

Ingeniería e Investigación
Journal

Abbreviated Journal Title: **Ing. Investig.**

Editor-in-chief

Sonia C. Mangones, PhD

Associate Editor

Andrés Pavas, PhD, MSc

Technical Editor

Lenin Alexander Bulla Cruz, PhD, MSc

Editorial Assistants

Julian Arcila-Forero, MSc, BSc

Ingri Gisela Camacho, BSc

Editorial Board

Paulo César Narváez Rincón, PhD

Universidad Nacional de Colombia - Bogotá

Cynthia Anbuselvi PhD

SEA College of Engineering and Technology, Indian

Fabio González, PhD

Universidad Nacional de Colombia - Bogotá

Weronika Kruszelnicka PhD

Bydgoszcz University of Science and Technology, Poland

Jayant Rajgopal, PhD

University of Pittsburgh, United States

Hilda Elizabeth Reynel-Ávila PhD

Instituto Tecnológico de Aguascalientes, Mexico

Gerasimos G. Rigatos PhD

Unit Industrial Automation, Industrial Systems Institute, Greece

Paulina Jaramillo PhD

Carnegie Mellon University, United States

Fausto Andrés Molina Gómez PhD

Universidad Politécnica de Madrid, Spain

Aquiles Enrique Darghan Contreras, PhD

Universidad Nacional de Colombia, Colombia

Frequency

Continuous periodicity (three issues per year)

Cover Layout and Community Manager

Ian Felipe Guarnizo Martínez

Proofreader

José Daniel Gutiérrez-Mendoza

Layout Artist

David Mauricio Valero

For additional information, please contact:

revij_bog@unal.edu.co

Bogotá - Colombia

2025

Table of Contents

Editorial

Artificial Intelligence in Engineering Education: Enhancing Human Capabilities, Not Replacing Them

Inteligencia artificial en la educación en ingeniería: potenciar lo humano, no reemplazarlo

Liz Karen Herrera-Quintero, Jenny Marcela Sánchez-Torres, Lorena Chaparro-Díaz, Sonia Patricia Carreño-Moreno, Yuly Edith Sánchez-Mendoza

Agricultural Engineering

A Hybrid Voting Ensemble Model for the Efficient Sorting and Classification of Date Fruit Varieties

Un modelo híbrido de votación por conjunto para la clasificación y ordenamiento eficiente de variedades de dátiles

Sofiane Abden, Mostefa Bendjima, Soumia Benkrama

Chemical, Food, and Environmental Engineering

Evaluation of the Known Sub-Sequence Algorithm (KSSA) for Optimal Use in Time Series of Tuna Fishing in the Pacific Ocean

Evaluación del algoritmo de subsecuencias conocidas (KSSA) para su uso óptimo en series de tiempo de pesca de atún en el Océano Pacífico

Julían Gómez, Iván Benavides, John Selvaraj

Assessing the Interaction between Water Erosion and SOC Storage in a Small Mexican Watershed

Evaluación de la interacción entre la erosión hídrica y el almacenamiento de COS en una microcuenca mexicana

Olimpya Talya Aguirre-Salado, Joel Pérez-Nieto, Carlos A. Aguirre-Salado, Alejandro Ismael Monterroso Rivas

Activation of Coal Fly Ash for Cadmium Wastewater Remediation

Activación de ceniza volante de carbón para la remediación de aguas residuales contaminadas con cadmio

Yareli Ginalis Medina, Roberto Flores

Bioelectricity Generation in Established Cacao (*Theobroma cacao*), Oil Palm (*Elaeis guineensis*), and Peruvian Amazon Grass (*Axonopus compressus*) Crops: Insights from Amazonian Soils

Generación de bioelectricidad en cultivos establecidos de cacao (*Theobroma cacao*), palma aceitera (*Elaeis guineensis*) y grama amazónica peruana (*Axonopus compressus*): perspectivas desde suelos amazónicos

Edwar Edinson Rubina-Arana, Letty Leonor Sandoval-Mendoza, Glendy Sánchez-Sunción, Dalia Carbonel, Grober Panduro-Pisco

Civil and Sanitary Engineering

Spatial and Temporal Analysis of Road Transport Emissions in Colombia

Análisis espacial y temporal de las emisiones del transporte carretero en Colombia

Julieth Alfonso, Dayron Bermúdez, Sonia Mangones, Nestor Rojas, Carlos Rivera, Aquiles Darghan, Mauricio Osses

Application of Graphene Oxide Nanoparticles to Cementitious Composites to Mitigate the Effects of Attacks by Aggressive Agents

Aplicación de nanopartículas de óxido de grafeno en compuestos cementosos para mitigar los efectos del ataque de agentes agresivos

Samuel Castro-Lopes, Barbara Simões, Sergio Peres, Viviane Rodrigues, Tiago Santos, Romildo Berenguer

Electrical, Electronic and Telecommunications Engineering

Active Disturbance Rejection Control of a DC Brushed Motor Using Simulink and Raspberry Pi

Control por rechazo activo de perturbaciones de un motor DC con escobillas utilizando Simulink y Raspberry Pi

Mario E. González N., Óscar H. Sierra H., Óscar I. Higuera M.

**Facultad de Ingeniería
Universidad Nacional de Colombia**

Sonia Esperanza Monroy Varela
Dean

Mario Enrique Velásquez Lozano
Vice Dean of Research and Extension

Iván Darío Gil Chaves
Vice Dean of Academic Affairs

Gloria Inés Beltrán Calvo
Director of the Students Welfare Service

Papers published in *Ingeniería e Investigación* are abstracted/indexed in

- Science Citation Index Expanded
- (SciSearch®), Clarivate Analytics
- Scopus - Elsevier
- Scientific Electronic Library Online - SciELO, Colombia
- Chemical Abstract
- Índice de Revistas Latinoamericanas
- en Ciencias Periódica
- Dialnet
- Sistema Regional de Información en Línea para
- Revistas Científicas de América Latina, el Caribe, España y Portugal - Latindex
- Ebsco Publishing
- DOAJ - Directory of Open Access Journals
- Redib - Red Iberoamericana de Innovación y Conocimiento Científico

The journal *Ingeniería e Investigación* was created in 1981. It is an entity in charge of spreading the teaching, Scientific, and technical research conducted at Universidad Nacional de Colombia's Department of Engineering and other national and international institutions. *Ingeniería e Investigación* deals with original, unedited scientific research and technological developments in the various disciplines related to engineering. *Ingeniería e Investigación* contributes the development of knowledge, generating a global impact on academia, industry, and society at large through an exchange of knowledge and ideas while maintaining a set of serious and recognized quality standards.

The content of the articles published in this journal does not necessarily reflect the opinions of the Editorial Team. These texts can be totally or partially reproduced provided a correct citation of the source.

Ingeniería e Investigación publications are developed for the academic community who is interested in research and engineering knowledge development. We invite readers to be part of this Journal and participate either as authors, peer reviewers, or subscribers.

For additional information, please contact:

www.revistas.unal.edu.co/index.php/ingevinv
E-mail: revii_bog@unal.edu.co
Tel: 57(1) 3 16 5000 Ext. 13674

Lightweight Deep Learning for Atrial Fibrillation Detection: Efficient Models for Wearable Devices

Aprendizaje profundo ligero para la detección de fibrilación auricular: modelos eficientes para dispositivos portátiles
Carlos A. Fajardo, Andrés S. Parra, Tania V. Castellanos-Parada

Mechanical Engineering, Mechatronics, and Materials Science

Influence of the Synthesis Route on the Structural and Magnetic Properties of La_{0.7}Sr_{0.3}MnO₃ Nanoparticles
Influencia de la ruta de síntesis en las propiedades estructurales y magnéticas de nanopartículas de La_{0.7}Sr_{0.3}MnO₃
Jenny Alejandra Mera Córdoba, Jhon F. Betancur-Pérez, Nicolás A. Salazar-Henao, Luis Alejandro Galeano, Andrés Rosales-Rivera, Diego F. Coral-Coral

Effect of Ship Size on the Hull Dimension Ratios of Purse Seiners
Efecto del tamaño de la embarcación sobre los coeficientes de dimensiones del casco de los barcos cerqueros
Dennys Dunker De La Torre Cortez

Development of Auto-Injection Systems through the TRIZ Problem-Solving Method
Desarrollo de sistemas de autoinyección mediante el método de resolución de problemas TRIZ
Burhan Şahin

Systems and Computer Engineering

Development of a Solid Waste Collector Robot for Cleaning in Public Areas
Desarrollo de un robot recolector de residuos sólidos para limpieza en áreas públicas
Dick Díaz-Delgado, Alexander Eduardo Inga Alva

Education in Engineering

Understanding How Software Project Managers Learn Using Kolb's Learning Style Inventory
Entendiendo cómo aprenden los gestores de proyectos de software mediante el inventario de estilos de aprendizaje de Kolb
Mauricio Hidalgo, Laura M. Castro, Hernán Astudillo

Artificial Intelligence in Engineering Education: Enhancing Human Capabilities, Not Replacing Them

L. K. Herrera-Quintero¹, J. M. Sánchez-Torres², L. Chaparro-Díaz³, S. P. Carreño-Moreno⁴
and Y. E. Sánchez-Mendoza⁵

The irruption of artificial intelligence (AI) in higher education – especially its generative forms, *i.e.*, models capable of producing text, images, or code – has rapidly transformed pedagogical, administrative, and assessment practices in our institutions. We are facing a reconfiguration of the university space that cannot be reduced to the adoption of tools; instead, it demands a critical examination of its epistemic, ethical, and pedagogical implications. Generative AI has established a set of practices that, if not continuously monitored, could compromise the very essence of education in engineering and other disciplines [1].

Today, engineering, historically positioned as the engine of technical and scientific development, faces a paradox: leading innovation without succumbing to thoughtless automation. AI cannot be assumed to be a cognitive shortcut or a substitute for thinking, but rather an opportunity to deepen human capabilities, *i.e.*, discerning, questioning, and proposing. This tension has been extensively documented in recent scientific literature [2]–[4], which warns about the risk of erosion of critical skills due to the superficial, uncritical, and decontextualized use of AI.

At Universidad Nacional de Colombia, we have taken on this challenge as an opportunity to rethink our mission-related functions. The proposal we share today, developed collaboratively and involving different disciplines, is based on the conviction that it is possible and necessary to incorporate AI into public universities in a strategic, ethical, and situated manner [5]. This implies not only regulating its use but also educating, researching, and extending from a paradigm that reaffirms the value of critical thinking, transparency, and social responsibility.

Within this framework, we propose three lines of action that are closely aligned with our university's academic and social dynamics. The first of them, CUIDAMOS-IA, proposes a pedagogical approach that recognizes student agency, promotes formative assessments, and strengthens the

necessary competencies to understand and engage with AI without becoming subordinate to it. As reported in [6], the vieJIAS initiative, an interdisciplinary academic collective, has led critical AI appropriation processes at Universidad Nacional de Colombia. This experience has integrated interdisciplinary workshops, teacher training spaces, and the development of AI-based conversational agents, involving more than 180 members of the university community. Its ethical, pedagogical, and collaborative approach embodies the type of educational environment that we promote in this line of action. This initiative has enabled the practical application of principles such as care, critical reflection, and shared responsibility in the use of these technologies within the university setting.

The aim is to train men and women engineers who can effectively interact with advanced technologies while remaining mindful of the ethical dilemmas, epistemic gaps, and social consequences associated with the use of AI. To this effect, it is necessary to create a learning environment that anticipates AI practices, preserves academic integrity, and fosters critical thinking.

The proper use of AI in teaching becomes possible in an environment of reflective pedagogical practice, which materializes through actions such as the inclusion of a transparent statement in course *syllabi* that details the use of AI, emphasizing how, why, and for what purpose this technology is used and how it is useful, but also what its risks are. The reality of using AI in higher education, whether its use is authorized or not, forces teachers to rethink their content and pedagogical strategies so that personal authorship, critical reflection, and human supervision are demanded, as no task carried out with the support of AI should overlook at least two reviews, one by the author and one by the teacher. Understanding the potential risks of indiscriminate, unsupervised, or unreflective AI usage is fundamental to preventing its use as a shortcut to the cognitive and reflective processes necessary for genuine learning.

¹ Department of Mechanical and Mechatronics Engineering, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá Campus. E-mail: lk Herrera@unal.edu.co

² Department of Systems and Industrial Engineering, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá Campus. E-mail: jmsanchezt@unal.edu.co

³ Department of Nursing, Faculty of Nursing, Universidad Nacional de Colombia, Bogotá Campus. E-mail: olchaparro@unal.edu.co

⁴ Department of Nursing, Faculty of Nursing, Universidad Nacional de Colombia, Bogotá Campus. E-mail: spcarrenom@unal.edu.co

⁵ Department of Physics, Faculty of Science, Universidad Nacional de Colombia, Bogotá Campus. E-mail: yesanchezm@unal.edu.co

Implementing a training system with AI support and developing it within a context of care should go beyond mere control. On the contrary, the goal is to rebuild trust in pedagogy through innovation. Therefore, it is not advisable to act as judges or executors regarding the use of AI, nor to monitor its use through AI detection platforms, as these platforms have their own biases and are prone to false positives. Instead, from a caring perspective, it is advisable to promote critical competencies to compare, validate, and debate the production of AI, e.g., by proposing exercises in critique and contrast of AI and human products, transforming the classroom into an ethical, reflective, and metacognitive laboratory.

The second line, REFLEXIONAMOS-IA, corresponds to the field of research. Through a set of recommendations organized around reflection, we have built an institutional guide that emphasizes ethical responsibility, academic integrity, transparency in documenting the use of AI, the prevention of AI-mediated plagiarism, and the care of sensitive data. This proposal does not intend to replace the researcher's judgment, but to strengthen it through clear criteria that position AI as a support tool, never as a hidden author or an intellectual replacement [5], [7].

Reflecting on AI support in research involves placing integrity and transparency at the center of the scientist's work. Among the processes that could be expedited with the use of AI are literature reviews and the projection of hypotheses and analysis pathways, for which it is necessary to thoroughly document the prompts used as well as each human intervention aimed at achieving the results. By explicitly declaring AI-mediated procedures, the traceability and reproducibility of research processes are favored, in addition to ensuring intellectual sovereignty.

REFLEXIONAMOS-IA is not only based on norms but also on the imperative of continuous training for research teams within a process of critical AI literacy, where the technical functioning of AIs is understood, as well as their limitations and political, economic, social, technological, environmental, and legal risks. In this sense, it is essential to take direct actions to avoid dependence on AI and, conversely, to promote a critical autonomy that encourages using AI as an assistant rather than a leader in generating knowledge. Thus, critical and reflective AI-supported research must take place in an environment nourished by human judgment, creativity, awareness, and responsibility.

In this context, it is essential to provide examples of how AI is used in engineering. Some of these include its application in additive manufacturing, automated microstructural analysis using computer vision, and the optimization of thermal and mechanical processes in materials engineering. These cases exemplify how AI can function as a technical assistant in simulation, classification, quality control, or experimental design tasks without replacing the researcher's judgment. Its inclusion strengthens the bridge between ethical reflection and everyday scientific practice in the field of engineering.

The third line, TEJEMOS-AI, responds to the need to think about university extension, also called the *third mission*, from an intercultural and territorial perspective [5]. Active student participation in co-creation processes with communities can be a key component of this line of action. Faced with the temptation of homogeneous and decontextualized solutions, we propose the use of AI in service of the social and environmental challenges of our regions. Critical extension is not limited to technological transfer; it enables co-creation processes with communities, recognizing their knowledge, languages, and needs. This is, perhaps, the most urgent dimension, given the deep digital inequalities that persist in our Latin American reality.

In this vein, proposals for the third university mission should begin with a participatory diagnosis that identifies the gaps and needs of communities. During these processes, AI can serve as a data collection, analysis, and reporting assistant, in addition to participating in the proposed solutions, generating responses to address the problems identified – that said, everything must be tailored to the level of AI access and appropriation exhibited by the community with which the work will be carried out. Weaving contextually appropriate solutions for and with communities assumes that AI is a bridge and not a barrier, always prioritizing the shared responsibility between the university and the community.

Weaving, as an exercise in AI-assisted co-creation between the university and the community, is a process that must take place within the framework of technological justice, wherein open and free access platforms should be used, preferably AI-supported solutions under the domain and governance of the university that aim for the social appropriation of knowledge, which transcends its mere transfer and advances towards the articulation of diverse, contextual knowledge based on real and experienced problems of the community. Thus, it is important to evaluate the impact of these initiatives through co-created indicators, document successful experiences, and ensure economic and environmental sustainability from the outset, which allows scaling up good practices and closing digital gaps. TEJEMOS-IA views AI as a tool to enhance care, solidarity, and innovation. Its aim is to build a living fabric of collaboration between academia and the community, which shapes citizens capable of transforming their reality with judgment and empathy.

These proposals do not emerge from improvisation; they are based on a comparative review of more than 350 policies and institutional guidelines on AI, in dialogue with frameworks such as the UNESCO recommendation [8], Colombia's CONPES 4144 [9], and other directives that guide the responsible use of emerging technologies. What distinguishes our approach is its internal coherence: five strategic axes (governance, training, curriculum, innovation, and sustainability) allowing for a progressive and transversal institutional implementation that is also capable of self-regulation.

Public universities cannot limit themselves to a passive adaptation to technological advancement; they are called to exercise ethical, pedagogical, and epistemic leadership. Therefore, incorporating AI cannot be a technical decision, but rather a collective, deliberative process that strengthens the innovative and ethical horizon of the university. Engineering, as an innovative discipline, holds an irreplaceable responsibility: to develop solutions that work and make sense, that build the future without relinquishing critical awareness [5].

If this moment teaches us anything, it is that the most powerful intelligence is not the one that automates at the highest speed, for the power lies within human beings who can stop and think, prioritizing situated and rigorous intelligence. This will, without a doubt, be reflected in the commitment to the responsible practice of engineering in the 21st century.

Final note. Members of the vieJIAs collective collaboratively wrote this document. This is a group of women professors from the Bogotá campus of Universidad Nacional de Colombia who are committed to a critical, pedagogical, and ethical reflection on the use of AI in university contexts.

REFERENCES

- [1] A. Abulibdeh, C. Chatti, A. AlKhereibi, and S. Menshawy, "A scoping review of the strategic integration of artificial intelligence in higher education: Transforming university excellence themes and strategic planning in the digital era," *Eur. J. Edu.*, vol. 60, no. 1, art. e12908, 2025. <https://doi.org/10.1111/ejed.12908>
- [2] S. Khairullah, S. Harris, H. Hadi, R. Sandhu, N. Ahmad, and M. Alshara, M. "Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in the higher education institutions," *Front. Edu.*, vol. 10, art. 1548104, 2025. <https://doi.org/10.3389/feduc.2025.1548104>
- [3] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," *Smart Learn. Environ.*, vol. 11, no. 1, 28, 2024.
- [4] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?" *Int. J. Educ. Tech. Higher Edu.*, vol. 16, no. 1, pp. 1-27, 2019.
- [5] J. M. Sánchez-Torres, L. Chaparro-Díaz, L. K. Herrera-Quintero, S. P. Carreño-Moreno, and Y. E. Sánchez-Mendoza, *Recomendaciones para el uso de la inteligencia artificial en actividades de docencia, investigación y extensión universitaria*. Bogotá DC, Colombia: Universidad Nacional de Colombia, 2025.
- [6] Agencia de Noticias UNAL, "ViejIAs lideran la apropiación crítica de la inteligencia artificial en la UNAL", Universidad Nacional de Colombia, 2025. [Online]. Available: <https://agenciadenoticias.unal.edu.co/detalle/viejias-lideran-la-apropiacion-critica-de-la-inteligencia-artificial-en-la-unal>
- [7] N. McDonald, A. Johri, A. Ali, and A. H. Collier, "Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines," *Comp. Human Behavior. Art. Humans*, vol. 3, art. 100121. <https://doi.org/10.1016/j.chbah.2025.100121>
- [8] UNESCO "Recommendation on the ethics of artificial intelligence," 2021. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [9] Departamento Nacional de Planeación, "Política nacional de transformación digital e inteligencia artificial, documento CONPES 4144," 2023. [Online]. Available: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf>

Inteligencia artificial en la educación en ingeniería: potenciar lo humano, no reemplazarlo

L. K. Herrera-Quintero¹, J. M. Sánchez-Torres², L. Chaparro-Díaz³, S. P. Carreño-Moreno⁴,
y Y. E. Sánchez-Mendoza⁵

La irrupción de la inteligencia artificial (IA) en la educación superior —en especial sus formas generativas, *i.e.*, los modelos capaces de producir textos, imágenes o código— ha transformado rápidamente las prácticas pedagógicas, administrativas y evaluativas en nuestras instituciones. Nos encontramos ante una reconfiguración del espacio universitario que no puede reducirse a la adopción de herramientas, sino que exige un examen crítico de sus implicaciones epistémicas, éticas y pedagógicas. La IA generativa ha instaurado una serie de prácticas que, si no se supervisan constantemente, podrían poner en peligro la mismísima esencia de la educación en ingeniería y otras disciplinas [1].

Hoy en día, la ingeniería, históricamente posicionada como motor del desarrollo técnico-científico, enfrenta una paradoja: liderar la innovación sin ceder a la automatización irreflexiva. La IA no puede asumirse como un atajo cognitivo ni como un sustituto del pensamiento, sino como una oportunidad para profundizar en las capacidades humanas, *i.e.*, discernir, cuestionar y proponer. Esta tensión ha sido extensamente documentada en la literatura científica reciente [2]–[4], que alerta sobre el riesgo de erosión de habilidades cruciales si se utiliza la IA de manera superficial, acrítica y descontextualizada.

Desde la Universidad Nacional de Colombia, hemos asumido este desafío como una oportunidad para repensar nuestras funciones misionales. La propuesta que presentamos hoy, creada de forma colaborativa y con diferentes disciplinas, se basa en la convicción de que es posible y necesario incluir la IA en la universidad pública de manera estratégica, ética y situada [5]. Esto implica no solo regular su uso, sino también formar, investigar y extender desde un paradigma que reafirme el valor del pensamiento crítico, la transparencia y la responsabilidad social.

En este marco, proponemos tres líneas de acción estrechamente articuladas con las dinámicas académicas y sociales de nuestra universidad. La primera de ellas,

CUIDAMOS-IA, plantea un enfoque pedagógico que reconoce la agencia estudiantil, promueve la evaluación formativa y fortalece las competencias necesarias para comprender y dialogar con la IA sin subordinarse a ella. Como se reportó en [6], la iniciativa vieJIAS, un colectivo académico interdisciplinario, ha liderado procesos de apropiación crítica de la IA en la Universidad Nacional de Colombia. Esta experiencia ha articulado talleres interdisciplinarios, espacios de formación docente y creación de agentes conversacionales con IA, involucrando a más de 180 integrantes de la comunidad universitaria. Su enfoque ético, pedagógico y colaborativo encarna el tipo de ambiente formativo que promovemos en esta línea. Esta iniciativa ha permitido poner en práctica los principios de cuidado, reflexión crítica y corresponsabilidad en el uso de estas tecnologías dentro del entorno universitario.

La meta es formar ingenieras e ingenieros que puedan interactuar de manera efectiva con tecnologías avanzadas sin perder de vista los dilemas éticos, los vacíos epistémicos y las consecuencias sociales asociadas al uso de la IA. Para ello, es necesario crear un ambiente de aprendizaje que se anticipe a las prácticas relacionadas con la IA, preserve la integridad académica y fomente el pensamiento crítico.

El uso adecuado de la IA en la docencia se viabiliza en un ambiente de práctica pedagógica reflexiva que se materializa a través de acciones tales como incluir en los *syllabi* de las asignaturas una declaración transparente con detalles sobre el uso de la IA, centrándose en cómo, por qué y para qué se usa esta tecnología y cuál es su utilidad, pero también cuáles son sus riesgos. La realidad del uso de las IAs en la formación, sea su uso autorizado o no, obliga al docente a repensar sus contenidos y estrategias pedagógicas, de manera que se exija la autoría personal, la reflexión crítica y la supervisión humana, pues ninguna tarea que se realice con apoyo de la IA debería pasar por alto al menos dos revisiones, la de su autor y la del profesor. Ser conscientes del riesgo del uso indiscriminado, sin supervisión e irreflexivo de la IA es fundamental para evitar que se utilice como un sustituto de

¹ Department of Mechanical and Mechatronics Engineering, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá Campus. E-mail: lkherreaq@unal.edu.co

² Department of Systems and Industrial Engineering, Faculty of Engineering, Universidad Nacional de Colombia, Bogotá Campus. E-mail: jmsanchezt@unal.edu.co

³ Department of Nursing, Faculty of Nursing, Universidad Nacional de Colombia, Bogotá Campus. E-mail: olchaparro@unal.edu.co

⁴ Department of Nursing, Faculty of Nursing, Universidad Nacional de Colombia, Bogotá Campus. E-mail: spcarrenom@unal.edu.co

⁵ Department of Physics, Faculty of Science, Universidad Nacional de Colombia, Bogotá Campus. E-mail: yesanchezm@unal.edu.co

los procesos de pensamiento y reflexión que son necesarios para un aprendizaje auténtico.

Implementar un sistema de formación con apoyo en IA y desarrollarlo en un contexto de cuidado debe ir más allá del mero control. Por el contrario, lo que se busca es reconstruir la confianza en la pedagogía mediante la innovación. Por ello, no es recomendable actuar como jueces o verdugos frente al uso de la IA, ni hacer vigilancia de su uso mediante plataformas detectoras de IA, ya que estas tienen sus propios sesgos y son susceptibles de falsos positivos. En su lugar, desde una visión del cuidado, lo recomendable es promover competencias críticas para comparar, validar y debatir la producción de la IA, e.g., proponiendo ejercicios de crítica y contraste de productos humanos y de IA, transformando el aula en un laboratorio ético, reflexivo y metacognitivo.

La segunda línea, REFLEXIONAMOS-IA, corresponde al ámbito de la investigación. Mediante un conjunto de recomendaciones organizadas en torno a la reflexión, hemos construido una guía institucional que hace énfasis en la responsabilidad ética, la integridad académica, la transparencia en el uso de IA, la prevención del plagio mediado por IA y el cuidado de los datos sensibles. Esta propuesta no pretende sustituir el juicio investigador, sino fortalecerlo mediante criterios claros que sitúen la IA como una herramienta de apoyo, nunca como autor oculto ni como reemplazo intelectual [5], [7].

Reflexionar acerca del apoyo de la IA en la investigación implica colocar la integridad y la transparencia en el centro del quehacer del científico. Dentro de los procesos que podrían agilizarse con el uso de IA están la revisión de literatura y la proyección de hipótesis y rutas de análisis, para los cuales se deben documentar detalladamente los *prompts* utilizados y cada intervención humana orientada hacia el logro de los resultados. Al declarar explícitamente los procedimientos mediados por IA, se favorece la trazabilidad y reproducibilidad de los procesos de investigación, además de asegurar la propiedad intelectual.

REFLEXIONAMOS-IA no solo se basa en normas, sino en el imperativo de formación continua del equipo de investigación en un proceso de alfabetización crítica de la IA donde se comprenda el funcionamiento técnico de las IAs, así como sus limitaciones y riesgos políticos, económicos, sociales, tecnológicos, ambientales y legales. Es fundamental tomar acciones directas para evitar la dependencia de la IA y, por el contrario, fomentar una autonomía crítica que promueva el uso de la IA como asistente y no como guía en la generación de conocimiento. Así, la investigación crítica y reflexiva con apoyo en IA debe darse en un ambiente nutrido por el juicio, la creatividad, la conciencia y la responsabilidad humanas.

