Journal of Rehabilitation Research and Development, vol. 48, no. 6, pp. 643–660, 2011.

[10].M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," Frontiers in Neurorobotics, vol. 10, p. 9, 2016.

[11].A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in emg pattern recognition-based control," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 12, pp. 2242–2251, 2018.

[12].A. Suberbiola, "Estudio de la relación entre señales emg y el ángulo de la articulación controlada," Ph.D. dissertation, Universidad del País Vasco, 2022. [Online]. Available: https://www.ehu.eus/ccwintco/uploads/6/67/Tesis doctoral suberbiola v13 (final).pdf

[13].B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, Robotics: Modelling, Planning and Control, ser. Advanced Textbooks in Control and Signal Processing. London: Springer-Verlag, 2009.

[14].A. Cohen, "Biomedical signals: Origin and dynamic characteristics; frequency-domain analysis," in The Biomedical Engineering Handbook, 2nd ed., J. D. Bronzino, Ed. Boca Raton: CRC Press LLC, 2000.

[15].Electrotens, "Colocación de los electrodos según los músculos del cuerpo," https://electrotens.com/colocacion-de-los-electrodos-segun-los-musculos-del-cuerpo/, 2025, accessed: 2025-11-14.

[16].P. Aagaard, E. B. Simonsen, J. L. Andersen, P. Magnusson, and P. Dyhre-Poulsen, "Coactivation of the elbow antagonist muscles is not affected by the speed of movement in isokinetic exercise," Muscle & Nerve, vol. 34, no. 2, pp. 241–248, 2006. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16307438/

[17].M. Gazzoni, N. Celadon, D. Mastrapasqua, M. Paleari, V. Margaria, and P. Ariano, "Quantifying forearm muscle activity during wrist and finger movements by means of multichannel electromyography," PLoS ONE, vol. 9, no. 10, p. e109943, 10 2014. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0109943

[18].Depositphotos, "Illustrations of resistance band exercises for upperlimb movement guidance," https://depositphotos.com/, 2025, includes illustrations of bicep curl, shoulder exercise, upper back reverse exercise, and lateral exercise. Accessed: 2025-11-14.

[19].J. Won and M. Iwase, "Highly responsive robotic prosthetic hand control considering electrodynamic delay," Sensors, vol. 25, no. 1, p. 113, 2024.

[20].R. Raj and K. Sivanandan, "Comparative study on estimation of elbow kinematics based on emg time domain parameters using neural network and anfis narx model," Journal of Intelligent & Fuzzy Systems, vol. 32, no. 1, pp. 791–805, 2017.

[21].F. Davarinia and A. Maleki, "Comparing the efficiency of recurrent neural networks to emg-based continuous estimation of the elbow angle," Neural Computing and Applications, vol. 36, no. 29, pp. 18 515–18 530, 2024.

[22].J. Sandoval, "Estimación de velocidad de la mano a partir de señales electromiográficas superficiales usando redes neuronales narx," Revista Ingeniería Biomédica, vol. 4, no. 8, pp. 39–46, 2010.

[23].C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," Journal of Experimental Social Psychology, vol. 49, no. 4, pp. 764–766, 2013.

[24].A. Ullah, S. Iqbal, and G. Jeon, "Effect of analysis window and feature selection on classification of hand movements using emg signal," in Advances in Computer Science and Ubiquitous Computing, ser. Lecture Notes in Electrical Engineering, vol. 702. Springer, 2020, pp. 325–334.

[25].Google AI Edge, "Pose landmark detection guide," 2026, accessed: 2026-04-27. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/pose landmarker

[26].OpenCV, "Detection of aruco markers," 2026, openCV Documentation, accessed 2026-04-27. [Online]. Available: https://docs.opencv.org/4.x/d5/dae/tutorial aruco detection.html

[27].Kinova Inc., KINOVA® Gen3 Lite Robot User Guide, Kinova Robotics, 2021, Together in robotics – Meet Gen3 Lite, revision R3.1. [Online]. Available: https://www.kinovarobotics.com

[28].F. P. Kendall, E. K. McCreary, and P. G. Provance, Muscles: Testing and Function with Posture and Pain, 5th ed. Baltimore, MD, USA: Lippincott Williams & Wilkins, 2005.

[29].D. C. Preston and B. E. Shapiro, Electromyography and Neuromuscular Disorders: Clinical-Electrophysiologic Correlations, 3rd ed. Philadelphia, PA, USA: Elsevier, 2013.

[30].D. A. Winter, Biomechanics and Motor Control of Human Movement, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2009.

[31].Noraxon USA Inc., ABC of EMG: A Practical Introduction to Kinesiological Electromyography, Noraxon USA Inc., Scottsdale, AZ, USA, 2014.

[32].R. Merletti and P. A. Parker, Electromyography: Physiology, Engineering, and Non-Invasive Applications. Hoboken, NJ, USA: John Wiley & Sons, 2004.

[33].National Aeronautics and Space Administration, "Technology readiness levels," https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/, 2023, accedido: agosto de 2026.

[34].NASA Earth Science Technology Office, "Technology readiness levels (trls)," https://esto.nasa.gov/trl/, 2020, accedido: agosto de 2026.

[35].L. O. Suplino, G. C. de Melo, G. S. Umemura, and A. Forner-Cordero, "Elbow movement estimation based on emg with narx neural networks," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 3767–3770.

[36].J. Wang, Q. Hao, X. Xi, J. Cao, A. Xue, and H. Wang, "Estimation of continuous joint angles of upper limb based on semg by using gaelman neural network," Computational Intelligence and Neuroscience, vol. 2020, p. 4065351, 2020.

[37].J. Liu, S. H. Kang, D. Xu, Y. Ren, S. J. Lee, and L.-Q. Zhang, "Emg-based continuous and simultaneous estimation of arm kinematics in able-bodied individuals and stroke survivors," Frontiers in Neuroscience, vol. 11, p. 480, 2017.

[38].T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)? arguments against avoiding rmse in the literature," Geoscientific Model Development, vol. 7, no. 3, pp. 1247–1250, 2014.

[39].L. N. Wimalasena, J. F. Braun, and Keshtkaran, "Estimating muscle activation from emg using deep learning-based dynamical systems models," Journal of Neural Engineering, vol. 19, no. 3, p. 036013, 2022.

[40].Q. Ding, A. Xiong, X. Zhao, and J. Han, "A novel emg-driven state space model for the estimation of continuous joint movements," Sensors, vol. 21, no. 13, p. 4966, 2021.

[41].P.-L. Kuo, D.-H. Lee, D. L. Jindrich, and N. Dounskaia, "Contribution of muscle synergies to biomechanical interaction effects during arm movements," Journal of Motor Behavior, vol. 47, no. 2, pp. 143–163, 2015.

[42].M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," Scientific Data, vol. 1, p. 140053, 2014.

[43].J. P. Eraso Sara, "Simulation of a bioinspired robotic arm controlled using deep learning techniques," https://github.com/SaraEraso/Bioinspired-Deep-Learning-Arm, 2026, accedido: 15-ene-2026.



Universidad de **Nariño**

FUNDADA EN 1904