En este contexto, es fundamental dar ejemplos de cómo se usa la IA en la ingeniería. Algunos de ellos incluyen su empleo en la fabricación aditiva, el análisis automático de microestructuras usando visión por computador y la optimización de procesos térmicos y mecánicos en

ingeniería de materiales. Estos casos ejemplifican cómo la IA puede actuar como asistente técnico en tareas de simulación, clasificación, control de calidad o diseño de experimentos sin reemplazar el juicio del investigador. Su inclusión fortalece el puente entre la reflexión ética y la práctica científica cotidiana en el campo de la ingeniería.

La tercera línea, TEJEMOS-IA, responde a la necesidad de pensar la extensión universitaria, también llamada *tercera misión*, desde una perspectiva intercultural y territorial [5]. La participación activa del estudiantado en procesos de co-creación con comunidades puede ser un componente clave de esta línea. Frente a la tentación de soluciones homogéneas y descontextualizadas, proponemos un uso de la IA al servicio de los desafíos sociales y ambientales de nuestras regiones. La extensión crítica no se limita a la transferencia tecnológica, sino que habilita procesos de co-creación con comunidades, reconociendo sus saberes, lenguajes y necesidades. Esta es, quizá, la dimensión más urgente, dadas las profundas desigualdades digitales que persisten en nuestra realidad latinoamericana.

En esta línea, las propuestas de la tercera misión universitaria deben partir de un diagnóstico participativo que identifique las brechas y las necesidades de las comunidades, procesos en los que la IA puede ser un asistente de captación, análisis y reporte de datos, además de participar en las soluciones propuestas, generando respuestas para hacer frente a los problemas identificados —eso sí, todo se debe adaptar al nivel de acceso y apropiación de IA que presente la comunidad con la que se va a trabajar. Tejer soluciones contextualmente apropiadas para y con las comunidades supone que la IA sea un puente y no una barrera, siempre privilegiando la corresponsabilidad entre la universidad y la comunidad.

Tejer, como un ejercicio de co-creación asistida por IA entre la universidad y la comunidad, es un proceso que debe darse en el marco de la justicia tecnológica, donde se usen plataformas de acceso abierto y gratuito, de preferencia soluciones apoyadas en IA que sean de dominio y gobierno de la universidad y que busquen la apropiación social del conocimiento, que trasciende la mera transferencia y avanza hacia la articulación de saberes diversos y contextuales basados en problemas reales y sentidos de la comunidad. Así, es crucial evaluar el impacto de estas iniciativas mediante indicadores co-creados, documentar experiencias exitosas y asegurar la sostenibilidad económica y ambiental desde la formulación, lo que permite escalar buenas prácticas y cerrar brechas digitales. TEJEMOS-IA ve la IA como una herramienta para mejorar el cuidado, la solidaridad y la innovación. Su meta es crear una red activa de colaboración entre académicos y la comunidad, formando a ciudadanos que puedan transformar su realidad con juicio y empatía.

Estas propuestas no emergen de la improvisación; se basan en un análisis de más de 350 políticas y pautas institucionales sobre IA, en conversación con marcos como la recomendación de la UNESCO [8], el CONPES 4144 de

Colombia [9] y otras guías que promueven el uso responsable de tecnologías emergentes. Lo que distingue nuestra ruta es su coherencia interna: cinco ejes estratégicos (gobernanza, formación, currículo, innovación y sostenibilidad) que permiten una implementación institucional progresiva, transversal y con capacidad de autorregulación.

La universidad pública no puede limitarse a adaptarse pasivamente al avance tecnológico; está llamada a ejercer liderazgo ético, pedagógico y epistémico. Por ello, incorporar la IA no puede ser una decisión técnica, sino un proceso colectivo y deliberativo que fortalezca el horizonte innovador y ético de la universidad. La ingeniería, como disciplina innovadora, tiene una responsabilidad intransferible: desarrollar soluciones que funcionen y tengan sentido, que construyan el futuro sin abdicar a la conciencia crítica [5].

Si algo nos enseña este momento es que la inteligencia más poderosa no es la que automatiza con mayor velocidad, pues el poder lo tienen los seres humanos que son capaces de detenerse y pensar, privilegiando la inteligencia situada y rigurosa. Esto, sin duda, se proyectará en el compromiso con el ejercicio responsable de la ingeniería en el siglo XXI.




Nota final. Este documento fue elaborado de manera colaborativa por integrantes del colectivo vieJIAS, un grupo de profesoras de la Universidad Nacional de Colombia, sede Bogotá, que están comprometidas con una reflexión crítica, pedagógica y ética sobre el uso de la IA en contextos universitarios.

REFERENCIAS

- [1] A. Abulibdeh, C. Chatti, A. AlKhereibi, and S. Menshawy, "A scoping review of the strategic integration of artificial intelligence in higher education: Transforming university excellence themes and strategic planning in the digital era," *Eur. J. Edu.*, vol. 60, no. 1, art. e12908, 2025. <https://doi.org/10.1111/ejed.12908>
- [2] S. Khairullah, S. Harris, H. Hadi, R. Sandhu, N. Ahmad, and M. Alshara, M. "Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in the higher education institutions," *Front. Edu.*, vol. 10, art. 1548104, 2025. <https://doi.org/10.3389/educ.2025.1548104>
- [3] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review," *Smart Learn. Environ.*, vol. 11, no. 1, 28, 2024.
- [4] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?" *Int. J. Educ. Tech. Higher Edu.*, vol. 16, no. 1, pp. 1-27, 2019.
- [5] J. M. Sánchez-Torres, L. Chaparro-Díaz, L. K. Herrera-Quintero, S. P. Carreño-Moreno, and Y. E. Sánchez-Mendoza, *Recomendaciones para el uso de la inteligencia artificial en actividades de docencia, investigación y extensión universitaria*. Bogotá DC, Colombia: Universidad Nacional de Colombia, 2025.
- [6] Agencia de Noticias UNAL, "ViejIAs lideran la apropiación crítica de la inteligencia artificial en la UNAL", Universidad Nacional de Colombia, 2025. [Online]. Available: <https://agenciadenoticias.unal.edu.co/detalle/viejias-lideran-la-apropiacion-critica-de-la-inteligencia-artificial-en-la-unal>
- [7] N. McDonald, A. Johri, A. Ali, and A. H. Collier, "Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines," *Comp. Human Behavior. Art. Humans*, vol. 3, art. 100121. <https://doi.org/10.1016/j.chbah.2025.100121>
- [8] UNESCO "Recommendation on the ethics of artificial intelligence," 2021. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [9] Departamento Nacional de Planeación, "Política nacional de transformación digital e inteligencia artificial, documento CONPES 4144," 2023. [Online]. Available: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/4144.pdf>

A Hybrid Voting Ensemble Model for the Efficient Sorting and Classification of Date Fruit Varieties

Un modelo híbrido de votación por conjunto para la clasificación y ordenamiento eficiente de variedades de dátiles

Sofiane Abden ¹, Mostefa Bendjima ², and Soumia Benkrama ³

ABSTRACT

Dates constitute one of Algeria's most important agricultural products, given their substantial health and economic advantages. Furthermore, they represent a vital export item beyond the hydrocarbon industry. The existing conventional techniques for classifying and sorting dates are ineffective, time-consuming, and labor-intensive, leading to a discrepancy between restricted exports and elevated production levels. This work presents an ensemble learning (EL) model that utilizes transfer learning (TL) strategies to overcome obstacles and improve date fruit classification. We assess the efficacy of four classifiers, *i.e.*, MobileNetV2, EfficientNet, DenseNet201, and an ensemble soft voting classifier that employs TL, with a dataset of 1619 photos representing 20 distinct varieties of Algerian dates. The dataset used is one of the greatest benchmarks for varietal diversity. The suggested hybrid model exhibits exceptional performance, with a validation accuracy of 99.07% and a classification accuracy of 99.93%. It establishes a new benchmark in agricultural technology by exceeding all assessed models in terms of precision, recall, and F1-score. These findings demonstrate the potential of this approach to completely transform date sorting and markedly improve agricultural production and efficiency.

Keywords: artificial intelligence, classification, machine learning, smart agriculture, date fruits

RESUMEN

Los dátiles constituyen uno de los productos agrícolas más importantes de Argelia debido a sus significativos beneficios para la salud y la economía. Además, representan un artículo de exportación vital más allá de la industria de los hidrocarburos. Las técnicas convencionales que existen para clasificar y ordenar dátiles son ineficaces, consumen mucho tiempo y requieren mucha mano de obra, lo que genera una discrepancia entre las exportaciones limitadas y los elevados niveles de producción. Este trabajo presenta un modelo de aprendizaje en conjunto (EL) que utiliza estrategias de aprendizaje por transferencia (TL) para superar obstáculos y mejorar la clasificación de frutos de dátiles. Evaluamos la eficacia de cuatro clasificadores, *i.e.*, MobileNetV2, EfficientNet y DenseNet201, y un clasificador de votación suave en conjunto que emplea TL, utilizando un conjunto de datos de 1619 imágenes que representan 20 variedades distintas de dátiles argelinos. El conjunto de datos utilizado es uno de los mayores referentes en diversidad varietal. El modelo híbrido propuesto presenta un rendimiento excepcional, con una precisión de validación del 99.07 % y una precisión de clasificación del 99.93 %, y establece un nuevo referente en la tecnología agrícola al superar a todos los modelos evaluados en términos de precisión, sensibilidad y puntuación F1. Estos hallazgos demuestran el potencial de este enfoque para transformar por completo la clasificación de dátiles y mejorar notablemente la producción y eficiencia agrícola.

Palabras clave: inteligencia artificial, clasificación, aprendizaje automático, agricultura inteligente, dátiles

Received: January 23rd, 2025

Accepted: May 14th, 2025

Introduction

Agriculture is fundamental to economic development and essential for sustaining the increasing global population [1]. Plants are crucial to ecosystems and human existence since they serve as the major source of food and provide essential resources such as oxygen and medicine. Nevertheless, the agricultural industry faces considerable hurdles [2], including satisfying increasing food demands and maintaining crop health in the face of evolving environmental conditions. Agricultural innovation is crucial for enhancing food security and economic stability, especially when it comes to improving crop quality and resilience through adequate plant health management [3].

The date palm (*Phoenix dactylifera* L.), one of the world's oldest cultivated fruit trees, originating from the region of Mesopotamia, has been grown since ancient times. With a robust root system that can reach considerable depths in well-drained, sandy soils, this tree thrives in the hot, arid

climates of the Middle East and North Africa [4]. Date palms are an essential crop in these regions, providing a vital source of food, income, and employment to millions. They hold not only economic value but also significant cultural and religious importance, making them a cornerstone of agricultural livelihoods in arid and semi-arid environments [5].

¹ Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: PhD student. Email: abdan.soufyane@univ-bechar.dz

² Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: Associate professor. Email: bendjima.mostefa@univ-bechar.dz

³ Department of Mathematics and Computer Science, Faculty of Exact Sciences, Laboratory of Information Processing and Telecommunications (LTIT), Tahri Mohammed University of Bechar, 08000, Algeria. Affiliation: associate professor. Email: benkrama.soumia@univ-bechar.dz



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Date palm fruits are a nutritious and affordable source of essential nutrients, providing numerous macro- and micro-elements. Rich in carbohydrates, particularly soluble sugars and dietary fiber [6], dates are an excellent source of energy. In addition to their high sugar content, they contain vital minerals, vitamins, and antioxidants, which contribute to their health benefits, such as supporting heart health and reducing the risk of diseases like cancer. This nutrient-dense profile makes dates an ideal, natural food with significant health-promoting properties [7].

Date palms hold great socioeconomic and environmental importance in many countries (e.g., Algeria). They ranked first in Saharan agriculture, with more than 17 million trees planted and more than 800 varieties, such as *Deglet Nour* (considered to be the best and most famous variety worldwide), *feggus*, *aadam*, *lanjouhar*, and *cherka* etc. Algeria occupies an important place among the date-producing and exporting countries [8]. However, there is a considerable gap between its production and exports due to its reliance on traditional production and sorting methods.

Sorting and classifying date palm fruits is traditionally done manually, a process that is labor-intensive, time-consuming, and costly, often requiring skilled workers. To address these challenges, there is a growing need for automated systems that can perform these tasks efficiently and accurately [6]. As a result, many studies are exploring intelligent methods for the rapid and fully automated sorting of date palm fruits, aiming to improve speed and accuracy and reduce dependence on manual labor.

Advancements in AI-driven image classification have enabled a highly precise, cost-effective, and rapid evaluation of date palm fruits, even under challenging environmental conditions [3]. Consequently, industries are increasingly adopting computer vision techniques to automatically grade and sort dates based on features such as color, texture, and size, tasks that were previously performed manually. With numerous types of date fruits displaying subtle differences in color, shape, and fleshiness, quality and price are often assessed through physical attributes, while nutritional value is evaluated based on chemical composition and sensory qualities, largely influenced by the fruit's variety and ripeness [9].

The goal of this study was to leverage machine learning (ML), particularly TL and EL, to enhance agriculture by using images of date palm fruits to identify and sort twenty different local date varieties (*Ajina*, *Deglet*, *Tanslit*, *Hamraya*, etc.). This classification was achieved by extracting features such as color, size, and shape from the images, facilitating an efficient and precise sorting. Unlike previous works, which largely focused on individual ML or TL models using public datasets from other countries, this study addressed a gap by introducing EL, combining CNNs pretrained on a local dataset. Several challenges were encountered during this study, including the large number of species used and the great similarity between some of them. Moreover, the scarcity of information on palm diseases compared to those of other plants is likely due to their occurrence in specific geographical areas. The main contributions of this paper include the application of a hybrid method combining TL on MobileNet, EfficientNet, and DenseNet201, as well as our proposal of EL on the same TL models. The performance of our EL approach, based on a voting meta-classifier, was

compared against the TL results using data collected from local markets of the Touggourt region, located in the south of Algeria.

Related works

With a view to improve efficiency in the food and agricultural sector, this section describes previous studies that have used AI approaches, specifically ML and deep learning (DL), to automate the categorization and sorting of date fruit species by means of images.

The authors of [7] suggested using supervised and unsupervised DL networks to classify different types of date fruit. They achieved effective feature fusion and dimensionality reduction by fusing features from a CNN (VGG-F) and PCANet using discriminant correlation analysis (DCA). The accuracy of their method was 99.32% with VGG-F (CCA) and 98.20% with VGG-F (DCA). To test the system, the authors presented one of the largest benchmark datasets used to this effect, consisting of 20 date variants. The findings support the usefulness of fused features for better categorization, as well as the efficacy of DCA.

By applying a CNN in MATLAB R2015 to three different types of dates (Aseel, Karbalain, and Kupro), [10] developed a system for date recognition based on color, shape, and size feature datasets of collected photos, achieving a 97.2% accuracy.

[11] collected dataset a containing 3228 photos corresponding to 27 date classes, reporting their experimental results based on five stages. ML techniques were used in the first step, followed by a deep TL model utilizing DenseNet in the second stage, a number tree in the fourth stage, and fine-tuning to produce the model's optimal classification configurations. Regularization was applied in the fifth stage. The best test accuracy was 95.21%, and the validation accuracy reached 97.21%.

For automatic categorization and qualitative comparison, [12] employed three distinct ML techniques: k-nearest neighbors (KNNs), support vector machines (SVMs), and artificial neural networks (ANNs). Concerning the six most popular date fruit types in Oman, the combination of color, shape, and size factors helped to reach the best accuracy. The ANN classifier exhibited a maximum classification accuracy of 99.2% according to trial results.

The you-only-look-once (YOLO) technique was used in [13] to identify and categorize date fruits using DL. Three YOLO models, namely YOLOv5, YOLOv7, and YOLOv8, were trained using 1735 photos of nine different kinds of date fruit. Metrics including the F1-score, recall, and precision were used to assess performance. With a mean recall of 99.0%, a precision of 99.1%, and a mean average precision (mAP) of 99.4%, the results show that YOLOv8 attained great accuracy, implying that it efficiently recognizes and categorizes date fruits according to their surface quality, assisting in productivity optimization by considering uneven ripening over several harvests.

[14] classified seven date fruit types using three ML approaches. From 898 images, 34 features related to morphology, shape, and color were extracted. Logistic regression (LR) and ANN models achieved accuracies of 91.0 and 92.2%, respectively, while a stacked model

combining both increased accuracy to 92.8%. These results demonstrate the effectiveness of ML for date fruit classification.

In order to train a MobileNetV2-based model, [15] created a dataset with eight different types of date fruit. To improve accuracy, a number of pre-processing methods were used, including hybrid weight adjustment, decaying learning rates, picture augmentation, and model checkpointing. With a high classification accuracy of 99.0%, the MobileNetV2 model outperformed others using AlexNet, VGG16, InceptionV3, and ResNet, proving its greater efficacy for classifying date fruits.

In order to improve the model performance, [16] applied several GAN designs for both classification and augmentation, increased the number of training epochs, and broadened the dataset. A CycleGAN-augmented dataset combined with the ResNet152V2 model exhibited a maximum classification accuracy of 96.8%, while a CNN model reached 94.3%. ResNet152V2 and CNNs, on the other hand, achieved 83.0 and 75.0% accuracy, respectively, when using the original dataset.

The aforementioned studies mainly focus on ML and DL-based classification of date fruits utilizing image features. However, novel metaheuristic methods like ant colony optimization (ACO) and the whale optimization algorithm (WOA) have shown considerable promise in enhancing classification efficacy via automated feature selection and model optimization. These techniques, especially when combined with neural networks, have been utilized to optimize parameters like weights and biases, resulting in improved diagnosis accuracy in medical picture classification [17].

Considering the parallels in visual pattern recognition between medical imaging and agricultural classification, future research could investigate the utilization of bio-inspired optimization techniques to enhance EL efficacy, parameter tuning, and feature resilience in date fruit classification.

Materials and methods

Google Colab was employed to perform date fruit classification studies via a soft ensemble technique. Utilizing its GPU and TPU capacities, together with up to 12 GB of RAM, Colab facilitated rapid model training and analysis. The dataset, located on Google Drive, was accessed within the Colab environment. Algorithms were formulated via libraries including Scikit-learn, NumPy, Pandas, Matplotlib, and Seaborn. The studies were conducted using an HP EliteBook G8 laptop equipped with an Intel Core i5-1165G7 CPU (2.6 GHz, four cores), 16 GB of RAM, and 256 GB of storage. A Python script was utilized to load pre-trained models, process the data, train each model, and merge the predictions through a soft ensemble method to assess performance.

Fig. 1 depicts a comprehensive hybrid approach using a voting classifier that incorporates three TL models: MobileNetV2, EfficientNet-B2, and DenseNet201. This method collects features including color, shape, and size from images to identify date fruit varieties. Its efficacy was evaluated against individual TL models alongside a test dataset from Algeria.

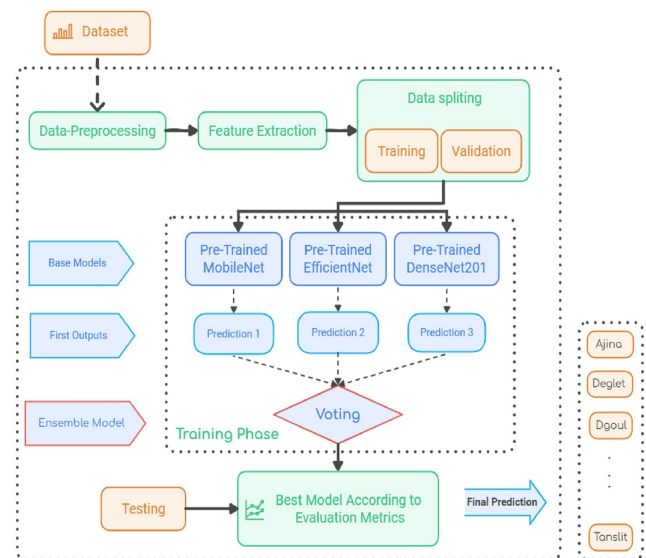


Figure 1. Flowchart of the proposed model

Source: Authors

Dataset description

The model presented herein was trained and assessed using a public dataset from Kaggle [7], which comprises 1619 images from 20 distinct Algerian date fruit varieties, including Adam Deglet Nour, Ajina, Ghars, Bayd Hmam, Bouaarus, Deglet, Deglabayda, Degletkahla, Degletghabia, Dfarlgat, Dgoul, Litima, Loullou, Hamraya, Tarmount, Tanslit, Tantbucht, Techbeh Tati, Tinisin, and Tivisyauin. These samples, sourced and captured from local markets in the Touggourt region in the southeast of Algeria, are presented in Table I, detailing the distribution across varieties. For image-based identification tasks, particularly during the training and evaluation phases, large datasets are essential to improving the accuracy of classification algorithms. This dataset was divided into training and validation sets to enhance the robustness of the model's evaluation. Each variety in the dataset exhibits distinct characteristics, such as shape, size, color, and hardness, further underscoring the need for a comprehensive dataset to accurately assess model performance.

The validation set comprised 20% of the complete date fruit dataset, whereas the training set included the remaining 80%. Each set contained 20 date fruit varieties. Examples of each category are presented in Fig. 2.

Table 1. Distribution of date fruit samples across different varieties in the dataset

Date variety	Number of samples
Adam Deglet Nour	86
Ajina	85
Chars	88
Bayd Hmam	87
Bouaamous	82
Deglet	38
Degla Bayda	95
Deglet Kahla	85
Deglet Ghabia	35
Dfar Lgat	86
Dgoul	103
Litima	85
Loullou	81
Hamraya	76
Tarmount	83
Tanslit	85
Tantbucht	76
Techbeh Tati	88
Tinisin	88
Tivisyauin	87

Source: Authors



Figure 2. Date varieties, arranged from top to bottom and left to right: Adam Deglet Nour, Ajina, Chars, BaydHmam, Bouaamous, Deglet, Deglabayda, Degletkahla, Degletghabia, Dfarlgat, Dgoul, Litima, Loullou, Hamraya, Tarmount, Tanslit, Tantbucht, Techbeh Tati, Tinisin, and Tivisyauin

Source: Authors

Transfer learning (TL)

An efficient ML technique called *TL* uses the knowledge gained from training a model on one task to improve performance on a different yet related one. For applications using little data, like plant disease classification, this makes it easier to reuse pre-trained models such as those developed on large datasets like ImageNet. By utilizing established patterns and information already in the pre-trained model, this technique reduces generalization error and training times. By sharing this knowledge, TL makes it possible to train DL models quickly and accurately, even with small

datasets, which is typically an issue for DL techniques that require large amounts of data.

For prediction, feature extraction, and fine-tuning purposes, this study used sophisticated pre-trained CNN models, including MobileNet, EfficientNet, and DenseNet201. These techniques, which will be further discussed in this document, were combined into ensembles to increase classification accuracy by utilizing the strengths of several models [18]. Each model was trained for 15 epochs with consistent parameters, including a dense layer with 512 units, ReLU activation, an input shape of $224 \times 224 \times 3$, and a dropout rate of 0.5 to avoid overfitting. To preserve the learned features, the initial ten layers of each model were frozen. After the training and validation stage, these TL models, which are detailed in this section, were tested on our local dataset.

MobileNetV2. In 2017, Google introduced MobileNet, a DL-based framework with an emphasis on accuracy, efficiency, and small size for mobile devices. Building upon this framework, MobileNetV2, which was put forth in 2018, improved the architecture even further with a more effective layout tailored to mobile apps. In order to improve feature extraction, MobileNetV2 uses an inverted residual structure, with residual connections between bottleneck layers, in contrast to CNNs. By separating spatial and depth operations, its fundamental building element, the depth-wise separable convolution, lowers computational complexity and increases efficiency without compromising performance. A 32-filter convolutional layer is the first of 19 bottleneck layers in the design. Depth-wise and point-wise convolution layers are then added to capture nonlinearity and select features [19]. MobileNetV2 is especially well suited for resource constrained contexts because it offers greater accuracy with fewer parameters than its predecessor. In order to maximize speed and accuracy in mobile visual recognition tasks, the model's backbone consists of an average pooling layer, a conventional 3×3 convolution, and 17 inverted residual bottlenecks [20].

EfficientNet-B2. The EfficientNet family of DL network designs strike a balance between computing efficiency and accuracy to optimize picture classification. EfficientNet, created by in 2019, offers a compound scaling technique that balances computational cost and performance by consistently scaling the model's depth, width, and resolution. EfficientNet uses fewer parameters while achieving great accuracy in comparison with models such as ResNet and Inception.

The EfficientNet family comprises versions B0 through B7. Each model version retains effective resource use thanks to the compound scaling coefficients, which are found using a grid search. Because of this, EfficientNet is perfect for implementation on devices with limited resources, including smartphones and edge platforms, where it reliably produces high accuracy with low processing requirements.

Interestingly, EfficientNet's initial three iterations use input sizes of 224×224 , ensuring interoperability with the other models in our ensemble approach. The main factor in our decision to use EfficientNet-B2 was its compatibility. EfficientNet has established itself as a leading option for effective DL applications by demonstrating good performance across image classification benchmarks such as ImageNet [21].

DenseNet201. As a member of the DenseNet family, this technique is an advanced DL architecture that uses its densely linked structure to address the vanishing gradient problem. DenseNet-201 has 201 levels and uses dense blocks, wherein each layer gives its outputs to later layers and receives inputs from all layers before it. The overall performance of the model is improved by this distinctive architecture, which encourages effective feature reuse and smooth gradient propagation.

In a variety of computer vision applications, such as semantic segmentation, object detection, and picture classification, DenseNet-201 has shown remarkable performance. This network is quite effective for complicated visual issues because of its deep architecture, which allows recording complex representations. For a variety of applications, academics and practitioners continue to favor DenseNet-201 because of its exceptional performance and versatility [22].

Ensemble learning (EL)

A powerful ML method termed *ensemble modeling* blends several models to improve generalization and prediction accuracy. EL performs better than single models by means of fusion techniques to correct individual model faults. Three TL models, MobileNetV2, EfficientNet-B2, and DenseNet201, were used in this work to categorize Algerian date fruit types. A voting classifier was used to obtain the final predictions [23]. Google Colab was used to train each model on the same dataset, and performance metrics like recall, accuracy, precision, and the F1-score were used to conduct assessments. Models with superior validation performance were given more weight in weighted soft voting, ensuring that they had a greater impact on the ensemble's final predictions.

In order to improve classification accuracy, EL uses a weighted soft voting process to aggregate class probabilities from the base models, leveraging their advantages while mitigating their disadvantages. This method improves performance in sorting date fruits by reducing the impact of overfitting and making use of the varied knowledge representations of pre-trained models. Through the combination of the results of several models, EL demonstrates that it can handle tough classification problems by making the results more accurate and reliable [24].

Evaluation metrics

Accuracy, precision, recall, the F1-score, and confusion matrices were used to assess the effectiveness of ML models for date fruit categorization. Using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), *accuracy* is the percentage of correctly identified samples. *Recall*, also called *sensitivity*, is the percentage of correctly identified real positives (TP) compared to false negatives (FN). *Precision*, on the other hand, looks at how reliable positive predictions are by focusing on TP and FP. Finally, the *F1-score* is a harmonic mean of precision and recall that strikes a balance between the two metrics. The following mathematical formulas were used to measure accuracy, precision, recall, and F1-score:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

These metrics, as defined in Eqs. (1)-(4), provide a thorough examination of the models' classification performance. Furthermore, the classification results were interpreted and visualized in terms of TP, TN, FP, and FN using confusion matrices [21].

Results and discussion

This section shows a detailed comparison of the proposed EL model's performance vs. that of other TL models using a local Algerian dataset. It also analyzes the training and validation curves of all models and evaluates additional metrics, such as precision, the F1-score, and recall.

This study utilized multiple ML models to sort and categorize various types of date fruit, showcasing the potential for automated sorting. To this effect, EL and three TL models were used on a dataset with 1619 photos divided into 20 classes. There is a notable paucity of research focusing on the application of EL to local date information, which is crucial for enhancing accuracy in real-world applications. By merging multiple TL models to enhance date fruit classification, this work fills this research gap. TL approaches employing pre-trained models, including EfficientNet-B2, MobileNet, and DenseNet201, were examined for categorizing date fruit varieties. Among these, the MobileNet model displayed remarkable efficiency during training, reaching a training accuracy of 98.46% and a validation accuracy of 88.54%. However, its performance on the local dataset was unsatisfactory, with a testing accuracy of 89%, the lowest among the models assessed. DenseNet201 achieved remarkable results, with training and validation accuracies of 99.81 and 98.14%, respectively, in addition to a testing accuracy of 98%. Similarly, the EfficientNet design performed well, obtaining training, validation, and testing accuracies of (97.98), (89.78), and (90%), respectively.

In this work, the EL technique outperformed all individual models. Utilizing a weighted soft voting classifier, EL integrated the predictions of the TL models while allocating weights depending on each model's performance and variables collected from the date images, such as size, shape, and color. This technique obtained the maximum performance, with training and validation accuracies of 99.93% and 99.07%, respectively, after 15 epochs. Fig. 3 illustrates the training and validation accuracies of all models, whereas Fig. 4 provides a comprehensive analysis of performance measures, including precision, recall, the F1-score, and accuracy. This comparative phase is necessary to identify the optimal forecasting models based on their evaluation metrics.

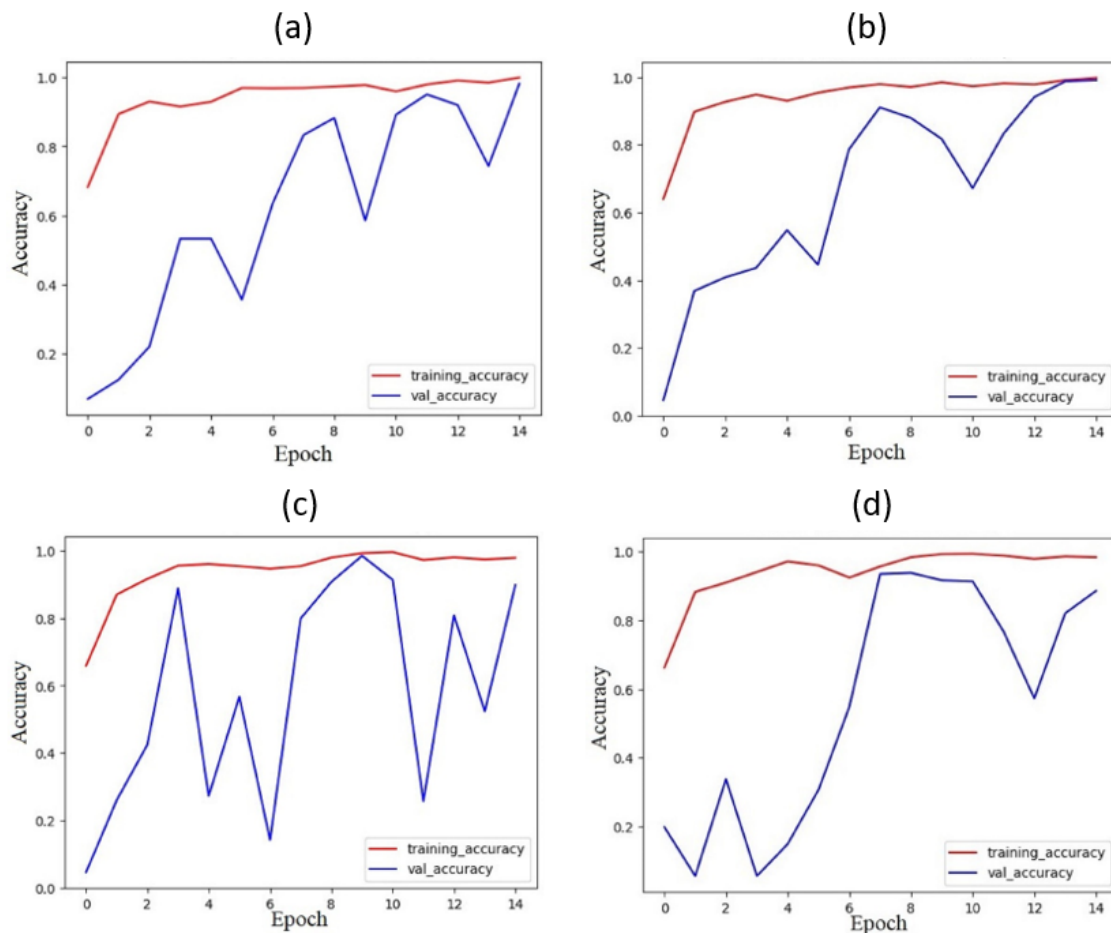


Figure 3. Training and validation accuracy: (a) EL, (b) EfficientNet, (c) DenseNet, (d) MobileNet
Source: Authors

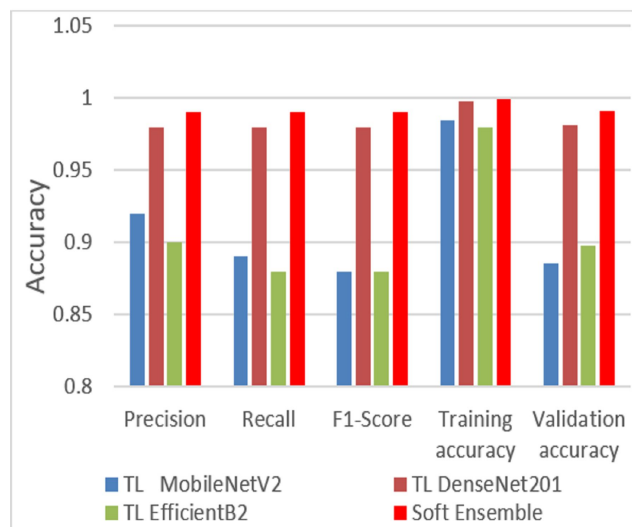


Figure 4. Model performance comparison in terms of evaluation metrics
Source: Authors

The EL model exhibited outstanding performance, attaining the highest testing accuracy (99%) among all experimental models. This outcome underscores its efficacy in precisely selecting and classifying local date fruit varieties. Fig. 5 displays the results of the confusion matrices for the test dataset, which includes 20 different date types, such as Ajina, Ghars, Deglet, Tanslit, and others. These

results further underscore the efficacy of our approach in managing intricate categorization tasks across several date fruit classifications.

Our approach exhibited a remarkable accuracy of 99.93%, surpassing previous research, as illustrated in Table II, while effectively tackling the complexities associated with a local dataset of 20 varieties, unlike some studies that depended on a limited number of species [10, 16]. Furthermore, only a small number of research studies, such as [14], have used EL stacking, but their accuracy ranges from 91.0 to 92.8%. Furthermore, [11], which included TL and covered a larger total of 27 classes, achieved an accuracy of 97.21%, lower than our model. The study most comparable to ours regarding the database is [7], despite minor discrepancies in the conclusions while employing significantly different methodologies. The implementation of weighted soft voting and the foundational pre-trained models in our EL model differentiates our research from others, constituting an innovative approach to harnessing the advantages of several selected TL models for enhanced outcomes. It offers a more resilient solution, improving overall performance while mitigating the danger of overfitting and misclassification. These findings corroborate our notion that EL will surpass individual models, especially when utilized on local datasets containing similar date varieties. The results demonstrate that incorporating several models for feature extraction can markedly enhance classification outcomes.

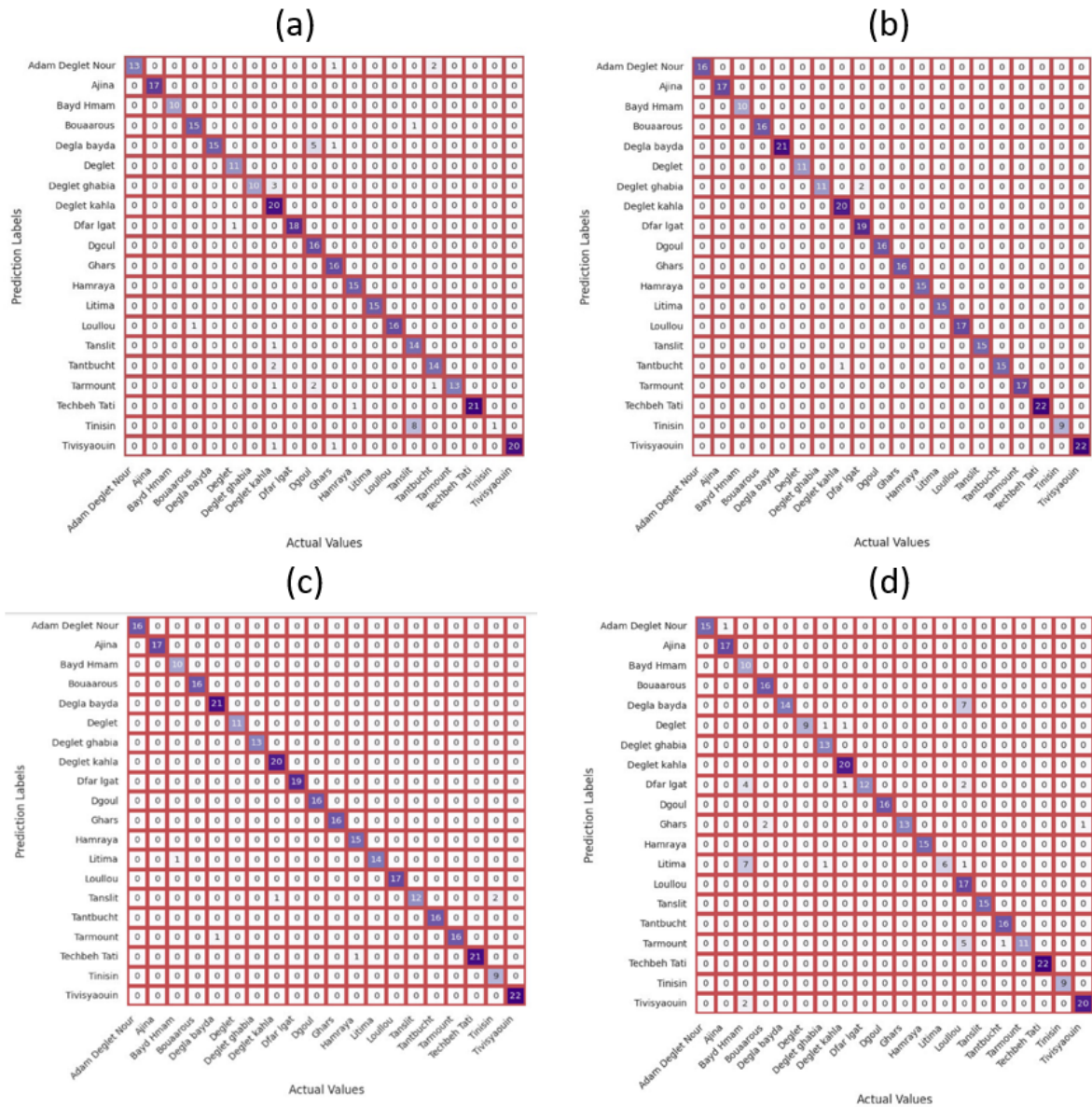


Figure 5. Confusion matrices of the models: (a) EL, (b) EfficientNet, (c) DenseNet, (d) MobileNet
Source: Authors

This work's ramifications go beyond the particular challenge of classifying date fruits. The EL model's effectiveness shows how it can be applied to various agricultural problems, providing opportunities to enhance crop management techniques, including date sorting and quality evaluation.

Future studies should concentrate on creating mobile applications for real-time classification, so that farmers can quickly implement these cutting-edge methods. Furthermore, the date industry's efficiency might be greatly improved by incorporating this model into automated sorting systems, which would increase productivity and expand export opportunities. These developments would allow the public and even novice farmers to recognize date varieties, evaluate their quality, and accurately determine their price or market value. Moreover, investigating alternative learning paradigms like self-supervised learning (SSL) could diminish the reliance on extensive labeled datasets, facilitating more scalable and adaptive classification systems in various agricultural contexts.

Conclusions

In this study, we proposed an EL model capable of classifying 20 popular varieties of date fruits in Algeria. Additionally, three TL techniques were evaluated for the dataset, *i.e.*, TL with EfficientNetB2, MobileNetV2, and Densenet201, in addition to a soft voting EL approach. The results demonstrate the potential of these models to assist farmers in effectively classifying date varieties, contributing to improved agricultural practices, enhanced marketing

Table II. Accuracy comparison of the proposed approach vs. previous literature

Ref	Number of images/classes	Dataset source	Techniques	Best accuracy (%)
[7]	1619 images, 20 varieties	Manually collected in Touggourt, Algeria	Supervised (VGG-F)	98.20
			Unsupervised (PCANet)	99.32
[10]	500 images, 3 varieties	Manually collected, Pakistan	CNN	97.20 95.21
[11]	3228 images, 27 varieties	Farms and shops	ML, TL (DenseNet), Fine-tuned	97.21
[12]	6 varieties	AL-Dhahirah, Oman	ANN, SVM, KNN	99.20
[13]	1735 images, 9 varieties	Kaggle (Saudi dataset)	YOLOv5	98.50
			YOLOv7	99.50
[14]	898 images, 7 varieties	Kaggle (combined dataset)	LR	91.00
			ANN	92.20
			Stacking	92.80
[15]	1750 images, 8 varieties	Saudi Arabia (manual)	MobileNetV2	99.00
[16]	628 images, 3 varieties	Online (augmented)	ResNet152V2	96.80
			CNN	94.30
Our EL Model	20 varieties (Algeria)	From [7]	DenseNet201, EfficientNet-B2, MobileNetV2	99.93
				99.07

Source: Authors

The findings align with and extend previous research, showcasing the value of EL in agricultural applications and its ability to address challenges such as limited local datasets and the similarity between some varieties. Our EL model, applied to a local database, represents a significant advancement in date fruit classification. It demonstrates the potential for integrating computer vision systems into agricultural processes such as automatic date sorting in packaging factories. Furthermore, the proposed approach can contribute to more efficient management in the agricultural sector.

Future research ought to concentrate on enhancing these models and creating practical instruments, such as mobile applications, for the real-time classification of date fruits. Furthermore, investigating alternative methods would enhance the applicability of the method across various agricultural environments. We recommend the inclusion of developing technologies such as generative adversarial networks (GANs) and transformers. Furthermore, SSL signifies a promising avenue, particularly in contexts with a scarcity of labeled data. Utilizing SSL techniques for preliminary tasks like image inpainting, contrastive learning, or clustering-based representation learning allows models to develop robust feature representations before fine-tuning, thereby diminishing reliance on extensive manual annotation and enhancing scalability in agricultural machine vision systems.

CRedit author statement

Sofiane Abden: Conceptualization, methodology, software, data collection, writing (original draft, review, and editing), and formal analysis. **Mostefa Bendjima:** investigation, validation, supervision, project administration, writing (review and editing), and approval of the final version. **Soumia Benkrama:** investigation, validation, supervision, project administration, and writing (review and editing).

References

- [1] S. Abden, M. Bendjima, and S. Benkrama, "Deep learning model based on VGG16 for tomato leaf diseases detection and categorization," in *Proc. 2nd Int. Conf. Electr. Eng. Autom. Control (ICEEAC)*, IEEE, May 2024, pp. 1–6. <https://doi.org/10.1109/ICEEAC61226.2024.10576347>
- [2] J. Rashid, I. Khan, I. Ahmed Abbasi, M. Rizwan Saeed, M. Saddique, and M. Abbas, "A hybrid deep learning approach to classify the plant leaf species," *Comput. Mater. Contin.*, vol. 76, no. 3, pp. 3897–3920, 2023. <https://doi.org/10.32604/cmc.2023.040356>
- [3] A. A. Al-Kharaz, A. B. A. Alwahhab, and V. Sabeeh, "Innovative date fruit classifier based on scatter wavelet and stacking ensemble," *HighTech Innov. J.*, vol. 5, no. 2, pp. 361–381, Jun. 2024. <https://doi.org/10.28991/HIJ-2024-05-02-010>
- [4] Y. Al-Mulla, A. Ali, and K. Parimi, "Detection and analysis of dubas-infested date palm trees using deep learning, remote sensing, and GIS techniques in Wadi Bani Kharus," *Sustainability*, vol. 15, no. 19, art. 14045, Sep. 2023. <https://doi.org/10.3390/su151914045>
- [5] M. Safran, W. Alrajhi, and S. Alfarhood, "DPXception: A lightweight CNN for image-based date palm species classification," *Front. Plant Sci.*, vol. 14, pp. 1–14, Jan. 2024. <https://doi.org/10.3389/fpls.2023.1281724>
- [6] P. Rybacki *et al.*, "Convolutional neural network (CNN) model for the classification of varieties of date palm fruits (*Phoenix dactylifera* L.)," *Sensors*, vol. 24, no. 2, art. 558, Jan. 2024. <https://doi.org/10.3390/s24020558>
- [7] O. Aiadi, B. Khaldi, M. L. Kherfi, M. L. Mekhalfi, and A. Alharbi, "Date fruit sorting based on deep learning and discriminant correlation analysis," *IEEE Access*, vol. 10, pp. 79655–79668, 2022. <https://doi.org/10.1109/ACCESS.2022.3194550>

- [8] S. Khirani, H. Boutaj, C. El Modafar, and A. O. E. Khelil, "Arbuscular mycorrhizal fungi associated with date palm in Ouargla region (southeastern Algeria)," *Plant Cell Biotechnol. Mol. Biol.*, vol. 21, no. 45–46, pp. 15–28, 2020. <https://www.researchgate.net/publication/344712109>
- [9] M. E. H. Chowdhury et al., "Automatic and reliable leaf disease detection using deep learning techniques," *AgriEngineering*, vol. 3, no. 2, pp. 294–312, May 2021. <https://doi.org/10.3390/agriengineering3020020>
- [10] A. Magsi, J. Ahmed Mahar, and S. H. Danwar, "Date fruit recognition using feature extraction techniques and deep convolutional neural network," *Indian J. Sci. Technol.*, vol. 12, no. 32, pp. 1–12, Aug. 2019. <https://doi.org/10.17485/ijst/2019/v12i32/146441>
- [11] A. Alsirhani, M. H. Siddiqi, A. M. Mostafa, M. Ezz, and A. A. Mahmoud, "A novel classification model of date fruit dataset using deep transfer learning," *Electronics*, vol. 12, no. 3, art. 665, Jan. 2023. <https://doi.org/10.3390/electronics12030665>
- [12] L. Khriji, A. Chiheb, and M. Awadalla, "Artificial intelligent techniques for palm date varieties classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, pp. 489–495, 2020. <https://doi.org/10.14569/IJACSA.2020.0110958>
- [13] A. Almutairi, J. Alharbi, S. Alharbi, H. F. Alhasson, S. S. Alharbi, and S. Habib, "Date fruit detection and classification based on its variety using deep learning technology," *IEEE Access*, vol. 12, pp. 190666–190677, 2024. <https://doi.org/10.1109/ACCESS.2024.3433485>
- [14] M. Koklu, R. Kursun, Y. S. Taspinar, and I. Cinar, "Classification of date fruits into genetic varieties using image analysis," *Math. Probl. Eng.*, vol. 2021, pp. 1–13, Nov. 2021. <https://doi.org/10.1155/2021/4793293>
- [15] K. Albarrak, Y. Gulzar, Y. Hamid, A. Mehmood, and A. B. Soomro, "A deep learning-based model for date fruit classification," *Sustainability*, vol. 14, no. 10, art. 6339, May 2022. <https://doi.org/10.3390/su14106339>
- [16] D. M. Ibrahim and N. M. Elshennawy, "Improving date fruit classification using CycleGAN-generated dataset," *Comput. Model. Eng. Sci.*, vol. 131, no. 1, pp. 331–348, 2022. <https://doi.org/10.32604/cmes.2022.016419>
- [17] Y. A. Mukhlif et al., "Ant colony and whale optimization algorithms aided by neural networks for optimum skin lesion diagnosis: A thorough review," *Mathematics*, vol. 12, no. 7, art. 1049, Mar. 2024. <https://doi.org/10.3390/math12071049>
- [18] D. Chen, X. Qi, Y. Zheng, Y. Lu, Y. Huang, and Z. Li, "Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition," *Comput. Electron. Agric.*, vol. 216, art. 108517, Jan. 2024. <https://doi.org/10.1016/j.compag.2023.108517>
- [19] J. Kotwal, R. Kashyap, P. M. Shafi, and V. Kimbahune, "Enhanced leaf disease detection: UNet for segmentation and optimized efficientnet for disease classification," *Softw. Impacts*, vol. 22, p. 100701, Nov. 2024. <https://doi.org/10.1016/j.simpa.2024.100701>
- [20] Z. Li et al., "Enhancing tea leaf disease identification with lightweight MobileNetV2," *Comput. Mater. Contin.*, vol. 80, no. 1, pp. 679–694, 2024. <https://doi.org/10.32604/cmc.2024.051526>
- [21] T. O'Halloran, G. Obaido, B. Otegbade, and I. D. Mienye, "A deep learning approach for maize lethal necrosis and maize streak virus disease detection," *Mach. Learn. with Appl.*, vol. 16, art. 100556, Jun. 2024. <https://doi.org/10.1016/j.mlwa.2024.100556>
- [22] Zia-ur-Rehman et al., "Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique," *PLoS One*, vol. 19, no. 9, art. e0304995, Sep. 2024. <https://doi.org/10.1371/journal.pone.0304995>
- [23] M. Azim Mim, N. Majadi, and P. Mazumder, "A soft voting ensemble learning approach for credit card fraud detection," *Heliyon*, vol. 10, no. 3, art. e25466, Feb. 2024. <https://doi.org/10.1016/j.heliyon.2024.e25466>
- [24] H. Bakır, "VoteDroid: A new ensemble voting classifier for malware detection based on fine-tuned deep learning models," *Multimed. Tools Appl.*, vol. 84, pp. 10923–10944, May 2024. <https://doi.org/10.1007/s11042-024-19390-7>

Evaluation of the Known Sub-Sequence Algorithm (KSSA) for Optimal Use in Time Series of Tuna Fishing in the Pacific Ocean

Evaluación del algoritmo de subsecuencias conocidas (KSSA) para su uso óptimo en series de tiempo de pesca de atún en el Océano Pacífico

Julián A. Gómez ¹, Iván F. Benavides ², and John J. Selvaraj ³

ABSTRACT

An important limitation in fisheries research using time series is the presence of missing data. An inadequate handling of these data can negatively impact the results of statistical analyses, leading to erroneous decision-making. A potential solution to this problem is the estimation of missing values through imputation methods, but none of them can be universally applied to all time series. In fact, their effectiveness largely depends on the structure of the data and the distribution of the missing values. Recently, a solution to this problem was proposed which employs the known sub-sequence algorithm (KSSA), a machine learning technique designed to compare the performance of different imputation methods and validate them within the time series that contains missing data. However, due to its recent development, there is no published evidence regarding its efficiency and reliability. This research aimed to assess the efficiency of the KSSA on tuna fisheries data for the Pacific Ocean by imputing simulated missing data for seven time series with distinct structures. The results demonstrated that the algorithm is robust for accurately validating a wide combination of properties, such as length, seasonality, trend, autocorrelation structure, and percentage of missing data. Additionally, the algorithm's hyperparameters can be easily adjusted to achieve optimal results for each time series.

Keywords: missing data, machine learning, imputation, data

RESUMEN

Una limitación importante en la investigación pesquera que utiliza series temporales es la presencia de datos faltantes. Un manejo inadecuado de estos datos puede afectar negativamente los resultados de los análisis estadísticos, conllevando una toma de decisiones errónea. Una posible solución a este problema es la estimación de valores faltantes mediante métodos de imputación, pero ninguno de ellos puede aplicarse universalmente a todas las series temporales. De hecho, su eficacia depende en gran medida de la estructura de los datos y de la distribución de los valores faltantes. Recientemente, se propuso una solución a este problema que emplea el algoritmo de sub-secuencias conocidas (KSSA), una técnica de aprendizaje automático diseñada para comparar el rendimiento de diferentes métodos de imputación y validarlos dentro de la serie temporal que contiene datos faltantes. Sin embargo, debido a su desarrollo reciente, no hay evidencia publicada sobre su eficiencia y fiabilidad. Esta investigación tuvo por objetivo evaluar la eficiencia del KSSA en datos de pesca de atún para el Océano Pacífico mediante la imputación de datos faltantes simulados para siete series temporales con estructuras distintas. Los resultados demostraron que el algoritmo es robusto para validar con precisión una amplia combinación de propiedades como la longitud, la estacionalidad, la tendencia, la estructura de autocorrelación y el porcentaje de datos faltantes. Además, los hiperparámetros del algoritmo pueden ajustarse fácilmente para lograr resultados óptimos en cada serie temporal.

Palabras clave: datos faltantes, aprendizaje automático, imputación, datos

Received: July 8th, 2024

Accepted: May 14th, 2025

Introduction

The term *time series* refers to a sequence of observations of a variable recorded at regular time intervals (daily, weekly, semiannual, yearly, etc.) [1]. This has diverse applications in various disciplines such as engineering, economics, and biology, among others. Data from time series can be utilized for forecasts that aid in implementing risk prevention policies and control measures or conducting exploratory analyses [2]. However, time series often exhibit missing

¹ Environmental engineer, Universidad Nacional de Colombia, Palmira campus, Colombia. Affiliation: Student, Universidad Nacional de Colombia, Medellín campus, Colombia. Email: jugomezp@unal.edu.co

² Hydrobiological Resources Research Group, Department of Engineering, Universidad Nacional de Colombia, Palmira campus, Carrera 32 No. 12-00, Palmira, Valle del Cauca, Colombia

³ Universidad Nacional de Colombia, Palmira campus, Department of Engineering, Faculty of Engineering and Administration - Carrera 32 No. 12-00 Chapinero, Vía Candelaria, Palmira, 763533 Valle del Cauca - Colombia



Attribution 4.0 International (CC BY 4.0) Share - Adapt

data (MD) due to instrument failures, record discontinuity, adverse weather conditions, and human error, among other factors [3]. Missing values can hinder statistical analysis or influence the generation of erroneous results, leading to inaccurate predictions or forecasts in different models [4]. Consequently, many analysis methods require time series to be free of MD. When there are MD, estimation or imputation is necessary, which involves the replacement of MD with logical values [5]. This issue is solved through simple calculations such as means, averages, and nearby values, or more sophisticated equations based on time series components like trends, seasonality, cyclicity, and autocorrelation structures [6].

It is crucial to identify the appropriate imputation method for each time series, as its efficacy depends on the structure of the existing data and the size and distribution of the MD [4], [7]. In this vein, it is important to note that there is no universally efficient MD imputation technique for all univariate time series [8]. All methods must be validated in order to determine the best approach for each particular time series, optimizing their applicability [9]. This is often achieved using external time series with no missing values, to which simulated MD are added. These MD are then imputed using different methods, and the imputed sequences are compared against the actual data, identifying the best imputation method by means of statistical performance metrics. However, considering that the efficacy of imputation methods depends on the structure of the time series and the MD, this validation approach may yield erroneous results, as the best method for imputing the complete time series used in validation is not necessarily the best option for that containing MD.

[10] offered a solution to this problem by developing a machine learning algorithm called the *known sub-sequence algorithm* (KSSA), which automatically identifies the best imputation method for each specific time series. The KSSA performs validation using the data available within the time series, eliminating the need for external time series. This algorithm leverages the sub-sequences located between the MD, which, when sufficiently informative, allow simulating new MD. These simulated MD can then be imputed using different methods and compared against the real data using performance metrics such as the root mean squared error (RMSE). This tool is used to support the selection of the best statistical imputation method for each time series, regardless of its structure, as the algorithm learns from the aforementioned sub-sequences.

The KSSA operates in six steps. The first step involves identifying the target time series and measuring its percentage of MD relative to the total data in the series. The second step corresponds to the imputation of the MD using all the available or desired imputation methods. This is referred to as *initial imputation* and is aimed at generating a complete time series for subsequent operations. In the third step, the time series is divided into equal segments, and, within each segment, a simulation window is placed whose

size and position are random but always avoid the positions of the initial imputation. Simulated MD are generated based on these positions and sizes, resulting in a new time series (step four), whose percentage of MD relative to the data in the time series is measured. In step five, the simulated MDs are imputed using the methods to be compared and validated, and, finally, performance metrics such as the RMSE are calculated (step six).

Steps three to six are repeated, randomly changing the size and position of the simulation windows at each iteration, which results in different combinations of MD sizes and positions within each segment. The hyperparameters that can be fine-tuned to optimize the results of the KSSA include the selection of initial imputation methods, the imputation methods to be compared, the number of segments, the number of iterations, the percentage of missing data, and the value of the random seed used to ensure reproducibility. This can be done using R's *kssa* package [7].

In this vein, the objective of this research was to evaluate the efficiency of the KSSA using various time series of tuna fishing in the Pacific Ocean, in comparison with a classical validation process. This validation simulated MD in complete time series, estimated them using different imputation methods, and employed performance metrics to confidently identify the best method. The classical validation approach was used as a reference to determine the reliability of the KSSA's results. Further details will be provided in the *Materials and methods* section. Optimal hyperparameter combinations were systematically sought for different time series sizes and structures, as well as for different failure thresholds, to serve as reference for future studies. The beneficiaries of these results will be resource managers, scientists, traders, and fishermen, as time series allow linking fishing information with associated environmental variables, generating risk forecasts, predicting potential fishing areas, reconstructing historical records of fishery resource abundance, and forecasting future behaviors.

Materials and methods

Data

Seven complete fisheries time series (without missing values) were collected from publicly available sources, i.e., the Inter-American Tropical Tuna Commission (IATTC) (<https://www.iattc.org/PublicDomainData/IATTC-Catch-by-species1.htm>) and the Fisheries and Aquaculture section of the Food and Agriculture Organization (FAO), through its FishStatJ application (<https://www.fao.org/fishery/en/statistics/software/fishstatj/en>) (Table I).

The IATTC time series corresponded to the monthly catch per unit effort (CPUE) of skipjack tuna (*Katsuwonus pelamis*) associated with dolphins (SKJ-DEL) and floating objects (SKJ-OBJ), as well as of free-swimming individuals (SKJ-NOA), in the Tropical Pacific Ocean. This time series also included

data on yellowfin tuna (*Thunnus albacares*) associated with dolphins (YFT-DEL) and floating objects (YFT-OBJ). The FAO time series included the annual catch tonnage of Pacific bluefin tuna (*Thunnus orientalis*) (PBF-USA) and southern bluefin tuna (*Thunnus maccoyii*) (SBF-USA) in the Pacific Ocean off the coast of the United States. Each time series was examined for outliers and analyzed to identify trends using Mann-Kendall tests with R's *trend* package [11]. Periodicity was assessed using Lomb periodograms with the *lomb* package [7]. This is shown in Table I and Fig. 1.

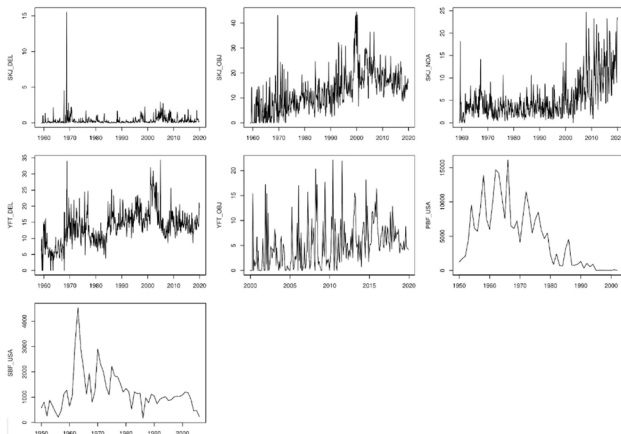


Figure 1. Time series plots used in this study. The vertical axis represents the CPUE.

Source: Authors

Experimental design

An R code was designed to compare the results of classical and KSSA validation. Classical validation involves taking a complete time series (without missing values), simulating the MD, imputing them using different methods, and selecting the best imputation method that simultaneously minimizes the error across various statistical performance metrics. This approach unequivocally identifies the best imputation method for each time series. In this vein, when the results obtained through classical validation and the KSSA converge with regard to the best or top-performing imputation method for each time series, it can be concluded that the KSSA results are reliable. In this work, we compared the performance of 11 imputation methods for MD in univariate time series available in the *kssa* package [7]. This can be seen in Table II.

Table I. Characteristics of the time series studied

Time series	Temporary extension	Length	Trend	Periodicity (dominant frequency)	Temporary resolution	Source
SKJ-DEL	1959-2019	732 months	Positive	100 months	Monthly	IACCT
SKJ-OBJ	1959-2019	732 months	Positive	100 months	Monthly	IACCT
SKJ-NOA	1959-2019	732 months	Positive	100 months	Monthly	IACCT

YFT-DEL	1959-2019	732 months	Positive	100 months	Monthly	IACCT
YFT-OBJ	2000-2019	240 months	Negative	12.5 months	Monthly	IACCT
SBF-USA	1950-2006	56 years	Null	9.09 years	Monthly	IACCT
PBF-USA	1950-2002	52 years	Negative	4.76 years	Monthly	IACCT

Source: Authors

Table II. Imputation methods used for classical and KSSA validation

Method	R function	Source in R
KALMAN smoothing over an ARIMA spatial representation	na_kalman(..., model="auto.arima", ...)	[5]
Kalman smoothing in state space structural models	na_kalman(..., model="StructTS", ...)	[5]
Exponential moving average	na_ma(..., weighting="exponential", ...)	[5]
Linear moving average	na_ma(..., weighting="linear", ...)	[5]
Simple moving average	na_ma(..., weighting="simple", ...)	[5]
Linear interpolation	na_interpolation(..., option="linear", ...)	[5]
Spline interpolation	na_interpolation(..., option="spline", ...)	[5]
Stineman interpolation	na_interpolation(..., option="stine", ...)	[5]
Last forward look	na_locf(..., option="locf", ...)	[5]
Seasonal decomposition with Kalman smoothing and state space models	na_seadec(..., algorithm="kalman", ...)	[5]
Robust decomposition by trend and seasonality with linear interpolation	na.interp	[12]

Source: Authors

Classical validation

In each time series, missing values were randomly simulated at levels of 10, 20, 30, 40, and 50%. To ensure that the positions of the missing values in each time series were different, each percentage used a different random seed by means of the *set.seed* function [13]. Subsequently, the missing values were imputed using the methods mentioned in Table II, and the RMSE was calculated.

KSSA validation

KSSA validation was conducted using different hyperparameter combinations in order to optimize the results for each time series. The imputation methods in Table II were used for both the initial imputation and the final comparison. For the series SKJ-NOA, SKJ-OBJ, SKJ-DEL, and YFT-DEL, 10, 20, and 30 segments were selected; for YFT-OBJ, 1-18 segments were used; for PBF-USA, 1-9 segments were selected; and, for SBF-USA, 1-6 segments were chosen. The percentages of MD were the same as in the classical validation. 10, 50, and 100 iterations were tested for all time series. The same random seeds used in the classical validation were employed to ensure reproducibility.

The results were recorded in a table that included the time series, the percentage of MD, the imputation method, and the performance metrics, with values arranged in ascending order. The best imputation method was the one with the lowest performance metrics. It should be highlighted that this study considered the results of the classical validation as a reference, as the simulation was carried out on a complete time series whose missing values could be accurately recovered. Fig. 2 provides a schematic representation of the classical and KSSA validation processes.

The following libraries or packages were employed by the algorithm: *sjmisc* [14], *missmethods* [15], *imputeTS* [5], *trend* [11], *forecast* [12], *Metrics* [16], *kssa* [7], and *haven* [17].

Data analysis

The KSSA is considered reliable and applicable when its results converge with those of classical validation regarding the best imputation method (i.e., the one that minimizes performance metrics). For instance, if the imputation method with the lowest RMSE (the best method) according to classical validation coincides with that exhibiting the lowest average RMSE after implementing the KSSA, the efficiency of the latter is 100%. This indicates that the algorithm accurately identified the best imputation method for the time series under study. If the best method according to classical validation is ranked second by the KSSA, its efficiency is 90%, and so on. An efficiency of 0% occurs when the best method obtained via classical validation is ranked last by the KSSA. Eq. (1) expresses this efficiency.

$$Ef = \frac{m - n}{m - 1} * 100 \quad (1)$$

where *Ef* represents the efficiency of KSSA imputation; *m* is the total number of imputation methods being compared (11 in this case); *n* is the position provided by the KSSA that coincides with the best method from classical validation; and 100 scales the value to a percentage. This formula may be regarded as a calculation of *Ef* that has been specifically adjusted for this research based on record linkage and data matching [18].

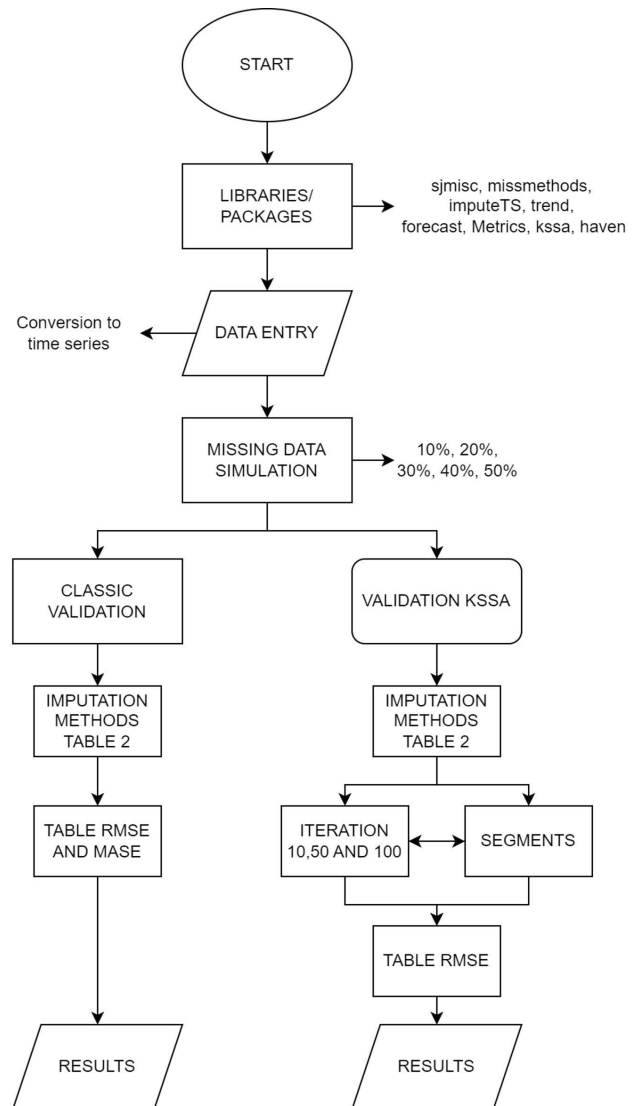


Figure 2. Flowchart summarizing the classical and KSSA validation processes

Source: Authors

The *Ef* values were analyzed using contour plots with the *filled.contour* function from R's *graphics* package [13]. Contour plots were generated for each combination of time series, number of segments, percentage of missing values, and number of iterations. The optimal efficiency area (OEA) was defined as the region within the 70% *Ef* contour, as there were no statistically significant differences regarding RMSE values between the first, second, third, and fourth positions of the KSSA (corresponding to 100, 90, 80, and 70% *Ef*, respectively). This was determined using a simple analysis of variance (ANOVA) with a significance level of 0.05. The OEA can be seen as a continuous white line in the contour plots in Fig. 3.

Finally, a multifactorial ANOVA was performed to quantify the variance explained by the number of iterations, the number of segments, and the percentage of missing values as fixed factors on *Ef*. This model considered all possible double interactions between the three factors. We selected a significance level of 0.05. These results are presented in Fig. 4.

Results and discussion

According to the results of this study, the time series exhibited a variety of structures, as detailed in Table II. This diversity included different lengths and patterns of temporal behavior, such as periodicity and trend. Specifically, periodicity and trend were observed in the SKJ-DEL, SKJ-OBJ, SKJ-NOA, YFT-DEL, and PBF-USA series, while YFT-OBJ and SBF-USA showed only periodicity. As mentioned by [19], time series can exhibit a wide variety of patterns. These characteristics can significantly affect the results, especially when there is no clear seasonality.

Starting at 40%, specific particularities were observed which differentiated each series, mainly in relation to the percentage of missing values, which affected the size, shape, and distribution pattern of the OEA. This finding is consistent with previous studies using imputation methods, wherein performance decreases when the percentage of MD is high. For example, [20] observed that, while the method generally maintains consistent performance, its accuracy deteriorates significantly when handling datasets with 50% missing values. Similarly, [21] found that the performance of imputation methods decreases when the percentage of MD varies between 5 and 30%.

The factor explaining most of the variance regarding efficiency (Ef) was the percentage of missing values – both additively (for series YFT-DEL, YFT-OBJ, SBF-USA, and PBF-USA) and multiplicatively in interacting with the number of segments (for series SKJ-DEL, SKJ-NOA, and SKJ-OBJ). Additionally, it was found that the maximum operating threshold for this factor was 50%. Above this value, the KSSA failed in all time series. Below this threshold, the efficiency increased as the percentage of MD approached 0%. The observed failure is consistent with other studies, which generally recommend working with time series exhibiting a maximum of 50% MD; it is assumed that, above this percentage, efficiency decreases dramatically. For example, [22] described a procedure to introduce random missing values into the first two datasets, varying the proportion from 2.5 to 50% in 2.5% increments. When this limit is exceeded, the precision of the algorithm is compromised.

Furthermore, increasing the number of iterations homogenized and stabilized the OEA in all time series, providing a better resolution and a more reliable decision regarding the best imputation method. This improved resolution was more noticeable in YFT-OBJ, PBF-USA, and SBF-USA. While the general structure of the OEA was similar in terms of size, position, and shape for any number of iterations, these three properties gained a better resolution with 50 and 100 iterations. This indicates that the decision on the best imputation method was more reliable with a higher number of iterations, especially when increasing from 10 to 50 – between the latter and 100, the differences were minimal or imperceptible.

Except for SKJ-DEL and SKJ-NOA, the OEA covered the majority of the contour plots, indicating the robustness of the KSSA when faced with a wide combination of factors and time series. In SKJ-DEL and SKJ-NOA, the OEA was more restricted, occupying approximately 25% of the contour plot. The issue with SKJ-DEL, whose OEA was located between 35 and 50% of MD, could be due to an atypical CPUE recorded for the 1960s, which exceeded the series average by more than 1000%. For SKJ-NOA, the OEA was located between 10 and 20% of MD. However, this series did not contain any noteworthy atypical data in comparison with SKJ-DEL. These cases highlight the importance of addressing outliers, as they may significantly impact the algorithm's efficiency.

It was observed that the OEA of shorter time series had less regularity. For example, the PBF-USA series, the shortest in length (52), showed more irregular tracings in its 70% Ef contour, even with 100 iterations. Similarly, SBF-USA and YFT-OBJ, with lengths of 56 and 240, respectively, showed patches of non-optimal Ef ($<70\%$) within the OEA for 10 and 50 iterations. Conversely, for longer time series (732), the OEA had a better resolution, even with 10 iterations in some cases (e.g., SKJ-OBJ).

It should be noted that optimal efficiency was achieved with time series exhibiting up to 50% of MD, albeit only with low or minimal segment numbers. For example, in SKJ-DEL and SKJ-OBJ, an optimal Ef at between 40 and 50% MD was reached only when the number of segments was 1 or 2. This may be related to the considerably high level of variance explained by the interaction between the number of segments and the percentage of MD in these series.

The data shown in Table I provide an overview of the various structures of the time series studied. Patterns such as periodicity and trend were identified in some of them, while others showed only periodicity. It is always important to corroborate the presence of these elements in the structure of the series, in order to better understand the functioning of the KSSA; as demonstrated in this study, its efficiency varies depending on the presence or absence of such structures. This is especially useful when seasonality, trend, cycles, and autocorrelations have been well identified.

An important observation is that the performance of the KSSA increases with longer time series but with a restricted number of segments. The algorithm exhibits total failure after surpassing the maximum threshold of this hyperparameter, which was fully identified in Table III but should be determined for each time series. The reason for this failure is that there is insufficient information in each segment for the algorithm to learn, which precludes its initialization.

These maximum thresholds varied mainly between the IATTC and FAO time series, which had monthly and annual resolutions, respectively, as well as different lengths. This suggests that both the length and the temporal resolution of the series influence the efficiency of the KSSA. This makes sense because monthly series have a greater capacity to detect

seasonal cyclicals, associated, for example, with climatic activity, which is highly relevant to fisheries. Fig. 4 shows a common pattern in all time series regarding the response of the OEA to the number of segments, percentage of MD, and iterations. This pattern is a decreasing plateau from left to right, with an inflection point starting at 40%, which is modified only by certain local particularities. The percentage of MD explained most of the of variation in the size, shape, and distribution of the OEA, highlighting its importance for the proper management of the algorithm. Thereupon, we suggest quantifying the percentage of MD before running the KSSA in order to assess the reliability of its results.

The maximum threshold of operation for this factor was 50% in all series, trials, and repetitions, which means that, when the number of MD is greater than that of existing data, the algorithm is unable to find enough information to learn and function. In this vein, we recommend that the results be cautiously interpreted and applied when the percentage of MD in the series of interest is between 40 and 50%. For this range, the decision regarding the best imputation method may have limited reliability. In any case, factors such as the structure of the time series, information on the origin and the recording process of the data, the process that generates MD (MCAR, MAR, or NMAR), and expert knowledge of the studied system should be jointly considered.

The number of iterations did not affect the size, shape, or distribution of the OEA, but it did improve its resolution in all time series. Increasing the number of iterations homogenized and stabilized the results, providing a better resolution and a more reliable decision regarding the best imputation method. We suggest using 50-100 iterations, although this will naturally depend on the processing capacity of each computer and the time available to obtain ready and reliable results. As a reference, on a standard personal-use computer with 8 GB DDR4, 2400 MHz RAM and a 2.0 GHz four-core processor, the KSSA can process a time series of 732 data points with 30% MD in five minutes using five segments.

The KSSA proved to be robust in most time series, as the OEA occupied most of the area in the contour plots. However, there were some exceptions, such as SKJ-DEL and SKJ-NOA, whose OEA was more restricted. These cases could be due to outliers in the data that affected the efficiency of the algorithm. Therefore, we suggest caution with the application of the KSSA in time series with one or a few outliers, as its results regarding the best imputation method may not be reliable. We also noted a direct relationship between the length of the time series and the regularity of the OEA: shorter series had less regular OEA, while longer ones showed a more defined OEA, even with few iterations.

This study demonstrated that the KSSA is an effective and reliable algorithm to optimally and automatically detect the best imputation method for any given time series, but caution should be exercised when the percentage of MD is between 40 and 50%. Before its publication as an R package in 2022 and as a scientific article in 2023, there was no such tool.

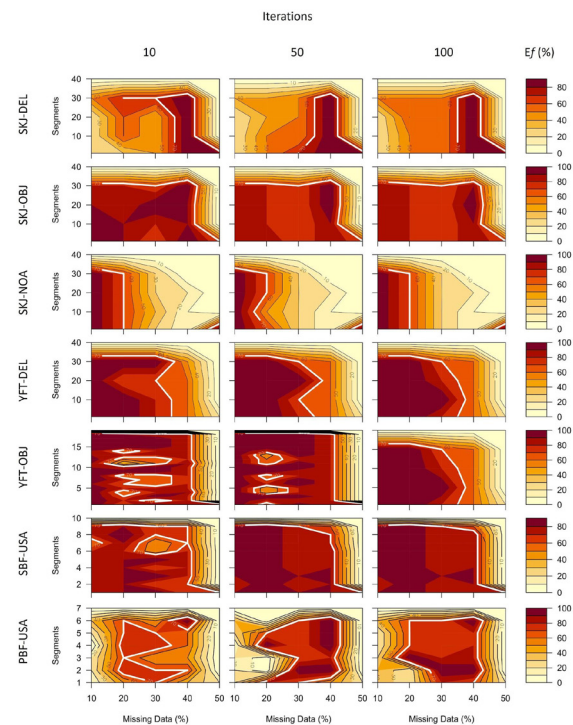


Figure 3. Contour plots representing the efficiency (E_f) of the KSSA in response to the time series, the number of segments, the percentage of missing data, and the number of iterations, using RMSE as the performance metric

Source: Authors

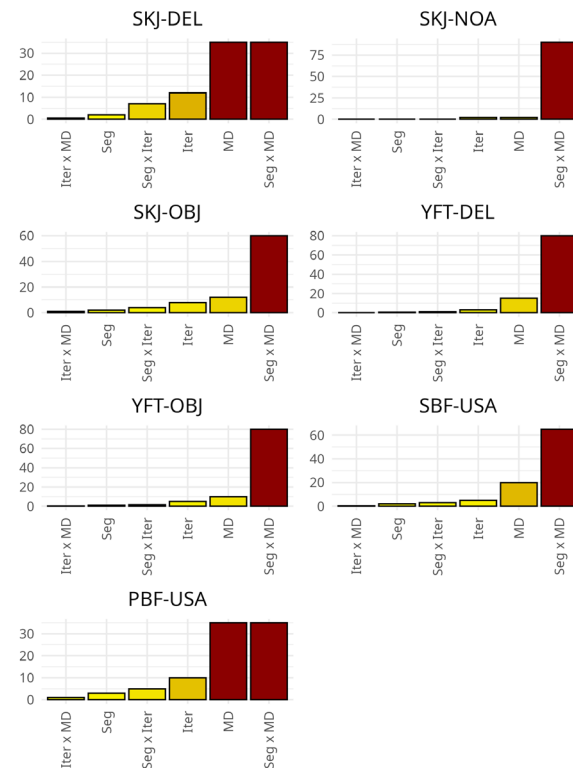


Figure 4. Portion of the variance explained by each source of variation and their double interactions regarding efficiency (E_f) for each time series. The values on the vertical axis represent the percentage of variance explained, according to a multifactorial ANOVA.

Source: Authors

Conclusions

The efficiency of the KSSA varies depending on the structure of the time series under study, including elements such as periodicity, trend, cycles, and autocorrelation. It performs better with longer time series but requires a restricted number of segments. The performance of the algorithm is significantly affected when the number of segments exceeds a certain threshold, as identified in this study. This failure occurs due to insufficient information in each segment for the algorithm to learn, which hinders its initialization.

Both the length and the temporal resolution of the time series influence the efficiency of the KSSA. Monthly series, for example, are more effective in detecting seasonal cycles that are relevant to fisheries. The results indicated that both length and temporal resolution are crucial for the algorithm's performance, as demonstrated by the differences between the IACCT and FAO time series, with monthly and annual resolutions, respectively.

The performance of the KSSA decreases significantly when the percentage of MD exceeds 50%. For MD percentages between 40 and 50%, the imputation results may be less reliable, underscoring the need to quantify this aspect before running the algorithm. Additionally, increasing the number of iterations improves the resolution and stability of the KSSA's results. We suggest using between 50 and 100 iterations depending on the processing capacity and time available.

The KSSA proved to be robust for most time series, with the optimal efficiency area (OEA) occupying most of the contour plots. However, caution is advised when applying the KSSA to time series with outliers, as these can affect the algorithm's efficiency. We also noted a direct relationship between the length of the time series and the regularity of the OEA, with shorter series showing a less regular OEA. This study demonstrated that the KSSA is an effective and reliable algorithm for optimally detecting the best imputation method for each time series of interest. However, caution should be exercised when the percentage of MD is between 40 and 50%. Prior to its publication as an R package in 2022 and as a scientific article in 2023, there was no tool that efficiently performed this task.

Given its early success, with 2340 downloads and three citations within a year of publication, it is likely that this algorithm will be widely used in time series research across various fields. Selecting the appropriate values for its parameters is crucial for obtaining optimal results, as each time series has unique characteristics. Future studies should include comprehensive analyses aimed at identifying additional sources of variation that may affect the algorithm's performance. For now, the use of KSSA is recommended for imputing MD in univariate tuna fishing time series.

Acknowledgements

The authors acknowledge the funding provided by Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA), as well as the logistical support from the Pacific Studies Institute of Universidad Nacional de Colombia, Tumaco campus, and from the authors of R's kssa package.

CRedit author statement

- Author 1 conceived the idea and was responsible for the statistical analysis, visualization, and original draft of the article.
- Author 2 supervised the work, was in charge of the methodology and data curation, and was involved in writing the manuscript.
- Author 3 supervised the work and was involved in writing and revising the manuscript.

Data availability

Link: <https://github.com/gjuliannp/KSSA.git>

References

- [1] F. Parra, "Estadística y machine learning con R," 2019. [Online]. Available: <https://bookdown.org/content/2274/series-temporales.html>
- [2] N. Bokde, M. W. Beck, F. M. Álvarez, and K. Kulat, "A novel imputation methodology for time series based on pattern sequence forecasting," *Pattern Recognit. Lett.*, vol. 116, pp. 88–96, 2018. <https://doi.org/10.1016/j.patrec.2018.09.020>
- [3] E. A. Yamoah, U. A. Mueller, S. M. Taylor, and A. J. Fisher, "Missing data imputation of high-resolution temporal climate time series data," *Meteorol. Appl.*, vol. 27, no. 1, Jan. 2020. <https://doi.org/10.1002/met.1873>
- [4] M. W. Beck, N. Bokde, G. Asencio-Cortés, and K. Kulat, "R package imputetestbench to compare imputation methods for univariate time series," *The R Journal*, vol. 10, no. 1, pp. 218–233, 2018. <https://doi.org/10.32614/rj-2018-024>
- [5] S. Moritz, A. Sardá, T. Bartz-Beielstein, M. Zaefferer, and J. Stork, "Comparison of different Methods for Univariate Time Series Imputation in R," 2015. [Online]. Available: <https://doi.org/10.48550/arXiv.1510.03924>
- [6] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Appl. Energy*, vol. 225, pp. 998–1012, Sep. 2018. <https://doi.org/10.1016/j.apenergy.2018.05.054>
- [7] I. F. Benavides, M. Santacruz, J. P. Romero-Leiton, C. Barreto, and J. J. Selvaraj, "kssa: Known sub-sequence algorithm," *Aquac. Fish.*, vol. 8, no. 5, pp. 587–599, Jun. 2022. <https://doi.org/10.1016/j.AAF.2021.12.013>
- [8] J. Honaker et al., "What to do about missing values in time-series cross-section data," *Am. J. Polit. Sci.*, vol. 54, no. 2, pp. 561–581, 2010.

- [9] N. Golyandina and A. Korobeynikov, "Basic singular spectrum analysis and forecasting with R," *Comput. Stat. Data Anal.*, vol. 71, pp. 934–954, Mar. 2014. <https://doi.org/10.1016/j.csda.2013.04.009>
- [10] I. F. Benavides, M. Santacruz, J. P. Romero-Leiton, C. Barreto, and J. J. Selvaraj, "Assessing methods for multiple imputation of systematic missing data in marine fisheries time series with a new validation algorithm," *Aquac. Fish.*, vol. 8, no. 5, pp. 587–599, Sep. 2023. <https://doi.org/10.1016/j.AAF.2021.12.013>
- [11] T. Pohlert, "Non-parametric trend tests and change-point detection [R package trend version 1.1.5]," 2023, [Online]. Available: <https://CRAN.R-project.org/package=trend>
- [12] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *J. Stat. Softw.*, vol. 27, no. 3, pp. 1–22, July 2008. <https://doi.org/10.18637/jss.v027.i03>
- [13] R. C. Team, "R: A language and environment for statistical computing," 2022. [Online]. Available: <https://www.r-project.org/>
- [14] D. Lüdtke, "sjmisc: Data and variable transformation functions," *J. Open Source Softw.*, vol. 3, no. 26, p. 754, Jun. 2018. <https://doi.org/10.21105/JOSS.00754>
- [15] R. Tobias, "missMethods: Methods for Missing Data. R package version 0.3.0," 2022. [Online]. Available: <https://cran.r-project.org/web/packages/missMethods/index.html>
- [16] H. Ben and F. Michael, "Metrics: Evaluation metrics for machine learning. R package metrics version 0.1.4," 2018, [Online]. Available: <https://CRAN.R-project.org/package=Metrics>
- [17] W. Hadley, M. Evan, and S. Danny, "haven: Import and export 'SPSS', 'Stata' and 'SAS' files," 2022, [Online]. Available: <https://CRAN.R-project.org/package=haven>
- [18] P. Christen, "Data linkage: The big picture," *Harv. Data Sci. Rev.*, vol. 1, no. 2, p. 2019, Nov. 2019. <https://doi.org/10.1162/99608F92.84DEB5C4>
- [19] S. Moritz, A. Sardá, T. Bartz-Beielstein, M. Zaefferer, and J. Stork, "Comparison of different methods for univariate time series imputation in R," 2015. [Online]. Available: <https://doi.org/10.48550/arXiv.1510.03924>
- [20] C. Yozgatligil, S. Aslan, C. Iyigun, and I. Batmaz, "Comparison of missing value imputation methods in time series: The case of Turkish meteorological data," *Theor. Appl. Climatol.*, vol. 112, Apr. 2012. <https://doi.org/10.1007/s00704-012-0723-x>
- [21] N. Savarimuthu and S. Karesiddaiah, "An unsupervised neural network approach for imputation of missing values in univariate time series data," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 9, art. e6156, 2021. <https://doi.org/10.1002/cpe.6156>
- [22] R. Wei et al., "Missing value imputation approach for mass spectrometry-based metabolomics data," *Sci. Rep.*, vol. 8, no. 1, art. 663, Jan. 2018. <https://doi.org/10.1038/s41598-017-19120-0>

Assessing the Interaction between Water Erosion and SOC Storage in a Small Mexican Watershed

Evaluación de la interacción entre la erosión hídrica y el almacenamiento de COS en una microcuenca mexicana

Olimpya T. Aguirre-Salado ¹, Joel Pérez-Nieto ², Carlos A. Aguirre-Salado ³, and Alejandro I. Monterroso-Rivas ⁴

ABSTRACT

Water erosion is a significant issue that impacts a substantial portion of Mexico. The purpose of this study is to establish a connection between soil erosion and soil organic carbon (SOC) reserves. This work was conducted within a small watershed in the Mixteca Alta region of Oaxaca, in order to examine the correlation between erosion intensity, determined via the revised universal soil loss equation (RUSLE), and SOC storage, calculated using spatial models. The results reveal erosion values between 0.19 and 266.99 Mg ha⁻¹ year⁻¹, with 305 693 t of erosion in the micro-watershed. The erosion patterns are closely linked to land use categories. The average SOC values (in Mg ha⁻¹) were associated with erosion, which was reclassified as null (31.79), light (22.36), moderate (16.19), and high (5.22). Kendall's *tau* coefficient showed a negative correlation of -0.39 between erosion and SOC. This inverse relationship can be attributed to the influence of erosive processes on the transport and exposure of SOC, the later replacement of carbon in the vegetation, and a reduced decomposition in deposition areas. Therefore, conservation practices, particularly terracing, have the potential to improve carbon storage.

Keywords: RUSLE, Kendall's tau coefficient, conservation practices, terracing, soil carbon stock

RESUMEN

La erosión hídrica es un problema significativo que afecta gran parte de México. El propósito de este estudio es establecer una conexión entre la erosión del suelo y las reservas de carbono orgánico del suelo (COS). Este trabajo fue realizado dentro de una pequeña cuenca hidrográfica en la región Mixteca Alta de Oaxaca para evaluar la correlación entre la intensidad de la erosión, determinada a través de la ecuación universal de pérdida de suelo revisada (RUSLE), y el almacenamiento del COS, calculado mediante modelado espacial. Los resultados revelan valores de erosión entre 0.19 y 266.99 Mg ha⁻¹ año⁻¹, con una erosión total de 305 693 t en la microcuenca. Los patrones de erosión están estrechamente vinculados a las categorías de uso del suelo. Los valores promedio de COS (en Mg ha⁻¹) se asociaron a la erosión, reclasificada como nula (31.79), ligera (22.36), moderada (16.19) y alta (5.22). El coeficiente *tau* de Kendall mostró una correlación negativa de -0.39 entre la erosión y el COS. Esta relación inversa puede atribuirse a la influencia de los procesos de erosión en el transporte y la exposición del COS, el posterior reemplazo de carbono de la vegetación y la reducción de la descomposición en las áreas de depósito. Por lo tanto, las prácticas de conservación, en particular las terrazas, tienen potencial para mejorar el almacenamiento de carbono.

Palabras clave: RUSLE, coeficiente tau de Kendall, prácticas de conservación, terrazas, almacenamiento de carbono en suelo

Received: February 18th, 2025

Accepted: May 19th, 2025

Introduction

Soil erosion entails soil displacement due to factors such as rain, runoff, wind, and gravity, leading to its deposition in depressions or water bodies [1]. This escalating issue threatens soil quality and environmental functions. Erosion contributes to organic matter (OM) loss through quicker decomposition and export from eroded sites [2]. Water erosion follows a four-stage process: initiation, splashing, transportation, and redistribution – as well as eventual deposition in depressions or water bodies [3]. This form of erosion is extremely damaging and leads to significant

¹ MSc in Hydrosiences, Colegio de Postgraduados, Mexico. Affiliation: Doctoral student, Universidad Autónoma Chapingo, Mexico. Email: oaguirres@chapingo.mx

² PhD in Soil Science, Colegio de Postgraduados, México. Affiliation: Professor (full), Universidad Autónoma Chapingo, Mexico. ORCID: Email: jperez@chapingo.mx

³ PhD in Natural Resource Management, Universidad Autónoma de Nuevo León, México. Affiliation: Professor (full), Universidad Autónoma de San Luis Potosí, Mexico. ORCID: Email: carlos.aguirre@uaslp.mx

⁴ PhD in Geography, Universidad Nacional Autónoma de México, México. Affiliation: Professor (full), Universidad Autónoma Chapingo, México. Email: aimrivas@correo.chapingo.mx



Attribution 4.0 International (CC BY 4.0) Share - Adapt

land degradation and nutrients, soil organisms, and organic carbon transport, thereby altering ecosystems.

Soil organic carbon (SOC) is the carbon that remains in the soil after the partial decomposition of organic matter [4]. Global SOC estimates range from 1500 to 2400 Pg C, making the soil the second-largest carbon reservoir [11]. SOC's relevance lies in its role in climate change mitigation through photosynthesis, which captures atmospheric carbon in the soil for long-term storage [5]. However, whether soil acts as a carbon sink or source depends on the SOC formation and decomposition rates.

[6] highlighted that approximately 50% of the SOC is concentrated within the upper 20-30 cm of soil depth. This implies that the first meter of soil holds two to three times more carbon than the atmosphere and Earth's entire vegetation combined [5]. Considering that water erosion displaces the topsoil, which is where most of the SOC is typically found, it is evident that this phenomenon has the potential to significantly impact SOC stocks. [7] explained that erosion processes modify SOC inventories by transporting SOC-enriched sediment away from a given landscape unit. Consequently, these processes not only result in the oxidation of SOC stocks and the release of carbon dioxide (CO₂) into the atmosphere; they also lead to SOC loss via surface runoff. Hence, erosion essentially redistributes SOC throughout the landscape.

[8] identified that erosion increases the risk of SOC loss. Although some studies have highlighted the impact of erosion on SOC loss, there is a noticeable scarcity of extensive research or case studies that directly correlate soil loss and SOC reserves at the landscape or watershed level in Mexico. Focused case studies could significantly enhance our understanding of the intricate relationship between water erosion and SOC storage dynamics.

The use of geographic information systems (GIS) is essential for predicting soil erosion in landscape and watershed studies. A variety of methodologies and software have been used in these investigations, but the revised universal soil loss equation (RUSLE) model remains the predominant choice for landscape-level erosion modeling [9].

In Mexico, water erosion has affected 52.86% of the nation's land area [10]. This raises a fundamental question: *is there a connection between water erosion and reduced SOC per unit area?* In this vein, the purpose of this study was to investigate the relationship between these factors. This was done by means of the RUSLE predictive model for estimating erosion values. Next, we examined the spatial relationship between erosion and SOC storage in a small watershed located in the state of Oaxaca, Mexico. This information could be used in assisting decision-makers with regard to resource management, in order to preserve the carbon content of soils. Our hypothesis is that water erosion significantly reduces the amount of SOC stored within the watershed per unit area.

Materials and methods

Study area

The study area, called *El Arenal*, is a small 44.6 ha watershed located within the Mixteca Alta region of San Miguel Tulancingo, Oaxaca (Fig. 1), at an altitude of 2200 m. The study area was selected due to its distinctive landscape and apparent erosion indicators, such as gullies and gutters, combined with soil and water conservation practices. According to the INEGI soil map (1:250 000 scale) [11], most of the micro-watershed is classified as a Vertisol. However, field data reveal greater edaphic heterogeneity, with the presence of sandy, sandy loam, and clay loam textures recorded at various locations. These textures suggest the coexistence of other soil types, such as Litosols and Castañozems.

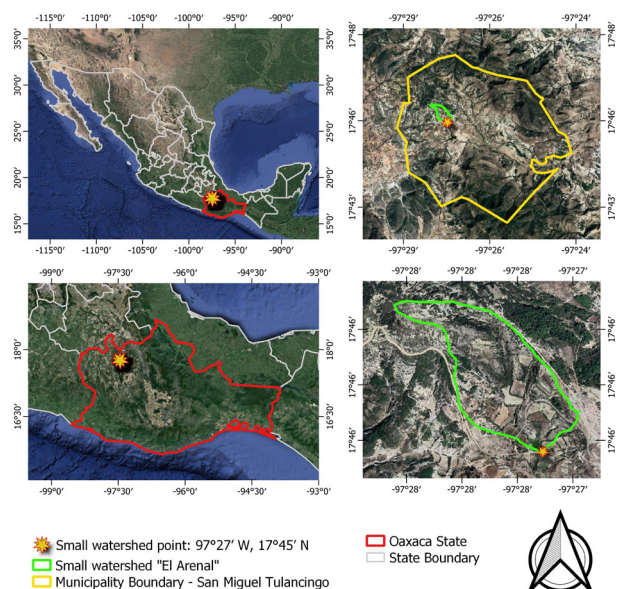


Figure 1. Location of the study area
Source: Authors

Data

Soil organic carbon data

The SOC map used in this study was obtained from a previously validated spatial model developed by [12], using the Smart Map plugin in QGIS, which applies machine learning-based interpolation techniques. In the aforementioned study, a total of 77 geopositioned soil samples were collected from El Arenal at a depth of 0-30 cm. Sampling locations were selected based on accessibility, land use diversity, and topographic variation in order to ensure a representation of the environmental heterogeneity within the watershed. The model reported a mean prediction error (ME) of 0.98 Mg SOC ha⁻¹, a root mean square prediction error (RMSE) of 3.77 Mg SOC ha⁻¹, and a coefficient of determination (R²) of 0.96, indicating a strong predictive accuracy. This SOC map was incorporated as an input dataset in this study, with the aim of analyzing the spatial relationship between erosion and SOC.

Climate data

Rainfall data for the past 30 years from 11 climate stations within the proximity of the watershed were accessed through the National Meteorological Service's CLICOM database [13].

Soil data

Across the study area, 11 geopositioned soil samples were collected to represent different land use and vegetation conditions. These samples were analyzed to determine particle size distribution – sand, silt, and clay (%) – using the Bouyoucos hydrometer method. In addition, SOC (g kg^{-1}) was measured via the Walkley and Black procedure [14].

Digital elevation model (DEM)

Slope calculations (%) utilized QGIS's slope tool with a 5 m spatial resolution DEM [15].

Satellite imagery

For the analysis, we used Sentinel-2 satellite imagery with a spatial resolution of 10 m, which was acquired on September 26, 2018, from the USGS EarthExplorer platform (<https://earthexplorer.usgs.gov/>). Atmospheric correction was performed using Luca Congedo's open-source semi-automatic classification plugin (SCP) within QGIS. Subsequently, bands 4 (red) and 8 (near-infrared) were clipped to the extent of the study area.

Methodology

This study was conducted in three phases. In the first phase, a SOC map was generated [12] (Fig. 2). The second phase focused on employing the RUSLE to model erosion estimation, and the third phase established correlations between erosion (X) and SOC (Y).

Spatial modeling of water erosion

Erosion was estimated using the RUSLE. This erosion model is tailored to predict the average annual soil loss due to slope runoff over the long term. The model is illustrated in Eq. (1) [16]–[17]:

$$A = R * K * L * S * C * P \quad (1)$$

where A represents the average soil loss rate ($\text{Mg ha}^{-1} \text{ year}^{-1}$); R denotes the precipitation erosivity factor ($\text{MJ.mm.h}^{-1}.\text{ha}^{-1}.\text{yr}^{-1}$); K represents the soil erodibility factor ($\text{t.h.MJ}^{-1}.\text{mm}^{-1}$); LS combines the factors of slope length and steepness (dimensionless); C denotes the factor for land use and cover (dimensionless); and P is the factor of management practices aimed at preserving the soil (dimensionless).

Rainfall erosivity factor

Rainfall erosivity (R) measures the kinetic energy of raindrops, which cause soil particle displacement [18]. The data show that soil losses from cultivated fields are proportional to storm energy (E) divided by a 30 min maximum intensity (I30) when other factors are constant [17]. The R factor was determined using Eq. (2) [19], as used by [9].

$$R = 0.0483 * P^{1.610} \quad (2)$$

where P is the annual precipitation (mm). To find the R factor, the annual precipitation was interpolated in QGIS using the multilevel B-spline algorithm [20], creating a representative rainfall map. Subsequently, the R factor map was computed using the raster calculator.

Soil erodibility factor

The soil erodibility (K) factor signifies the soil's susceptibility to water-induced erosion.

The K factor was computed using Eq. (3) [16]–[19], as used by [9].

$$K = 27.66 * m^{1.14} * 10^{-8} * (12-a) + 0.0043 * (b-2) + 0.0033 * (c-3) \quad (3)$$

where K is the soil erodibility factor ($\text{t ha h ha}^{-1} \text{ MJ}^{-1} \text{ mm}^{-1}$); $m = (\text{Silt \%} + \text{Sand \%}) \times (100 - \text{clay \%})$; a refers to the percentage of organic matter; and b is a structure code: 1) very structured, 2) fairly structured, 3) slightly structured, and 4) solid. Additionally, c is the profile permeability code: 1) rapid, 2) moderate to rapid, 3) moderate, 4) moderate to slow, 5) slow, and 6) very slow.

To collect soil data and calculate the K factor, 11 strategically positioned points in the study area were initially selected for soil sampling. The samples were analyzed following Walkley and Black's method [14] for assessing OM, in addition to Bouyoucos method for determining the texture (% sand, silt, and clay).

The K factor was computed and interpolated in QGIS using the IDW method [21]. It exhibited a focal trend, accentuating intense local maxima that coincided with the sampling points.

Slope length and steepness factor (LS)

Steepness and slope length are topographic elements that critically influence soil erosion dynamics. The LS factor is a composite parameter that encompasses the combined effects of slope length (L) and gradient (S) [18]. This factor was computed using a modified adapted from [16], as utilized by [9]. This is shown in Eq. (4).

$$LS = (\text{flow acc} * \text{cellsize} / 22.13)^{0.4} * [(\sin(\text{slope} * 3.14 / 180)) / 0.0896]^{1.3} \quad (4)$$

Flow accumulation data were first acquired using QGIS's *r.fill* tool from the GRASS module, eliminating depressions in the DEM. Subsequently, the *r.watershed* function was employed to generate the accumulation raster. Calculating the watershed's slope necessitated a slope tool, so a clipped watershed DEM was employed as input in QGIS. Utilizing the raster calculator, LS was then computed using the *flow acc* and *slope* layers. This comprehensive approach to LS factor computation provided a more nuanced understanding of the topographical influences on erosion within the study area.

Crop management factor

The crop management (C) factor is a measure of the percent soil loss potential based on the protection afforded by different land-cover types. Higher C factor values indicate a lack of vegetation cover, thereby increasing the risk of significant soil loss. Vegetation cover is a key factor in erosion control since it intercepts precipitation, enhances infiltration, and mitigates the erosive impact of rainfall [18]. To quantify the C factor, we employed Eq. (5), introduced by [22] and employed by [9].

$$C = 0.431 - 0.805 * NDVI \quad (5)$$

The normalized difference vegetation index (NDVI) was fundamental in this calculation. This index was defined by [23] and employed in this work by means of Eq. (6).

$$NDVI = (NIR - RED) / (NIR + RED) \quad (6)$$

Here, NIR stands for the reflectance in the near-infrared band (%), while RED is the reflectance in the red band (%). B8 Sentinel-2 imagery corresponds to the NIR band, and B4 corresponds to the R band, both with a spatial resolution of 10 m. The NDVI was computed via QGIS's *i.vi* module, and the C factor, obtained using the Eq. (5), was obtained using the raster calculator. This approach offered a comprehensive understanding of how different land-cover types contribute to erosion risk within the study area.

Conservation practices factor

The conservation practices (P) factor assesses the efficacy of soil protection against erosion through management practices, as these practices reduce runoff and consequently limit soil loss [18]. In this work, the calculation of this factor involved the manual digitization of a Google Earth image within QGIS. A subsequent field visit was conducted to validate areas with soil and water conservation initiatives. Moreover, digitization was refined using land knowledge obtained during field campaigns. In QGIS, reclassification was conducted using the percentage values of the slope map, as well as zone-based statistics, to compute the mean slope value for each digitized polygon. In the attribute table of the generated polygons provided the corresponding P values. These values were determined based on Table I, i.e., according to the average slope grades and their

corresponding P factor values, as proposed by [24] and adopted by [25]. This methodological approach captured the nuanced interplay between management practices, slope gradients, and erosion mitigation within the study area.

Table I. P factor values for contouring, strip-cropping, and terracing focusing on slope classes

Slope (%)	Conservation Practices (P Factor)		
	Contouring	Strip Cropping	Terracing
0.0-7.0	0.55	0.27	0.10
7.0-11.3	0.60	0.30	0.12
11.3-17.6	0.80	0.40	0.16
17.6-26.8	0.90	0.45	0.18
>26.8	1	0.50	0.20

Source: [24]

Correlation between erosion and SOC storage

Soil samples were gathered during field visits in order to quantify SOC levels. From this sampling, 50 georeferenced points were randomly selected (Fig. 2) and positioned on the erosion map.

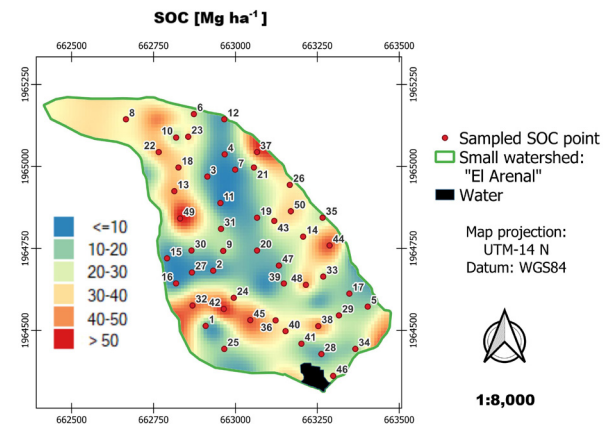


Figure 2. SOC map in Mg ha^{-1} . The 50 georeferenced points correlating erosion and SOC are shown in red.

Source: [12]

Using QGIS's point sampling tool, the estimated erosion values were derived for each point and associated with their corresponding erosion classification values (null, light, moderate, and high). The correlation analysis considered X (erosion, $\text{Mg ha}^{-1} \text{ year}^{-1}$) and Y (SOC, Mg ha^{-1}). To initiate the examination, a descriptive statistical analysis was performed, accompanied by the elaboration of a scatter plot (X,Y), a histogram paired with a normal curve, and a Q-Q plot to assess data normality. Normality was evaluated using the Shapiro-Wilk test [26]–[27]. Since our data did not follow a normal distribution, the Kendall *tau* correlation coefficient

test (KTCCT) was applied. The Kendall correlation coefficient and its corresponding p-values guided the decision between the two hypotheses, i.e., H_0 (no association) and H_1 (correlation). The Kendall coefficient ranges from -1 to 1, where positive values indicate a positive association and negative ones denote a negative association. The statistical computations presented in this section were conducted using the R Studio software [28]. This rigorous approach ensured a comprehensive analysis of the relationship between SOC and erosion levels, thereby yielding insights into their dynamic interdependence.

Results

The R factor is shown in Fig. 3a. The minimum, maximum, and average values for R and precipitation were (1302, 565), (1356, 579), and (1325, 571) in $\text{MJ}\cdot\text{mm}\cdot\text{h}^{-1}\cdot\text{ha}^{-1}\cdot\text{yr}^{-1}$ and mm, respectively. The upper regions of the small watershed exhibited the highest R values, which correlated with the zones of maximum precipitation.

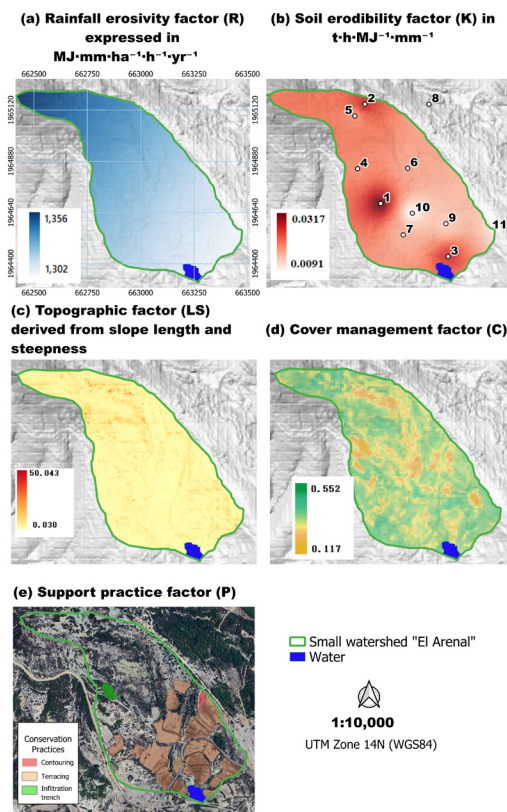


Figure 3. Spatial representation of the five input factors used in the RUSLE model to estimate soil erosion in the El Arenal micro-watershed
Source: Authors

In Fig. 3b, the K factor shows variation within the small watershed, with values ranging from 0.0091 to 0.0317 $\text{t}\cdot\text{ha}\cdot\text{h}\cdot\text{MJ}^{-1}\cdot\text{mm}^{-1}$ and an average of $0.0197 \text{ t}\cdot\text{ha}\cdot\text{h}\cdot\text{MJ}^{-1}\cdot\text{mm}^{-1}$.

Table II provides an overview of the sampled soil points, their corresponding land use and vegetation, the soil texture values, and the estimated K factor. Notably, the latter was the lowest in the bare soil areas and the highest in the agricultural zones.

Table II. K values estimated by soil texture class

ID	Land use and vegetation	Texture	K-RUSLE
1	Rainfed agriculture	Loam	0.0317
2	Thorn scrub	Clay loam	0.0264
3	Rainfed agriculture	Sandy clay loam	0.0263
4	Adult pine plantation	Sandy clay loam	0.0199
5	Gallery vegetation	Sandy clay loam	0.0196
6	Gallery vegetation	Sandy loam	0.0179
7	Rainfed agriculture	Clay loam	0.0177
8	Adult pine plantation	Sandy clay loam	0.0159
9	Rainfed agriculture	Sandy loam	0.0150
10	Bare soil	Sandy loam	0.0111
11	Bare soil	Sandy loam	0.0091

Source: Authors

The LS factor (Fig. 3c) shows values ranging from 0.030 to 23 045, with an average of 2.31. Notably, the elevated values were concentrated in regions denoted as *flow accumulation zones*, mainly along river courses (Fig. 4).

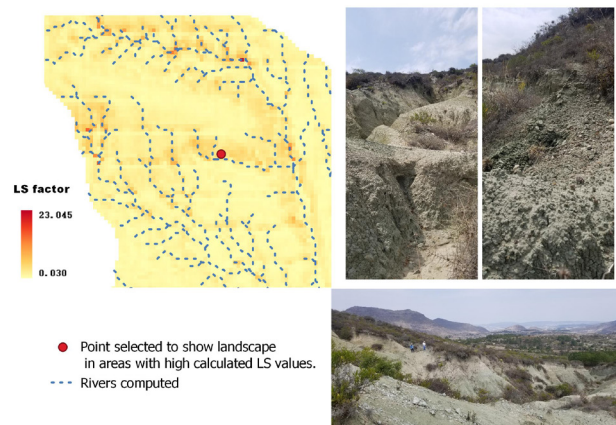


Figure 4. Point selected to show a landscape with high values of LS factor
Source: Authors

The C factor (Fig. 3d) exhibits a range of values spanning from 0.117 to 0.552, with an average of 0.322. These C values, corresponding to diverse soil and vegetation uses in the small watershed, indicate the following categories: bare soil (0.488), gallery vegetation (0.314), thorn scrub (0.328), rainfed agriculture (0.295), and adult pine plantation (0.129).

The P factor in Fig. 3e corresponds to slope-related values. A total of 18 polygons were digitized under field supervision to obtain these values. Table III illustrates the direct relationship between an increasing slope and a higher P factor.

Table III. P values associated with slope (%)

Polygon	PCWS	Slope (%)	P factor
1	None	16.1	1
2	Land terrace	26.48	0.2
3	Land terrace	28.97	0.2
4	Land terrace	17.02	0.18
5	Land terrace	17.44	0.18
6	Land terrace	17.82	0.18
7	Land terrace	11.13	0.16
8	Contouring	12.08	0.16
9	Land terrace	14.63	0.16
10	Land terrace	8.81	0.12
11	Land terrace	9.13	0.12
12	Land terrace	9.55	0.12
13	Land terrace	5.24	0.1
14	Land terrace	5.87	0.1
15	Land terrace	6.02	0.1
16	Land terrace	6.74	0.1
17	Infiltration trench	7.05	0.1

Source: Authors

Spatial modelling of erosion estimation

Fig. 5 shows the spatial distribution of soil erosion in the study area. The minimum, maximum, and average values were 0.19, 266.99, and 17.44 $\text{Mg ha}^{-1} \text{ year}^{-1}$, respectively. The cumulative erosion for the entirety of the watershed was estimated at 305 693 t. The average erosion-related SOC values (Mg ha^{-1}) by classification are as follows: null (31.79), light (22.36), moderate (16.19), and high (5.22).

The average erosion values (in $\text{Mg ha}^{-1} \text{ year}^{-1}$) associated with different soil and vegetation types are as follows: bare soil (55.02), thorn scrub (29.50), gallery vegetation (17.84), adult pine plantation (5.76), and rainfed agriculture (2.31). The lowest erosion value is linked to flat slopes within agricultural areas that have implemented soil and water conservation practices such as land terracing. On the other hand, erosion values exceeding 200 $\text{Mg ha}^{-1} \text{ year}^{-1}$ are concentrated in the northern zone of the small watershed, characterized by high R and LS values, as well as by the absence of conservation practices.

Correlation between erosion and SOC storage

The dataset, comprising 50 points, can be observed in Fig. 6 and Table V, which detail the estimated erosion and measured SOC values. A visual comparison of erosion and SOC is depicted in Fig. 6 by means of a bar graph. This graph shows that in 35 out of 50 instances (70% of the sample), as erosion decreases, the SOC tends to increase.

Estimated Soil Erosion Rates (RUSLE) [$\text{Mg ha}^{-1} \text{ yr}^{-1}$]

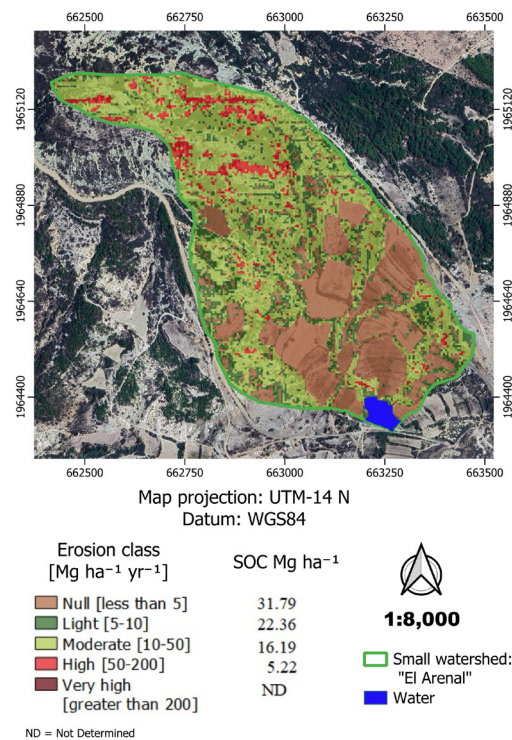


Figure 5. Spatial distribution of the estimated soil erosion rates ($\text{Mg ha}^{-1} \text{ yr}^{-1}$) and soil organic carbon (SOC) content (Mg ha^{-1}) across the El Arenal micro-watershed

Source: Authors

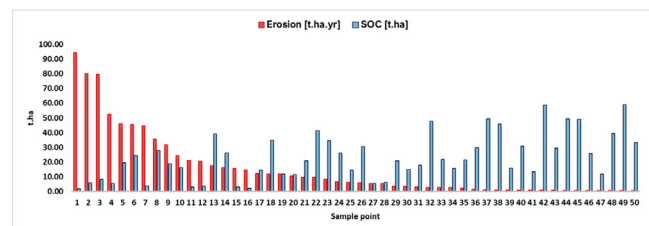


Figure 6. Graphical comparison of erosion and SOC

Source: Authors

A scatter plot was created to illustrate the connection between erosion and SOC. This plot reveals a relatively weak correlation, with an R^2 value of 0.21. Descriptive statistics for both variables are presented Table IV. Notably, the mean for both variables exceeds the median, indicating an asymmetric distribution.

Table IV. Descriptive statistics and Shapiro-Wilk test for erosion and SOC

Statistics	Erosion (X)	SOC (Y)
Number	50.00	50.00
Addition	770.63	1 158.21
Half	15.41	23.16
Median	6.10	20.60
Stdvd (pop)	22.12	15.53
Stdvd (sample)	22.34	15.69
Minimum	0.19	1.58

Maximum	94.40	59.07
Range	94.21	57.49
Minority	0.19	1.58
Most	0.19	1.58
Variety	50.00	50.00
Q1	0.87	11.57
Q3	17.62	33.32
IQR	16.75	21.75
Missing (null) values	0.00	0.00
P-value (Shapiro W)	6.907e-9	0.006948
W (Shapiro W)	0.6956	0.9326

Source: Authors

Fig. 7 shows the statistical relationship between soil erosion and SOC. A moderate negative correlation was observed ($r=-0.46$), with the SOC decreasing as erosion increased (Fig. 7a). The erosion data exhibited a right-skewed distribution (Fig. 7b), which was confirmed by the deviation from normality in the Q-Q plot (Fig. 7c). In contrast, the SOC values followed a more symmetrical and near-normal distribution, as shown in Figs. 7d and 7e.

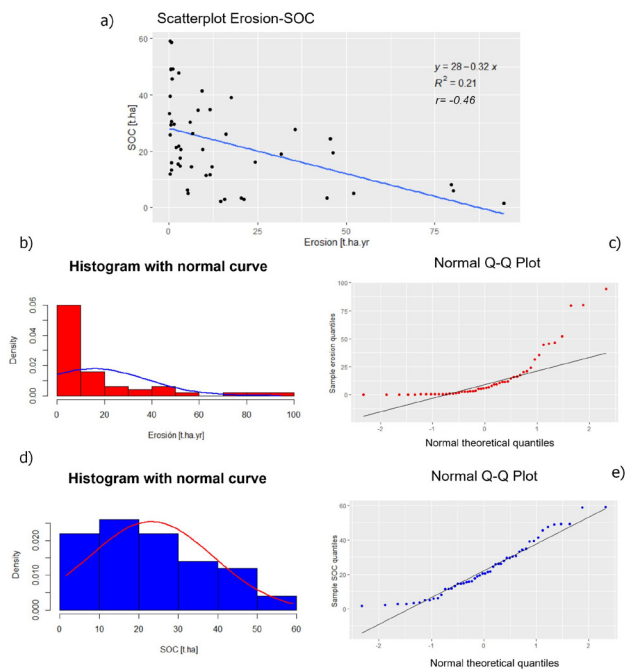


Figure 7. An asymmetric distribution was observed in both the histograms and Q-Q plots. a) Scatterplot illustrating the negative linear correlation between soil erosion ($\text{Mg ha}^{-1} \text{ yr}^{-1}$) and SOC (Mg ha^{-1}), including the regression line, the equation, the coefficient of determination ($R^2=0.21$), and Pearson's correlation coefficient ($r=-0.46$). b) Histogram with the density curve of the erosion values, indicating a right-skewed distribution. c) Normal Q-Q plot showing that the erosion values deviate from normality, particularly in the upper quantiles. d) Histogram with a density curve of SOC values, approximating a normal distribution. e) Normal Q-Q plot confirming that the SOC values closely follow a normal distribution.

Source: Authors

Regarding the normality test (Shapiro-Wilk), the results for the variables X and Y were as follows: $W(50) = 0.70$, $p <$

0.001 ; $W(50) = 0.93$, $p = 0.007$. This indicates that the data do not follow a normal distribution in either case. Therefore, the data can be analyzed using non-parametric tests.

As for the non-parametric data, Kendall's rank correlation coefficient was utilized. The results for Kendall's τ are as follows: $z = -3.9566$, $p\text{-value} = 7.603\text{e-}05$, and $\tau = -0.386$. This value of τ supports the alternative hypothesis, as it is not equal to zero. The results demonstrate a negative correlation, implying that the variables move in opposite directions

Table V presents the details of the 50 sampled points, i.e., data on erosion, SOC storage, land use, vegetation, and soil and water conservation practices. This table reveals that areas with higher erosion rates tend to exhibit a lower SOC storage. Specifically, points 38 to 50 show a higher frequency of terraces, which aligns with agricultural areas and corresponds to reduced soil loss.

Table V. Sampled points and their corresponding data

ID	E	SOC	LSV	SWCP	ID	E	SOC	LSV	SWCP
1	94.40	1.58	Bs	N	26	5.88	30.38	App	N
2	80.20	6.02	Ts	N	27	5.33	5.11	Ts	N
3	79.53	8.17	Ts	N	28	5.27	6.24	Ra	T
4	52.11	5.10	Gv	N	29	3.43	20.56	Ra	T
5	46.25	19.35	Ts	N	30	3.15	14.80	Ts	N
6	45.56	24.37	Ts	N	31	3.12	17.68	Gv	N
7	44.62	3.52	Ts	N	32	2.79	47.66	Ra	T
8	35.63	27.78	Ts	N	33	2.69	21.79	Ra	T
9	31.66	18.88	Gv	N	34	2.62	15.45	Ra	T
10	24.28	16.16	Gv	N	35	2.03	21.36	Ra	C
11	21.05	2.91	Ts	N	36	1.39	29.63	Gv	N
12	20.30	3.49	Ts	N	37	1.08	49.20	App	N
13	17.62	39.13	Ts	N	38	0.87	45.68	Ra	T
14	16.10	26.09	Ts	N	39	0.77	15.83	Ra	T
15	15.65	2.88	Bs	N	40	0.66	30.62	Ra	T
16	14.54	2.17	Ra	N	41	0.65	13.35	Ra	T
17	12.11	14.52	App	N	42	0.63	58.69	Ra	T
18	11.65	34.82	Ts	N	43	0.59	29.48	Ra	T
19	11.53	11.57	Gv	N	44	0.55	49.23	Ra	T
20	10.51	11.38	Gv	N	45	0.50	49.03	Ra	T
21	9.51	20.65	App	N	46	0.35	25.77	Ra	T
22	9.30	41.33	Ts	N	47	0.30	11.82	Ra	T

23	8.16	34.44	Gv	N	48	0.29	39.45	Ra	T
24	6.65	26.20	Ra	N	49	0.25	59.07	App	IT
25	6.33	14.56	Ts	N	50	0.19	33.32	Ra	T

ID: identification of the sampled point, E: erosion ($\text{Mg ha}^{-1} \text{ year}^{-1}$), SOC: soil organic carbon (Mg ha^{-1}), LSV: land use and vegetation, SWCP: soil and water conservation practices, Bs: bare soil, Ts: thorn scrub, Gv: gallery vegetation, Ra: rainfed agriculture, App: adult pine plantation, N: none, T: terracing, C: contouring, IT: infiltration trench.

Source: Authors



Figure 8. View of some representative sites.

Source: Authors

Discussion

The results of this study provided significant insights into the erosion-SOC relationship. Kendall's τ value (-0.39), for instance, revealed a negative correlation. This implies that, as erosion increases, SOC tends to decrease. This negative correlation can be attributed to the erosion process, including initiation, transport, and deposition, which can

impact the loss and gain of SOC within the watershed. [1] highlighted the significant impact of water erosion on the transport of organic carbon. This impact begins with the decomposition of aggregates, leading to the exposure of SOC. During this process, colloidal fractions such as fine silt, clay, and SOC are eliminated. The sediment, along with carbon, undergoes redistribution and is eventually deposited in depression areas. In our study area, this phenomenon was primarily observed in areas with greater soil loss, which tend to coincide with concave zones such as slopes. Regarding these sites, [29] indicated that a high cumulative intensity of water erosion reduces the vegetation cover through the formation of streams or ravines, decreasing the thickness of the soil and eliminating the upper layer of fertile soil, which in turn leads to a decrease in SOC. The RUSLE model identified these areas as having higher LS and C values, predominantly consisting of bare soil and thorn scrub (Fig. 3).

Studies suggest an initial loss of SOC during erosion. Nevertheless, this lost carbon is transported to flatter areas [1], [29], [30]. In our study area, this trend is aligned with the lower part of the watershed where agricultural areas are located. According to [31] and [32], these locations also correspond to the presence of soil and water conservation practices, which create conditions for carbon storage.

Previous research conducted in the same watershed [12] demonstrated that SOC storage is strongly influenced by land cover and soil type. For example, gallery vegetation and pine plantations exhibited higher SOC values (up to 31.8 Mg ha^{-1}), while intensively cultivated areas stored much lower values (around 5.2 Mg ha^{-1}). These patterns coincided with soil classification: Vertisols with fine textures supported a greater SOC retention, whereas Litosols and sandy soils were associated with a reduced SOC content. These findings are consistent with broader studies. [6] reported that nearly 50% of Mexico's SOC is concentrated in the upper 30 cm of the soil, which makes it highly vulnerable to erosion. Similarly, [7] emphasized that soil erosion processes can significantly deplete SOC stocks through the detachment and oxidation of OM. [31] also observed higher SOC values in vegetated zones when compared to croplands, and [32] identified land use as a key factor influencing SOC accumulation, even in terraced systems. Our findings align with these observations: areas with high erosion – mainly bare soils and scrublands – are linked to low SOC values, while depositional zones with conservation practices such as terracing promote SOC accumulation. These outcomes underscore the importance of considering both edaphic and vegetation factors in land management strategies aimed at improving carbon retention and reducing erosion at the watershed scale.

This is line with [2]: erosion-induced carbon storage is supported by two key mechanisms. First, dynamic replacement involves the replacement of eroded carbon with new carbon, such as OM from vegetation. This is especially relevant in cultivated areas. Second, reduced decay rates in depositional environments contribute to long-term carbon storage.

This increased carbon storage takes place through a combination of dynamic replacement and reduced decomposition, which ultimately contributes to an overall carbon storage increase, despite the initial carbon loss due to erosion, as seen in sites 1 and 3 (Fig. 7 and Table V). These dynamics reinforce the notion that erosion culminates in deposition, particularly in agricultural zones. Conservation practices like terracing amplify carbon storage benefits, as exemplified in sites 38 and 42. These findings align with [30], who explains how erosion and deposition redistribute sediment and SOC in agricultural landscapes.

However, it must be acknowledged that, while the correlation exists, it is relatively weak. Some cases, such as sites 39 and 47, showed a lower carbon content in agricultural areas with terraces. Here, the absence of dynamic replacement due to limited recent cultivation (uncultivated agricultural land) prevented SOC storage. Despite soil and water conservation efforts, these sediment-filled areas retain less SOC, which underscores the importance of active soil management in promoting dynamic carbon replacement. According to the study conducted by [32], SOC storage resulting from terracing is primarily influenced by land use.

This discussion highlights the intricate interactions between erosion, sedimentation, and carbon storage within landscapes, emphasizing that natural processes can yield unexpected carbon dynamics. It is crucial to consider the importance of accounting for both erosion and deposition when evaluating carbon budgets. However, one limitation of this research lies in the need to validate the spatial erosion model in the field.

Future research could investigate whether the SOC within sediment layers buried deep in soil and water conservation areas qualifies as a CO₂ sink. This proposition arises from the stabilization of carbon within oxygen-limited environments, facilitated by organisms responsible for decomposition, in line with [33].

Conclusion

This study examined the spatial relationship between water erosion and SOC storage in a small watershed using the RUSLE model and geostatistical analysis. According to the initial hypothesis, erosion intensity is negatively related to SOC levels. In approximately 70% of our data, lower erosion rates corresponded to higher SOC values, with a Kendall tau correlation coefficient of -0.39. This inverse trend was attributed to erosion-induced aggregate breakdown and SOC detachment, followed by its redistribution across the landscape.

Despite the observed relationship, the correlation was moderate, suggesting that other variables such as vegetation cover, land use, and soil texture also influence SOC dynamics. Our findings demonstrate that erosion does not always lead to an irreversible loss of SOC. Instead, it can contribute to the storage of carbon in depositional zones

under favorable conditions. In low-oxygen environments, mechanisms such as dynamic carbon replacement from vegetation and reduced decomposition rates enhance SOC retention, particularly in areas managed with conservation practices such as terracing.

The findings of this study demonstrate the importance of integrating erosion control and carbon management into land use planning. In the future, it would be beneficial to explore whether the SOC stored in buried sediment layers under conservation structures can act as a long-term CO₂ sink, which could contribute to the mitigation of climate change.

CRediT author statement

- OTAS conceived and elaborated the manuscript, conducted field data collection, performed spatial and statistical data processing, and wrote the original draft.
- JPN provided overall conceptual direction, supervised the methodological framework, and critically revised the manuscript.
- CAAS actively contributed to the content review and substantially improved the structure and clarity of the article.
- AIMR supported the review process, particularly regarding results interpretation and scientific framing.

Conflicts of interest

There is no conflict of interest to declare.

Data availability

The data used in this study can be provided upon request to the corresponding author.

References

- [1] R. Lal, "Fate of soil carbon transported by erosional processes," *Appl. Sci.*, vol. 12, no. 1, p. 48, Dec. 2021. <http://doi.org/10.3390/app12010048>
- [2] S. Doetterl, A. A. Berhe, E. Nadeu, Z. Wang, M. Sommer, and P. Fiener, "Erosion, deposition and soil carbon: A review of process-level controls, experimental tools and models to address C cycling in dynamic landscapes," *Earth-Sci. Rev.*, vol. 154, pp. 102–122, Mar. 2016. <http://doi.org/10.1016/j.earscirev.2015.12.005>
- [3] R. Lal, "Accelerated soil erosion as a source of atmospheric CO₂," *Soil Tillage Res.*, vol. 188, pp. 35–40, May 2019. <http://doi.org/10.1016/j.still.2018.02.001>
- [4] C. Lefèvre, F. Rekik, V. Alcantara, and L. Wiese, *Soil organic carbon: The hidden potential*. Rome, Italy: Food and Agriculture Organization of the United Nations, 2017.
- [5] H. Burbano Orjuela, "El carbono orgánico del suelo y su papel frente al cambio climático," *Rev. Cienc. Agríc.*, vol. 35, no. 1, p. 82, Jun. 2018. <http://doi.org/10.22267/rcia.183501.85>

- [6] F. Paz and J. Etchevers, "Distribución a profundidad del carbono orgánico en los suelos de México," *Terra Latinoam.*, vol. 34, no. 3, pp. 339–335, 2016. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57792016000300339&lng=es&tlng=es
- [7] K. R. Olson, M. Al-Kaisi, R. Lal, and L. Cihacek, "Impact of soil erosion on soil organic carbon stocks," *J. Soil Water Conserv.*, vol. 71, no. 3, pp. 61A–67A, May 2016. <http://doi.org/10.2489/jswc.71.3.61A>
- [8] L. Cui, X. Li, J. Lin, G. Guo, X. Zhang, and G. Zeng, "The mineralization and sequestration of soil organic carbon in relation to gully erosion," *CATENA*, vol. 214, art. 106218, Jul. 2022. <http://doi.org/10.1016/j.catena.2022.106218>
- [9] D. B. Tiruwa, B. R. Khanal, S. Lamichhane, and B. S. Acharya, "Soil erosion estimation using geographic information system (GIS) and revised universal soil loss equation (RUSLE) in the Siwalik Hills of Nawalparasi, Nepal," *J. Water Clim. Change*, vol. 12, no. 5, pp. 1958–1974, Aug. 2021. <http://doi.org/10.2166/wcc.2021.198>
- [10] INEGI (National Institute of Statistics and Geography), "Soil erosion dataset, scale 1:250 000 series I, national continuum," 2014. [Online]. Available: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825004223>
- [11] INEGI (National Institute of Statistics and Geography), "Soil map, scale 1:250,000. National continuum," 2002-2006. [Online]. Available: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=794551131916>
- [12] O. Aguirre-Salado, J. Pérez-Nieto, C. Aguirre-Salado, and A. Monterroso-Rivas, "Factors regarding the spatial variability of soil organic carbon in a Mexican small watershed," *Rev. Fac. Agron. Univ. Zulia*, vol. 41, no. 1, art. e244101, Dec. 2023. [http://doi.org/10.47280/RevFacAgron\(LUZ\).v41.n1.01](http://doi.org/10.47280/RevFacAgron(LUZ).v41.n1.01)
- [13] CLICOM, "Daily climate data from the CLICOM of the SMN through its CICESE web platform," 2023. [Online]. Available: <https://cucapa-clicom.cicese.mx/mapa.html>
- [14] A. Walkley and C. A. Black, "An examination of the Degtjareff method for determining soil organic matter and a proposed modification of chromic acid titration method," *Soil Sci.*, vol. 37, pp. 29–38, 1934.
- [15] INEGI (National Institute of Statistics and Geography), "Digital surface-type elevation model with 5m resolution derived from satellite and airborne remote sensing data," 2017. [Online]. Available: <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463542605>
- [16] W. H. Wischmeier and D. D. Smith, *Predicting rainfall erosion losses – A guide to conservation planning*. Washington, DC, USA: Department of Agriculture, Science and Education Administration, 1978.
- [17] K. G. Renard, G. R. Foster, G. A. Weesies, D. K. McCool, and D. C. Yoder, *Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (RUSLE)*. Washington, DC, USA: US Government Printing Office, 1997.
- [18] C. A. Aguirre-Salado et al., "Improving identification of areas for ecological restoration for conservation by integrating USLE and MCDA in a GIS-environment: A pilot study in a priority region northern Mexico," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 9, art. 262, Aug. 2017. <http://doi.org/10.3390/ijgi6090262>
- [19] K. G. Renard and J. R. Freimund, "Using monthly precipitation data to estimate the R-factor in the revised USLE," *J. Hydrol.*, vol. 157, no. 1, pp. 287–306, 1994.
- [20] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel B-splines," *IEEE Trans. Vis. Comput. Graph.*, vol. 3, no. 3, pp. 228–244, Sep. 1997. <http://doi.org/10.1109/2945.620490>
- [21] N. Efthimiou, "The new assessment of soil erodibility in Greece," *Soil Tillage Res.*, vol. 204, art. 104720, Oct. 2020. <http://doi.org/10.1016/j.still.2020.104720>
- [22] S. M. De Jong, L. C. Brouwer, and H. T. Riezebos, "Erosion hazard assessment in the La Peyne Catchment, France," Department of Physical Geography, University of Utrecht, Utrecht, The Netherlands, 1998. [Online]. Available: <https://research.wur.nl/en/publications/erosion-hazard-assessment-in-the-la-peyne-catchment-france>
- [23] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, pp. 127–150, 1979.
- [24] G. Shin, "The Analysis of Soil Erosion Analysis in Watershed Using GIS," PhD dissertation, Gang-Won National University, Chuncheon, Korea, 1999. [Online]. Available: <https://www.scirp.org/reference/referencespapers?referenceid=3870069>
- [25] F. Karamage, C. Zhang, T. Liu, A. Maganda, and A. Isabwe, "Soil Erosion Risk Assessment in Uganda," *Forests*, vol. 8, no. 2, p. 52, Feb. 2017. <http://doi.org/10.3390/f8020052>
- [26] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.
- [27] H. Hernandez, "Testing for Normality: What is the Best Method?," 2021. [Online]. Available: <https://doi.org/10.13140/RG.2.2.13926.14406>
- [28] RStudio Team, "RStudio: Integrated development for R," 2023. [Online]. Available: <http://www.rstudio.com/>
- [29] L. Liu, Q. Zhang, Q. Liu, and Z. Li, "Is soil an organic carbon sink or source upon erosion, transport and deposition?," *Eur. J. Soil Sci.*, vol. 74, no. 1, art. e13344, Jan. 2023. <http://doi.org/10.1111/ejss.13344>
- [30] F. M. S. A. Kirkels, L. H. Cammeraat, and N. J. Kuhn, "The fate of soil organic carbon upon erosion, transport and deposition in agricultural landscapes – A review of different concepts," *Geomorphology*, vol. 226, pp. 94–105, Dec. 2014. <http://doi.org/10.1016/j.geomorph.2014.07.023>
- [31] E. Bojago, M. W. Delango, and D. Milkias, "Effects of soil and water conservation practices and landscape position on soil physicochemical properties in Anuwa watershed, Southern Ethiopia," *J. Agric. Food Res.*, vol. 14, art. 100705, Dec. 2023. <http://doi.org/10.1016/j.jafr.2023.100705>
- [32] D. Chen, W. Wei, S. Daryanto, and P. Tarolli, "Does terracing enhance soil organic carbon sequestration? A national-scale data analysis in China," *Sci. Total Environ.*, vol. 721, art. 137751, Jun. 2020. <http://doi.org/10.1016/j.scitotenv.2020.137751>
- [33] M. Mekonnen and M. Getahun, "Soil conservation practices contribution in trapping sediment and soil organic carbon, Minizir watershed, northwest highlands of Ethiopia," *J. Soils Sed.*, vol. 20, no. 5, pp. 2484–2494, May 2020. <http://doi.org/10.1007/s11368-020-02611-5>

Activation of Coal Fly Ash for Cadmium Wastewater Remediation

Activación de ceniza volante de carbón para la remediación de aguas residuales contaminadas con cadmio

Yareli Ginalis Medina ¹ and Roberto Flores ²

ABSTRACT

Industrial wastewater contamination by heavy metals is a major environmental issue. Various techniques and materials have been proposed to address this problem, including those related to the adsorption process. However, new, improved, and low-cost materials must be developed and proposed in order for these strategies to be competitive. This study aims to investigate the remediation of water contaminated with cadmium, *i.e.*, Cd(II), using activated coal fly ash, a low-cost sorbent, as it is the byproduct of an industrial process. Coal fly ash was chemically treated in acidic and alkaline mediums and activated using ultrasonic energy to enhance the materials and increase the uptake of Cd(II). Controlling the pH was found to be crucial, as the maximum sorption capacity occurred at pH 6. The materials activated in an alkaline medium with ultrasound were able to adsorb more significant amounts of Cd(II) under the studied experimental conditions and over four cycles of adsorption experiments. Finally, the kinetics of the adsorption process were analyzed, and some mathematical kinetics models were proposed to simulate the experimental data. After statistical discrimination, the Elovich isotherm was selected to represent the adsorption of Cd(II) in the different materials studied.

Keywords: coal fly ash, ultrasonic activation in alkaline mediums, cadmium wastewater remediation, pH effect, kinetics parameter estimation, Elovich isotherm

RESUMEN

La contaminación de aguas residuales industriales por metales pesados es un problema medioambiental importante. Se han propuesto varias técnicas y materiales para abordar esta problemática, incluyendo aquellas relacionadas con el proceso de adsorción. No obstante, es necesario desarrollar y proponer materiales nuevos, mejorados y de bajo costo para que estas estrategias sean competitivas. Este estudio tiene como objetivo investigar la remediación de agua contaminada con cadmio, *i.e.*, Cd(II), utilizando cenizas volantes de carbón activado, un sorbente de bajo costo, pues es un subproducto de un proceso industrial. Las cenizas volantes de carbón se trataron químicamente en medios ácidos y alcalinos y se activaron mediante energía ultrasónica para mejorar su capacidad de adsorción de Cd(II). El control del pH resultó crucial, ya que la máxima capacidad de adsorción se produjo a un pH de 6. Los materiales activados en medio alcalino mediante ultrasonido pudieron adsorber cantidades más significativas de Cd(II) bajo las condiciones experimentales estudiadas y en cuatro ciclos de adsorción. Finalmente, se analizó la cinética del proceso de adsorción y se propusieron algunos modelos cinéticos matemáticos para simular los datos experimentales. Después de una discriminación estadística, se seleccionó la isoterma de Elovich para representar la adsorción de Cd(II) en los diferentes materiales estudiados.

Palabras clave: ceniza volante de carbón, activación con ultrasonido en medio alcalino, remediación de agua de desecho con cadmio, efecto de pH, estimación de parámetros cinéticos, isoterma de Elovich

Received: May 29th, 2024

Accepted: May 8th, 2025

Introduction

Cadmium – Cd(II) – contamination in agricultural soils has been shown to affect plant growth and health [1]–[5]. High levels of Cd(II) significantly reduce crop yield, potentially resulting in plant mortality. Cd(II) also alters plant height, root length, and the fresh weight of roots, stems, and leaves. In addition, it changes the uptake and transport of essential nutrients, resulting in nutrient deficiencies within the plants. Cd(II) toxicity impedes seed germination, plant growth, photosynthetic efficiency, and plant protein homeostasis, and it has the propensity to induce oxidative damage and generate reactive oxygen species within plant tissues.

Soil Cd(II) pollution significantly affects microbial communities and ecological processes, leading to soil microbial diversity and composition alterations. Bacterial diversity is more severely impacted compared to fungal diversity [6], [7]. The accumulation of Cd(II) in the soil also impedes the uptake and translocation of water and essential nutrients in plants, adversely affecting their health and

¹ PhD in Engineering – Chemical Engineering, Texas A&M University, USA. Affiliation: Associate profesor, Universidad Autónoma del Estado de Morelos, Facultad de Ciencias Químicas e Ingeniería, México. Email: roberto.flores@uaem.mx

² BSc in Engineering – Chemical Engineering, Universidad Autónoma del Estado de Morelos, México. Email: yareli.medina@uaem.edu.mx



Attribution 4.0 International (CC BY 4.0) Share - Adapt

development. This is attributed to the harmful modulation of genes associated with metal resistance, carbon fixation, nitrification, and denitrification [3], [8]. The consequences of Cd(II) contamination extend beyond chemical toxicity and influence complex biological interactions and ecological dynamics within the soil's ecosystem.

Cd(II) pollution impacts human health, with occupational and environmental exposure leading to various ailments and diseases, including nephrotoxicity, osteotoxicity, teratogenicity, endocrine disruption, hepatotoxicity, and cancers of the breast, lung, prostate, nasopharynx, pancreas, and kidney [9], [10]. Cd(II) disrupts the body's antioxidant defense system, inducing oxidative stress and cellular or molecular alterations, thus contributing to liver and kidney diseases [11], [12].

The maximum concentration limit of Cd(II) in wastewater varies depending on the source and prevailing regulatory standards. Generally, permissible Cd(II) concentrations in wastewater fall within the limits set by regulatory bodies such as the WHO and the EPA [13]. OSHA has established a legal exposure limit of $5 \mu\text{g m}^{-3}$ of Cd(II) as an eight-hour time-weighted average. In Mexico, NOM-201-SSA1-2015, referencing the WHO, sets the maximum permissible limits for Cd(II) in water for human consumption at 0.003 mg L^{-1} .

Wastewater contaminated with Cd(II) can originate from various sources such as mining, residues from industrial activity (e.g., pharmaceuticals), atmospheric deposition, metal plating, sedimentation, agricultural activities, and stormwater runoff [12], [14]–[16].

Several technologies have been developed for the remediation of Cd(II)-contaminated waters, including adsorption [17]–[21], biosorption [22]–[24], electrochemical precipitation [25]–[27], ion exchange [28]–[30], and bioremediation [16], [24], [31]. However, Cd(II) wastewater remediation faces various challenges. Firstly, as a non-biodegradable heavy metal, Cd(II) persists in the environment, making its removal a challenging task. Additionally, conventional Cd(II) removal techniques, such as chemical precipitation and ion exchange, are costly and produce toxic sludges as secondary contaminants. Furthermore, the methods of existing wastewater treatment plants need to be improved to effectively remove Cd(II).

Cd(II)-containing wastewater can be remediated using fly ash as an adsorbent. Numerous studies have investigated the use of modified fly ash – either through acid or alkaline activation, or combined with other species – for cadmium removal from aqueous solutions, demonstrating its significant potential with efficiencies exceeding 90% [32]–[34].

Fly ash offers several advantages for wastewater remediation as a low-cost, recyclable material that can be easily obtained as a byproduct of thermal power plants. Its high porosity, surface area, and chemical composition make it an efficient material for removing contaminants such as heavy metals, organic compounds, and dyes from wastewater [35],

[36]. The acid activation of fly ash significantly affects its properties by increasing Si/Al ratios, eliminating impurities and dealuminations, and increasing the total pore volume and surface area [37], [38]. On the other hand, the alkaline activation of fly ash leads to the formation of a sodium aluminosilicate geopolymer with enhanced heavy metal adsorption capacity compared to normal ash [39], [40]. Another alternative to improving fly ash properties for heavy metal sorption corresponds to the application of ultrasonic energy [41]–[43].

This study investigates the potential use of coal fly ash as an adsorbent material for treating wastewater contaminated with Cd(II). Coal fly ash has been chosen because it is a waste product from thermoelectric power plants that could be utilized as a raw material in a new process, which aligns with the principles of the circular economy. The coal fly ash used in this work was treated with ultrasound and chemicals in acidic and alkaline media to increase its reactivity.

Materials and methods

Our coal fly ash was obtained from a pulverized coal-fired power plant in the north of Mexico, and it underwent multiple washes with distilled water to remove all soluble compounds. Afterwards, it was dried at 120°C overnight and stored at room temperature. Its chemical composition has been reported elsewhere [42].

A set of coal fly ash samples was subjected to ultrasonic treatment for 4 h at room temperature using a Cole Parmer 8890R-MTH ultrasonic bath. 10 g of fly ash were placed in 100 mL of distilled water. Ultrasonic energy was applied at a frequency of 47 kHz and an intensity of 147 W for 20 min on and 20 min off. Following ultrasonic treatment, the samples were dried at 120°C overnight and stored at room temperature.

Another group of samples was activated in an acidic medium by suspending 10 g in a 100 mL 5 M nitric acid (HNO_3) solution for 4 h, with continuous mechanical stirring or ultrasonic energy. After activation, the samples were washed several times with distilled water to remove all remaining chemicals. Then, they were dried at 120°C overnight and stored at room temperature.

The last set of samples was subjected to alkaline activation by mixing 10 g in a 100 mL 5 M sodium hydroxide (NaOH) solution. This process was similar to the acidic activation method, with the only difference being that NaOH was used instead of an HNO_3 solution. The samples were prepared using either mechanical stirring or ultrasonic energy.

Synthetic wastewater was prepared by dissolving $\text{Cd}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ (purchased from Aldrich and used as received) in distilled water to obtain a nominal Cd(II) composition of 500 ppm, simulating mining wastewater that is highly contaminated before it reaches freshwater bodies.

The experiments were conducted in duplicate as follows. 100 mL of synthetic wastewater solution were mixed with 5 g of sorbent material in a beaker for 60 min. At different time intervals, 1 mL aliquots were taken and analyzed in accordance with the ASTM D 4691-02 Method (*Standard practice for measuring elements in water by flame atomic absorption spectrophotometry*) to determine the Cd(II) concentration in the solution. Different pH conditions were studied and adjusted using a 20 wt% HNO_3 solution.

The Cd(II) sorption (q_t) in mg Cd g^{-1} sample was estimated as follows:

$$q_t = \frac{(C_i - C_t) \cdot V}{m} \quad (1)$$

where C_i is the initial Cd(II) concentration in mg L^{-1} ; C_t is the Cd concentration in mg L^{-1} at time t ; V is the volume of Cd(II) solution in mL; and m is the weight of the sorbent in g.

Results

Effect of pH on normal and ultrasonically activated fly ash

Cd(II) adsorption experiments were conducted on the original fly ash at different pH levels. In these experiments, samples of the contaminated water were taken at various intervals (between 0 and 60 min), and the Cd(II) concentration was measured. Through the calculations explained earlier, the quantity of Cd(II) adsorbed on the fly ash per unit of mass was determined. The results are presented in Fig. 1. A slight increase is observed as the pH increases from 4 to 6. At a pH of 6, the adsorption reaches its maximum, and, after 60 min, the amount adsorbed on the fly ash is 2.9 mg Cd g^{-1} of fly ash. When the pH increases to 7, the amount of Cd(II) adsorbed decreases, and, after 60 min, the adsorbed amount is 2.4 mg Cd g^{-1} of fly ash. This is consistent with the results presented in previous studies, which have shown the adsorption of Cd(II) in fly ash to be most effective for pH values between 4 and 7 [12], [44]–[46].

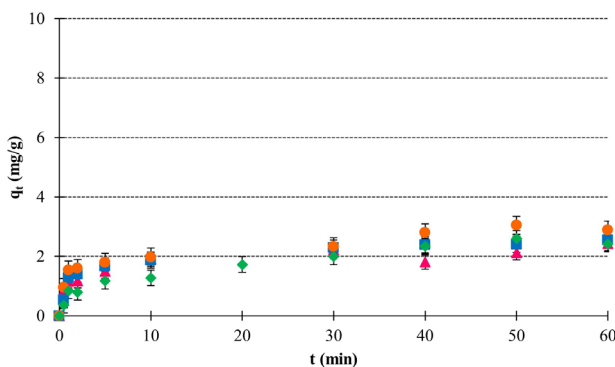


Figure 1. Experimental results of Cd adsorption on Normal fly ash. ▲ pH 4, ■ pH 5, ● pH 6, ◆ pH 7.
Source: Authors

When fly ash was activated with ultrasonic energy, the amount of Cd(II) adsorbed increased significantly in comparison with its normal counterpart (Fig. 2). However, the behavior of the solution's pH was similar to that observed with normal fly ash. As the pH increased from 4 to 6, the adsorbed amount of Cd(II) increased, reaching a maximum of 9.8 mg Cd g^{-1} of fly ash after 60 min. When the pH increased to 7, the adsorbed amount of Cd(II) decreased, and, after 60 min, it was 6.7 mg Cd g^{-1} of fly ash. The cavitation phenomenon, which enhances the textural characteristics of fly ash, is responsible for the observed increase in the adsorption capacity of Cd(II) in fly ash when activated with ultrasonic energy, as previously observed [41]–[43].

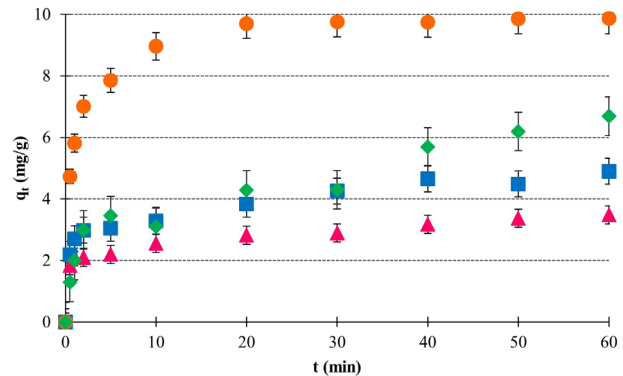


Figure 2. Experimental results of Cd adsorption on fly ash activated with ultrasonic energy. ▲ pH 4, ■ pH 5, ● pH 6, ◆ pH 7.
Source: Authors

Activation of fly ash in acidic and alkaline medium by applying ultrasonic energy

As previously explained, fly ash samples were also activated in both acidic and alkaline mediums. Two activation methods were used: in one method, the samples were mechanically shaken, and, in the other, ultrasonic energy was applied during activation. After this process, the samples were dried at 120°C overnight. The Cd(II) adsorption capacity of the fly ash samples prepared at pH values of 6 and 7 was studied, as these pH levels had shown the best Cd(II) adsorption results in the original fly ash and in the samples activated with ultrasonic energy alone. The results obtained are presented in Figs. 3 and 4.

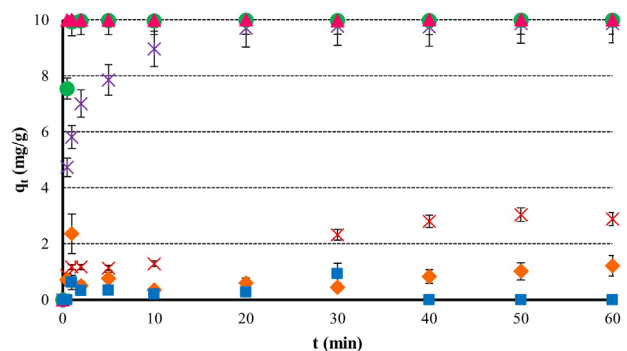


Figure 3. Experimental results of Cd adsorption at pH 6. × Normal fly ash; * US fly ash, ◆ Normal fly ash + HNO_3 , ■ US fly ash + HNO_3 , ● Normal fly ash + NaOH , ▲ US fly ash + NaOH .
Source: Authors

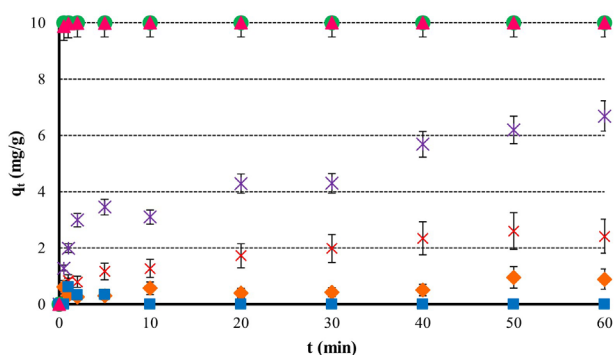


Figure 4. Experimental results of Cd adsorption at pH 7. × Normal fly ash; × US fly ash, × Normal fly ash + HNO_3 , × US fly ash + HNO_3 , × Normal fly ash + NaOH, × US fly ash + NaOH.

Source: Authors

According to the experimental results, the treatment with HNO_3 significantly reduces the adsorption capacity of Cd(II) in fly ash when compared to normal fly ash and that activated solely with ultrasound. In the experiments conducted at pH 7, adsorption was almost non-existent. Previous studies indicate that the acid treatment of fly ash increases its porosity and surface area, making it more effective in adsorbing contaminants. This treatment also promotes dealumination and increases the concentration of silanol groups on the surface, creating more acidic sites [47], [48]. However, these changes make it difficult for Cd(II) to be adsorbed onto the surface of the fly ash.

On the other hand, it was observed that fly ash treated with NaOH can adsorb almost all the Cd(II) present in an aqueous solution. According to the material balance, if the 5 g of adsorbent material can adsorb all the Cd(II) present in the 100 mL of solution, the load is 10 mg g^{-1} . The method used to treat the fly ash, i.e., mechanical stirring or ultrasonic energy, did not affect the adsorption capacity of Cd(II). Both methods showed a practically immediate adsorption of Cd(II). The behavior of both materials was also similar at pH values of 6 and 7; all the Cd(II) was adsorbed at the beginning of the experiment. Previous studies have reported that the alkaline activation of fly ash can lead to the formation of NaAlSiO_2 polymorphs that can quickly transform into zeolite structures. Furthermore, when combined with ultrasonic energy, zeolite A is produced, with improved crystallinity and a smaller particle size. This technology can effectively treat contaminated wastewater [49], [50].

Reuse of adsorbent materials

The materials activated in an alkaline medium were used in four adsorption cycles at pH 6. The results obtained are shown in Figs. 5 and 6. The materials prepared with mechanical stirring showed a gradual loss of adsorption after each cycle. The maximum adsorption of 10 mg g^{-1} was achieved in less than 1 min during cycle 1. However, during cycle 2, it took 60 min to reach the same level of adsorption. Similarly, in cycle 3, the adsorption achieved was 5 mg g^{-1} within 60 min; in cycle 4, it was only 2 mg g^{-1} .

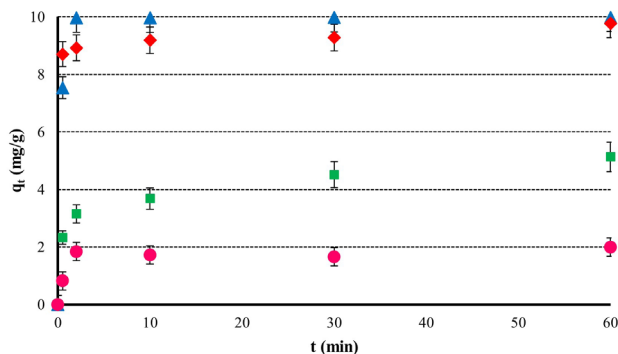


Figure 5. Reuse cycles of activated fly ash in alkaline medium with mechanical stirring at pH 6. ▲ cycle 1, ◆ cycle 2, ■ cycle 3, ● cycle 4.

Source: Authors

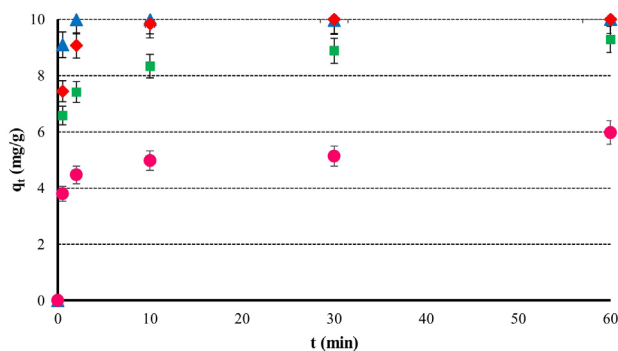


Figure 6. Reuse cycles of fly ash activated in alkaline medium with ultrasonic energy at pH 6. ▲ cycle 1, ◆ cycle 2, ■ cycle 3, ● cycle 4.

Source: Authors

When fly ash is activated in an alkaline medium using ultrasonic energy, it shows a greater adsorption capacity in each of the cycles. The maximum adsorption capacity of 10 mg g^{-1} was achieved in less than 10 min in cycles 1 and 2. Additionally, the adsorption achieved in cycle 3 was 9.4 mg g^{-1} in 60 min. In cycle 4, it was 6 mg g^{-1} , which is significantly greater than the value achieved with fly ash activated via mechanical agitation.

Kinetic studies of Cd(II) adsorption on fly ash

Experimental data were analyzed to determine the adsorption kinetics of Cd(II) on normal fly ash and ultrasound-activated fly ash. The kinetic study was not carried out on fly ash activated in an alkaline medium because the adsorption process was nearly instantaneous in the initial cycles. Table I presents the selected kinetic models already integrated and linearized. The first-order Lagergren model assumes that the limiting stage of the adsorption process is the mass transfer of Cd(II) ions from the solution to the adsorbent's surface, where there are homogeneous active sites. Therefore, the adsorption is uniform and involves physical forces. The second-order Lagergren model explains that adsorption occurs at specific sites on the adsorbent surface, and that the probability of occupying these sites is proportional to the square of the fraction of unoccupied sites. This model also assumes that the intraparticle diffusion of Cd(II) through the adsorbent material's pores is the limiting stage of the adsorption process. On the other hand, the Elovich model

is derived from the Langmuir theory, but it differs in that the active sites of the adsorbent are heterogeneous and exhibit different activation energies. This means that the rate of adsorption of a solute molecule at a given active site depends on the activation energy of that site.

Table I. Kinetics models proposed for the heavy metal sorption on solid materials

Name of the model	Integrated model
First-order Lagergren	$q_t = q_e \cdot \left[1 - e^{(-k \cdot t)} \right]$
Second-order Lagergren	$q_t = \frac{t}{\frac{1}{k \cdot q_e^2} + \frac{t}{q_e}}$
Elovich	$q_t = k_1 + k_2 \cdot \ln(t)$

Source: Authors

The experimental data were analyzed using the PolymathPlus software, and nonlinear regression was performed for each proposed model. The R^2 adjusted criterion for the sorption of Cd(II) on coal fly ash materials is presented in Table II. The results show that the Elovich adsorption isotherm is the kinetic model that best fits the experimental data. The fittings of the Elovich isotherm with the experimental data shown in Figs. 7 and 8 provide compelling evidence of their strong correlation.

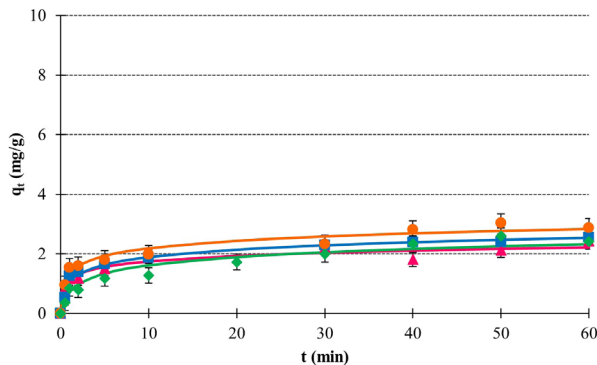


Figure 7. Fitting experimental data for Cd adsorption on normal fly ash with the Elovich isotherm. ▲ pH 4, ■ pH 5, ● pH 6, ◆ pH 7.

Source: Authors

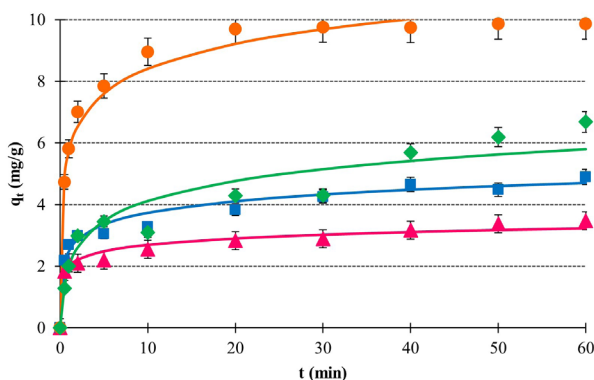


Figure 8. Fitting experimental data for Cd adsorption on fly ash activated with ultrasonic energy with the Elovich isotherm. ▲ pH 4, ■ pH 5, ● pH 6, ◆ pH 7.

Source: Authors

Table II. R^2 adjusted criterion for the sorption of Cd(II) on coal fly ash materials

Model	R ² _{adjusted}								Average
	Normal fly ash				Activated fly ash				
	pH = 4	pH = 5	pH = 6	pH = 7	pH = 4	pH = 5	pH = 6	pH = 7	
First-order Lagergren	0.8432	0.9014	0.7990	0.8668	0.7682	0.7700	0.9364	0.8668	0.8440
Second-order Lagergren	0.9146	0.9602	0.8891	0.9146	0.8760	0.8563	0.9853	0.8371	0.9042
Elovich	0.9456	0.9569	0.9604	0.9359	0.9694	0.9555	0.9806	0.9132	0.9522

Source: Authors

Discussion

Effect of pH on adsorption capacity

The pH of the solution affects the surface charge of the adsorbent, which in turn influences the adsorption capacity. At lower pH values, the adsorbent surface is more positively charged, leading to a reduced adsorption of positively charged Cd(II) ions due to electrostatic repulsion. However, as the pH increases, the surface charge becomes increasingly negative, leading to a more significant electrostatic attraction of Cd(II) ions [51], [52]. The adsorption capacity of the fly ash is enhanced as the pH increases from 4 to 6 and decreases at pH 7 (Figs. 1 and 2). At low pH levels, the active sites of fly ash are subjected to an intense competition from H^+ and Cd(II) ions for adsorption. In a similar manner, the surface of fly ash can receive a partially positive charge, causing it to repel Cd(II) ions. The latter is due to the presence of SiO_2 and Al_2O_3 in the fly ash structure, whose charge depends on the pH of the medium. The number of H^+ ions decreases as the pH of wastewater increases, resulting in less competition for the active sites of the fly ash with Cd(II) ions. Similarly, with the increase in pH, the surface charge of the fly ash tends to be negative, which attracts Cd(II) ions [17], [18], [20], [21], [51]. On the other hand, at greater pH levels (>6) there is the possibility of forming $Cd(OH)_2$, which may precipitate and negatively affect its adsorption by blocking active sites [37], [51].

Activation of fly ash in an acidic medium

The activation of fly ash in an acidic medium has been used to improve different physical, chemical, and textural properties of fly ash [53]. The acidic medium helps to remove impurities, such as unburned carbon residues, sulfur, and volatile compounds. It also increases the surface area by dissolving metal oxides and increasing the pore size. However, in this work, the acid activation of fly ash was ineffective in improving the Cd(II) adsorption capacity, regardless of whether ultrasonic energy was applied. The low adsorption capacity is similar to that reported in previous studies [37]. This may be due to the acidic character that fly

ash acquires after acid treatment, making it unattractive to Cd(II) ions, which, according to these studies, behave like soft Lewis acids.

Activation of fly ash in an alkaline medium and its reuse

The activation of fly ash with NaOH, a process that induces structural changes on its surface by destroying its glass phase structure and increasing its surface area, has significant practical implications. The chemical interactions between NaOH and the active species of Al_2O_3 and SiO_2 make them more susceptible to increasing their adsorption capacity. For instance, once activation has been carried out, the Al sites tend to have a negative charge and quickly attract positive ions [18], [40]. This has led to a significant improvement in the adsorption capacity of fly ash, particularly regarding Cd(II) ions, and it is the main reason for the high adsorption efficiency of the materials synthesized in this work. Said efficiency is maintained for up to four reuse cycles, a situation that has also been noted in other studies [19], [54], [55].

Activation of fly ash using ultrasonic energy

On the other hand, it was demonstrated that using ultrasonic energy on fly ash significantly enhances its ability to adsorb chemical compounds, even when it has been treated in an acidic solution – where its effectiveness in removing Cd(II) is minimal. Several studies have demonstrated that activating fly ash with ultrasonic energy effectively removes contaminants from wastewater. These studies consistently show a significant increase in the adsorption capacity of fly ash after ultrasound treatment, regardless of the specific contaminant (e.g., heavy metals, dyes, and toxic organic chemicals), the pH, the temperature, and the amount of adsorbent used in the wastewater [43], [56], [57]. The improvement in the characteristics of fly ash can be ascribed to the acoustic cavitation produced when ultrasonic energy is applied. Ultrasonic waves generate gas micro-bubbles in the liquid, which rapidly expand and burst in response to varying sound pressure. During the collapse of the microbubbles, there are microjets and extremely high local pressure and temperature conditions (exceeding 1000 atm and 5000 °C). These conditions lead to significant changes in the structure and morphology of the fly ash, causing it to break down into smaller particles. This results in an increase in surface area, pore size, and the availability of active sites. Consequently, the external resistance to mass transfer during the adsorption process is reduced. Additionally, sonic cavitation can aid in the elimination of contaminants within the material which hinder the organization of active sites [57], [58]. Additionally, the use of ultrasonic energy can significantly enhance the reactions between fly ash and NaOH in an alkaline environment. At specific points where high pressure and temperature conditions are reached, a synergistic effect is triggered. This effect is a result of the combination of alkaline activation and ultrasound, which plays a crucial role in increasing

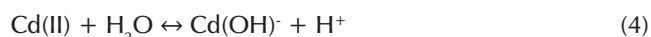
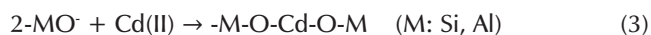
the surface area and forming fused Na-fly ash species with negative charge. Consequently, the adsorption capacity is significantly enhanced.

Kinetics studies of Cd(II) adsorption on fly ash

When carrying out kinetic studies of Cd(II) adsorption on fly ash, the Elovich model exhibited a better statistical fit with the experimental data than the pseudo-first order and pseudo-second order models. Elovich's theory demonstrates the heterogeneity of the active sites in fly ash. Some these sites may be derived from Al and others from Si, and each of them can have different ionic strengths, thus exhibiting different activation energies. Other studies have fitted the experimental data to the pseudo-second order model [17], [19], [37], [51], [54], [59]. However, they did not raise the possibility of fitting the data to the Elovich model. It is important to note that the pseudo-second-order model also considers that adsorption occurs at specific sites on the surface of the adsorbent, but it does not include the variation in these sites' activation energy.

This is in line with the characterization results reported by some authors [17]–[20], [54]. FTIR characterization identified that Cd(II) ions adsorb by ion exchange with the -OH, Si-O, and Al-O species, forming Si-O-Cd species – this was confirmed through XRD spectra. However, they did not relate these findings to the Elovich model, which considers that adsorption takes place in various heterogeneous sites with different activation energies, and adjustments were made mainly with the Lagergren first- and second-order models.

Based on the above-presented results, the following reaction mechanism was proposed by [20], where the presence of more than one adsorbent species is observed, again in accordance with the Elovich model, with multiple active site with different activation energies:



Conclusions

This study shows that coal fly ash, a cost-effective material, can remove cadmium from industrial wastewater. However, the activation of coal fly ash in an acidic medium is ineffective; after this chemical treatment, the material did not adsorb cadmium. Activating coal fly ash by applying ultrasonic energy in an alkaline medium enhances its adsorption capacity. pH plays a critical role in adsorption, with the highest cadmium uptake occurring at pH 6. Statistical analysis indicates that the Elovich isotherm provides a better fit for the experimental data on cadmium adsorption across all studied materials and under various experimental conditions.

CRediT author statement

Roberto Flores conceived the idea, developed the methodology, and supervised the experimental work. Yareli G. Medina conducted the experimental work, collected the experimental data, and validated the results. Both authors collected bibliographic information, formally analyzed the experimental results, and provided critical feedback. Roberto Flores contributed to writing the definitive version of the manuscript.

Conflicts of interest

The authors declare that they have no conflict of interest that could influence the contents of this manuscript.

References





- [1] T. El Rasafi *et al.*, "Cadmium stress in plants: A critical review of the effects, mechanisms, and tolerance strategies," *Crit. Rev. Environ. Sci. Technol.*, vol. 52, no. 5, pp. 675–726, Mar. 2022. <https://doi.org/10.1080/10643389.2020.1835435>
- [2] Y. Huang, S. Mubeen, Z. Yang, and J. Wang, "Cadmium contamination in agricultural soils and crops," in *Theories and Methods for Minimizing Cadmium Pollution in Crops*, Z. Yang, C. He, and J. Xin, Eds., Singapore: Springer Nature Singapore, 2022, pp. 1–30. https://doi.org/10.1007/978-981-16-7751-9_1
- [3] U. Zulfiqar *et al.*, "Cadmium phytotoxicity, tolerance, and advanced remediation approaches in agricultural soils; A comprehensive review," *Front. Plant Sci.*, vol. 13, art. 773815, Mar. 2022. <https://doi.org/10.3389/fpls.2022.773815>
- [4] M. Tunçtürk *et al.*, "Safflower (*Carthamus tinctorius* L.) response to cadmium stress: Morpho-physiological traits and mineral concentrations," *Life*, vol. 13, no. 1, art. 135, Jan. 2023. <https://doi.org/10.3390/life13010135>
- [5] W. Fan *et al.*, "The rhizosphere microbiome improves the adaptive capabilities of plants under high soil cadmium conditions," *Front. Plant Sci.*, vol. 13, art. 914103, Oct. 2022. <https://doi.org/10.3389/fpls.2022.914103>
- [6] Y. Guo *et al.*, "Copper and cadmium co-contamination affects soil bacterial taxonomic and functional attributes in paddy soils," *Environ. Pollut.*, vol. 329, art. 121724, Jul. 2023. <https://doi.org/10.1016/j.envpol.2023.121724>
- [7] Q. Meng, T. Diao, L. Yan, and Y. Sun, "Effects of single and combined contamination of microplastics and cadmium on soil organic carbon and microbial community structural: A comparison with different types of soil," *Appl. Soil Ecol.*, vol. 183, art. 104763, Mar. 2023. <https://doi.org/10.1016/j.apsoil.2022.104763>
- [8] N. Verma, R. Gill, K. Priya, and A. Kumar, "Plants and microbes assisted remediation of cadmium-contaminated soil," in *Hazardous and Trace Materials in Soil and Plants*, M. Naeem, T. Aftab, A. A. Ansari, S. Singh Gill, and A. Macovei, Eds. Cambridge, MA, USA: Academic Press, 2022, pp. 283–296. <https://doi.org/10.1016/B978-0-323-91632-5.00008-2>
- [9] G. Genchi, M. S. Sinicropi, G. Lauria, A. Carocci, and A. Catalano, "The effects of cadmium toxicity," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, art. 3782, May 2020. <https://doi.org/10.3390/ijerph17113782>
- [10] I. Suhani, S. Sahab, V. Srivastava, and R. P. Singh, "Impact of cadmium pollution on food safety and human health," *Curr. Opin. Toxicol.*, vol. 27, pp. 1–7, Sep. 2021. <https://doi.org/10.1016/j.cotox.2021.04.004>
- [11] K. Bhattacharyya *et al.*, "Pathophysiological effects of cadmium(II) on human health-a critical review," *J. Basic Clin. Physiol. Pharmacol.*, vol. 34, no. 3, pp. 249–261, May 2023. <https://doi.org/10.1515/jbcpp-2021-0173>
- [12] H. Thu Ha, T. Dinh Minh, and H. Minh Nguyet, "Application of green nanocomposite to adsorb cadmium ion in wastewater," *VNU J. Sci. Earth Environ. Sci.*, vol. 37, no. 1, Mar. 2021. <https://doi.org/10.25073/2588-1094/vnuees.4564>
- [13] Y. Deng *et al.*, "Purification and water resource circulation utilization of Cd-containing wastewater during microbial remediation of Cd-polluted soil," *Environ. Res.*, vol. 219, art. 115036, Feb. 2023. <https://doi.org/10.1016/j.envres.2022.115036>
- [14] M. Irfan, X. Liu, K. Hussain, S. Mushtaq, J. Cabrera, and P. Zhang, "The global research trend on cadmium in freshwater: A bibliometric review," *Environ. Sci. Pollut. Res.*, vol. 30, no. 28, pp. 71585–71598, Apr. 2021. <https://doi.org/10.1007/s11356-021-13894-7>
- [15] M. A. Irshad *et al.*, "Green and eco-friendly synthesis of TiO₂ nanoparticles and their application for removal of cadmium from wastewater: reaction kinetics study," *Z. Für Phys. Chem.*, vol. 236, no. 5, pp. 637–657, May 2022. <https://doi.org/10.1515/zpch-2021-3171>
- [16] Z. Yuan *et al.*, "Tracing anthropogenic cadmium emissions: From sources to pollution," *Sci. Total Environ.*, vol. 676, pp. 87–96, Aug. 2019. <https://doi.org/10.1016/j.scitotenv.2019.04.250>
- [17] X. Zhao *et al.*, "Effect and mechanisms of synthesis conditions on the cadmium adsorption capacity of modified fly ash," *Ecotoxicol. Environ. Saf.*, vol. 223, art. 112550, Oct. 2021. <https://doi.org/10.1016/j.ecoenv.2021.112550>
- [18] H. Zhao *et al.*, "Possibility of removing cadmium pollution from the environment using a newly synthesized material coal fly ash," *Environ. Sci. Pollut. Res.*, vol. 27, no. 5, pp. 4997–5008, Feb. 2020. <https://doi.org/10.1007/s11356-019-07163-x>
- [19] X. Huang *et al.*, "Optimization of preparation technology for modified coal fly ash and its adsorption properties for Cd²⁺," *J. Hazard. Mater.*, vol. 392, art. 122461, Jun. 2020. <https://doi.org/10.1016/j.jhazmat.2020.122461>
- [20] R. Qiu, F. Cheng, and H. Huang, "Removal of Cd²⁺ from aqueous solution using hydrothermally modified circulating fluidized bed fly ash resulting from coal gangue power plant," *J. Clean. Prod.*, vol. 172, pp. 1918–1927, Jan. 2018. <https://doi.org/10.1016/j.jclepro.2017.11.236>
- [21] L. Ma *et al.*, "Removal of cadmium from aqueous solutions using industrial coal fly ash-nZVI," *R. Soc. Open Sci.*, vol.

- 5, no. 2, art. 171051, Feb. 2018. <https://doi.org/10.1098/rsos.171051>
- [22] B. Devi and H. P. Sarma, "Equilibrium isotherm and kinetic study of biosorption of cadmium from synthetic water using wastes leaves of *Averrhoa carambola*," *J. Appl. Nat. Sci.*, vol. 15, no. 2, pp. 826–843, Jun. 2023. <https://doi.org/10.31018/jans.v15i2.4598>
- [23] S. Gupta, D. Garg, and A. Kumar, "Cadmium biosorption using *Aloe. barbadensis* Miller leaves waste powder treated with sodium bicarbonate," *Clean. Waste Syst.*, vol. 3, art. 100032, Dec. 2022. <https://doi.org/10.1016/j.clwas.2022.100032>
- [24] A. Rezaee and S. Ahmady-Asbchin, "Removal of toxic metal Cd (II) by *Serratia bozhouensis* CdIW2 using in moving bed biofilm reactor (MBBR)," *J. Environ. Manage.*, vol. 344, art. 118361, Oct. 2023. <https://doi.org/10.1016/j.jenvman.2023.118361>
- [25] Y. Kim, K. Kim, H. H. Eom, X. Su, and J. W. Lee, "Electrochemically-assisted removal of cadmium ions by redox active Cu-based metal-organic framework," *Chem. Eng. J.*, vol. 421, art. 129765, Oct. 2021. <https://doi.org/10.1016/j.cej.2021.129765>
- [26] R. Rajumon, S. P. Aravind, S. Bhuvaneshwari, J. Ranjitha, and P. Mohanraj, "Removal of cadmium heavy metal ions from wastewater by electrosorption using modified activated carbon felt electrodes," *Water Sci. Technol.*, vol. 82, no. 7, pp. 1430–1444, Oct. 2020. <https://doi.org/10.2166/wst.2020.425>
- [27] M. A. Salem and N. Majeed, "Removal of cadmium from industrial wastewater using electrocoagulation process," *J. Eng.*, vol. 26, no. 1, pp. 24–34, Dec. 2019. <https://doi.org/10.31026/j.eng.2020.01.03>
- [28] M. M. Abou-Mesalam, M. R. Abass, E. S. Zakaria, and A. M. Hassan, "Metal doping silicates as inorganic ion exchange materials for environmental remediation," *Silicon*, vol. 14, no. 13, pp. 7961–7969, Aug. 2022. <https://doi.org/10.1007/s12633-021-01568-5>
- [29] Z. Qing, L. Guijian, P. Shuchuan, and Z. Chuncai, "The simultaneous removal of cadmium (II) and lead (II) from wastewater with the application of green synthesized magnesium silicate hydrate," *Front. Earth Sci.*, vol. 10, art. 1074687, Jan. 2023. <https://doi.org/10.3389/feart.2022.1074687>
- [30] Y. Zalm, A. Benayada, and Z. El Ahmadi, "Cadmium removal from cadmium-containing apatites by ion-exchange reactions," *ChemistrySelect*, vol. 7, no. 35, art. e202201862, Sep. 2022. <https://doi.org/10.1002/slct.202201862>
- [31] V. Ponia, H. Kumar, S. Ishtiyag, R. Goswami, and S. Bhaduria, "Remediation of cadmium ions from the contaminated aqueous media by using dead biomass of *Bacillus subtilis*," *Int. J. Adv. Res.*, vol. 10, no. 05, pp. 344–359, May 2022. <https://doi.org/10.21474/IJAR01/14715>
- [32] P. Chen, J. Wu, L. Li, Y. Yang, and J. Cao, "Modified fly ash as an effect adsorbent for simultaneous removal of heavy metal cations and anions in wastewater," *Appl. Surf. Sci.*, vol. 624, art. 157165, Jul. 2023. <https://doi.org/10.1016/j.apsusc.2023.157165>
- [33] A.-E. Segneanu et al., "Highly efficient engineered waste eggshell-fly ash for cadmium removal from aqueous solution," *Sci. Rep.*, vol. 12, no. 1, art. 9676, Jun. 2022. <https://doi.org/10.1038/s41598-022-13664-6>
- [34] K. Singh, A. K. Singh, A. Kumar, and A. Agarwal, "Fly ash and TiO₂ modified fly ash as adsorbing materials for removal of Cd(II) and Pb(II) from aqueous solutions," *J. Hazard. Mater. Adv.*, vol. 10, art. 100256, May 2023. <https://doi.org/10.1016/j.hazadv.2023.100256>
- [35] N. B. Singh, A. Agarwal, A. De, and P. Singh, "Coal fly ash: An emerging material for water remediation," *Int. J. Coal Sci. Technol.*, vol. 9, no. 1, art. 44, Dec. 2022. <https://doi.org/10.1007/s40789-022-00512-1>
- [36] V. K. Yadav, A. Amari, A. Gacem, N. Elboughdiri, L. B. El-tayeb, and M. H. Fulekar, "Treatment of fly-ash-contaminated wastewater loaded with heavy metals by using fly-ash-synthesized iron oxide nanoparticles," *Water*, vol. 15, no. 5, art. 908, Feb. 2023. <https://doi.org/10.3390/w15050908>
- [37] G. Buema et al., "Eco-friendly materials obtained by fly ash sulphuric activation for cadmium ions removal," *Materials*, vol. 13, no. 16, art. 3584, Aug. 2020. <https://doi.org/10.3390/ma13163584>
- [38] Z. Hussain et al., "Modification of coal fly ash and its use as low-cost adsorbent for the removal of directive, acid and reactive dyes," *J. Hazard. Mater.*, vol. 422, art. 126778, Jan. 2022. <https://doi.org/10.1016/j.jhazmat.2021.126778>
- [39] A. Fernández-Jiménez, M. Monzó, M. Vicent, A. Barba, and A. Palomo, "Alkaline activation of metakaolin-fly ash mixtures: Obtain of Zeoceramics and Zeocements," *Microporous Mesoporous Mater.*, vol. 108, no. 1–3, pp. 41–49, Feb. 2008. <https://doi.org/10.1016/j.micromeso.2007.03.024>
- [40] A. Purbasari, D. Ariyanti, and S. Sumardiono, "Preparation and application of fly ash-based geopolymer for heavy metal removal," presented at the *2nd Int. Conf. Chem. Process Prod. Eng.*, Semarang, Indonesia, 2020, art. 050006. <https://doi.org/10.1063/1.5140918>
- [41] S. Boycheva, I. Marinov, S. Miteva, and D. Zgureva, "Conversion of coal fly ash into nanozeolite Na-X by applying ultrasound assisted hydrothermal and fusion-hydrothermal alkaline activation," *Sustain. Chem. Pharm.*, vol. 15, art. 100217, Mar. 2020. <https://doi.org/10.1016/j.scp.2020.100217>
- [42] R. Flores, "Kinetics studies on the process of zn removal from wastewater using ultrasonically activated sorbents," *Chem. Biochem. Eng. Q.*, vol. 31, no. 1, pp. 123–130, Apr. 2017. <https://doi.org/10.15255/CABEQ.2015.2267>
- [43] B. Ileri and D. Sanliyüksel Yucel, "Metal removal from acid mine lake using ultrasound-assisted modified fly ash at different frequencies," *Environ. Monit. Assess.*, vol. 192, no. 3, art. 185, Mar. 2020. <https://doi.org/10.1007/s10661-020-8150-4>
- [44] P. Kucharski, B. Białecka, and M. Thomas, "Removal of cadmium ions from polluted waters using low-cost adsorbents: process optimization study," *Desalination Water Treat.*, vol. 256, pp. 114–124, Apr. 2022. <https://doi.org/10.5004/dwt.2022.28375>
- [45] T. T. Nguyen, T. M. L. P. Thi, T. N. Thi, T. T. Le, C. T. N. Thi, and N. H. Nguyen, "Adsorption optimization for the removal of cadmium in water by aluminum (hydr)oxide on cation exchange resin," *Curr. Appl. Sci. Technol.*, vol. 23, no.

- 2, Jul. 2022. <https://doi.org/10.55003/cast.2022.02.23.012>
- [46] Y. Xu, H. Xia, Q. Zhang, G. Jiang, W. Cai, and W. Hu, "Adsorption of cadmium(II) in wastewater by magnesium oxide modified biochar," *Arab. J. Chem.*, vol. 15, no. 9, art. 104059, Sep. 2022. <https://doi.org/10.1016/j.arab-jc.2022.104059>
- [47] A. Chatterjee, X. Hu, and F. L.-Y. Lam, "Modified coal fly ash waste as an efficient heterogeneous catalyst for dehydration of xylose to furfural in biphasic medium," *Fuel*, vol. 239, pp. 726–736, Mar. 2019. <https://doi.org/10.1016/j.fuel.2018.10.138>
- [48] N. Wang, L. Hao, J. Chen, Q. Zhao, and H. Xu, "Adsorptive removal of organics from aqueous phase by acid-activated coal fly ash: Preparation, adsorption, and Fenton regenerative valorization of 'spent' adsorbent," *Environ. Sci. Pollut. Res.*, vol. 25, no. 13, pp. 12481–12490, May 2018. <https://doi.org/10.1007/s11356-018-1560-y>
- [49] Z. Adamczyk, M. Cempa, and B. Białecka, "The influence of ultrasound on fly ash zeolitisation process efficiency," *Miner. Process. Extr. Metall. Rev.*, vol. 43, no. 4, pp. 427–439, May 2022. <https://doi.org/10.1080/08827508.2021.1897585>
- [50] M. Echeverri-Aguirre, J. Molina, A. A. Hoyos-Montilla, H. H. Carvajal, and J. S. Rudas, "Heat flow modelling of the alkaline activation of fly ash with sodium hydroxide in the presence of portlandite," *Constr. Build. Mater.*, vol. 357, art. 129248, Nov. 2022. <https://doi.org/10.1016/j.conbuildmat.2022.129248>
- [51] F. A. Al-Khaldi, B. Abu-Sharkh, A. M. Abulkibash, and M. A. Atieh, "Cadmium removal by activated carbon, carbon nanotubes, carbon nanofibers, and carbon fly ash: a comparative study," *Desalination Water Treat.*, vol. 53, no. 5, pp. 1417–1429, Feb. 2015. <https://doi.org/10.1080/19443994.2013.847805>
- [52] K. Seffah, A. Zafour-Hadj-Ziane, A. T. Achour, J.-F. Guillet, P. Lonchambon, and E. Flahaut, "Adsorption of cadmium ions from water on double-walled carbon nanotubes/iron oxide composite," *Chem. J. Mold.*, vol. 12, no. 2, pp. 71–78, Nov. 2017. <https://doi.org/10.19261/cjm.2017.412>
- [53] S. K. Malpani and A. Rani, "A greener route for synthesis of fly ash supported heterogeneous acid catalyst," *Mater. Today Proc.*, vol. 9, pp. 551–559, 2019. <https://doi.org/10.1016/j.matpr.2018.10.375>
- [54] G. Buema et al., "Performance assessment of five adsorbents based on fly ash for removal of cadmium ions," *J. Mol. Liq.*, vol. 333, art. 115932, Jul. 2021. <https://doi.org/10.1016/j.molliq.2021.115932>
- [55] N. Mladenović Nikolić et al., "Adsorption efficiency of cadmium (ii) by different alkali-activated materials," *Gels*, vol. 10, no. 5, art. 317, May 2024. <https://doi.org/10.3390/gels10050317>
- [56] B. Çalışkan and E. Şayan, "A brief overview of the effects of ultrasound on the adsorption/desorption process: A review," *Int. J. Environ. Anal. Chem.*, vol. 104, no. 16, pp. 3821–3851, Dec. 2024. <https://doi.org/10.1080/03067319.2022.2093641>
- [57] A. Mary Ealias and M. P. Saravanakumar, "A critical review on ultrasonic-assisted dye adsorption: Mass transfer, half-life and half-capacity concentration approach with future industrial perspectives," *Crit. Rev. Environ. Sci. Technol.*, vol. 49, no. 21, pp. 1959–2015, Nov. 2019. <https://doi.org/10.1080/10643389.2019.1601488>
- [58] O. Dere Ozdemir and S. Piskin, "A novel synthesis method of zeolite x from coal fly ash: Alkaline fusion followed by ultrasonic-assisted synthesis method," *Waste Biomass Valor.*, vol. 10, no. 1, pp. 143–154, Jan. 2019. <https://doi.org/10.1007/s12649-017-0050-7>
- [59] F. A. Olabemiwo, B. S. Tawabini, F. Patel, T. A. Oyehan, M. Khaled, and T. Laoui, "Cadmium removal from contaminated water using polyelectrolyte-coated industrial waste fly ash," *Bioinorg. Chem. Appl.*, vol. 2017, pp. 1–13, 2017. <https://doi.org/10.1155/2017/7298351>

Bioelectricity Generation in Established Cacao (*Theobroma cacao*), Oil Palm (*Elaeis guineensis*), and Peruvian Amazon Grass (*Axonopus compressus*) Crops: Insights from Amazonian Soils

Generación de bioelectricidad en cultivos establecidos de cacao (*Theobroma cacao*), palma aceitera (*Elaeis guineensis*) y grama amazónica peruana (*Axonopus compressus*): perspectivas desde suelos amazónicos

Edwar E. Rubina-Arana ¹, Letty L. Sandoval-Mendoza ², Glendy Sánchez-Sunció ³, Dalia Carbonel ⁴, and Grober Panduro-Pisco ⁵

ABSTRACT

As urban populations continue to grow, the global energy demand is expected to rise accordingly. Bioelectricity generation constitutes a promising and environmentally friendly alternative for sustainable energy production. This study evaluated the energy performance of biocells with galvanized graphite (GG) and copper-aluminum (CA) electrodes, which were installed in soils cultivated with three typical Amazonian crops: cacao (*Theobroma cacao*), oil palm (*Elaeis guineensis*), and Peruvian amazon grass (*Axonopus compressus*). Voltage and current measurements were recorded twice a day over a seven-day period. According to the results, the cacao-cultivated soil with GG electrodes achieved the highest electricity generation, with a voltage of 537 mV, a current of 0.17 mA, and power density of 26.2 mW/m². In comparison, the oil palm soil with GG electrodes reached a maximum voltage of 444 mV and a power density of 10.8 mW/m². CA electrodes showed lower energy yields across all crop types, reinforcing the importance of electrode material selection. By demonstrating significant bioelectricity generation in Amazonian agro-industrial crops and ornamental grass, this research highlights the potential of agricultural soils as renewable energy sources. As the first study to assess electricity generation in established Amazonian soils, our work provides novel insights into soil-based bioelectricity, an underexplored avenue for sustainable energy production. These findings pave the way for further research on optimizing bioelectricity generation in tropical soils, offering an innovative perspective on integrating renewable energy solutions into agricultural landscapes.

Keywords: biocells, bioenergy, electrodes

RESUMEN

A medida que las poblaciones urbanas continúan creciendo, se espera que la demanda global de energía aumente proporcionalmente. La generación de bioelectricidad constituye una alternativa prometedora y respetuosa con el medio ambiente para la producción de energía sostenible. Este estudio evaluó el rendimiento energético de bioceldas con electrodos de grafito galvanizado (GG) y cobre-aluminio (CA) instaladas en suelos cultivados con tres cultivos típicos de la Amazonía: cacao (*Theobroma cacao*), palma aceitera (*Elaeis guineensis*) y pasto amazónico peruano (*Axonopus compressus*). Se realizaron mediciones de voltaje y corriente dos veces al día durante un período de siete días. Según muestran los resultados, el suelo cultivado con cacao y electrodos GG logró la mayor generación de electricidad, con un voltaje de 537 mV, una corriente de 0.17 mA y una densidad de potencia de 26.2 mW/m². En comparación, el suelo de palma aceitera con electrodos GG alcanzó un voltaje máximo de 444 mV y una densidad de potencia de 10.8 mW/m². Los electrodos CA mostraron menores rendimientos energéticos en todos los tipos de cultivo, reforzando la importancia de la selección del material del electrodo. Al demostrar una generación significativa de bioelectricidad en cultivos agroindustriales amazónicos y césped ornamental, esta investigación destaca el potencial de los suelos agrícolas como fuentes de energía renovable. Como el primer estudio en evaluar la generación de electricidad en suelos amazónicos establecidos, nuestro trabajo proporciona conocimientos novedosos sobre la bioelectricidad basada en suelos, un campo poco explorado en la producción de energía sostenible. Estos hallazgos abren el camino para futuras investigaciones orientadas a optimizar la generación de bioelectricidad en suelos tropicales, ofreciendo una perspectiva innovadora para la integración de soluciones energéticas renovables en los paisajes agrícolas.

Palabras clave: bioceldas, bioenergía, electrodos

Received: December 18th, 2023

Accepted: April 1st, 2025

¹ Environmental engineer, Universidad Nacional de Ucayali, Peru. Affiliation: specialist in economic assessment of environmental impact, National Environmental Certification Service for Sustainable Investments, Peru. Email: eera.rubina@gmail.com

² Environmental engineer, Universidad Nacional de Ucayali, Peru. MSc in Environment, Sustainable Management, and Social Responsibility, Universidad Nacional de Ucayali, Peru. Affiliation: professor, Universidad Nacional de Ucayali, Peru. Email: letty_sandoval@unu.edu.pe

³ Agricultural engineer, Universidad Nacional de Ucayali, Peru. PhD in Public Management and Governance, Universidad César Vallejo, Lima - Peru. Affiliation: professor, Universidad Nacional de Ucayali, Peru. Email: glendy_sanchez@unu.edu.pe

⁴ Forest engineer, Universidad Nacional Agraria La Molina, Peru. MSc in Environmental Engineering with a mention in Water Treatment and Waste Reuse, Universidad Nacional de Ingeniería, Peru. Affiliation: Researcher. Email: dcarbonelr@uni.pe

⁵ Agricultural engineer, Universidad Nacional de Ucayali, Peru. PhD in Environmental Sciences and Renewable Energy, Universidad Nacional de San Agustín, Peru. Affiliation: researcher professor, Universidad Nacional de Ucayali, Peru. Email: grober_panduro@unu.edu.pe



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Introduction

According to the United Nations, the global urban population was approximately 4.5 billion in 2022 and is projected to reach 6.6 billion by 2050 [1]. This increase is expected to escalate product consumption and energy demand. Currently, the primary energy sources are non-renewable, which significantly contributes to environmental pollution through greenhouse gas emissions [2]. The looming threat of climate change, the dependence on energy imports, and the finite nature of fossil fuels have raised widespread concern regarding energy security and the environmental impacts associated with energy production and consumption. Consequently, the United Nations have urged governments to implement strategies geared towards sustainable growth and development [3].

In response, numerous countries are increasing investments in renewable energy. A novel form of renewable energy that has attracted attention in recent years is bioelectricity generated by soil-dwelling microorganisms and plant root exudates. This growing interest arises from the global trend towards sustainable energy production and the capability of bioelectrochemical systems to degrade organic matter and perform bioremediation [4]. Biocells, *i.e.*, devices that utilize microorganisms to convert the chemical energy of substrates into electricity, represent an emerging technology with significant potential to mitigate the energy crisis [5]. During photosynthesis, plants produce compounds they do not fully utilize and release them through their roots. When these exudates interact with microorganisms in the rhizosphere that are capable of oxidizing carbohydrates under anoxic conditions, electrons are produced in the soil, facilitating organic matter degradation [6]. The potential of biocells is considerable, and previous studies have demonstrated their effectiveness [7], [8]. For instance, a study conducted by [9] using microbial fuel cells with ornamental plants in wetlands achieved outputs of 750 mV, and [10] investigated electric power generation in house-grown potted plants, reaching maximum voltages of 340 mV.

Reflecting on the progress of field-based bioelectricity generation research, significant advances have been made in harnessing the electrical potential of plant-microbial ecosystems. [11] demonstrated the utility of plant-microbial fuel cells in diverse agricultural settings. These findings highlight the practicality of integrating bioelectrochemical systems into existing agricultural infrastructures to simultaneously enhance crop yield and renewable energy production. [12] further explored the deployment of these systems in field conditions, identifying key factors such as plant species selection and the optimization of reactor components that significantly influence bioelectricity generation efficiency. These insights are particularly relevant for this study, as they provide a foundation for evaluating the bioelectric potential of Amazonian soils and the performance of different electrode materials in an agro-industrial context.

Carbon-based materials and metal electrodes have recently been employed in biocell systems. These materials are promising candidates due to their cost-effectiveness, high electrical conductivity, large specific surface area, significant pore volume, and stability [13]. Galvanized graphite (GG) as an anode material is particularly suitable for soil-based microbial fuel cells because of its high conductivity and compatibility with microbial life, which is crucial for maximizing energy capture from soil microbial activity. Copper and aluminum (CA) electrodes are also promising; their exceptional conductivity may further enhance bioelectric potential in microbial fuel cells. Notably, copper supports the formation of highly active biofilms essential for bioenergy production [14]. Moreover, [15] suggest that coating metal electrodes with biocompatible materials could improve their performance, an approach that is beneficial for the CA electrodes used in our study.

Although these findings are encouraging, a noticeable research gap remains. To date, only a limited number of studies has evaluated electricity generation in soils with established crops. Consequently, this study seeks to estimate the potential for electricity generation in soils cultivated with two agro-industrial crops and one ornamental grass. The goal is to quantify the energy derived from the metabolic activity of microorganisms in soils planted with cacao (*Theobroma cacao*), oil palm (*Elaeis guineensis*), and Torourco grass (*Axonopus compressus*). Additionally, we aim to assess the performance of two electrode types (GG and CA) and understand the influence of environmental and soil conditions on voltage and current production. Significantly, this research offers a novel contribution to the field of bioelectricity, as it constitutes the first investigation into electricity generation in Amazonian soils cultivated with agro-industrial crops and ornamental grass. Thus, it provides a new perspective on the renewable energy potential inherent in these under-researched soils, marking a critical step forward in this innovative field and offering valuable insights despite its relatively short duration.

Materials and methods

To provide a clear and concise overview of the research methodology, a flowchart is included below. This flowchart outlines the key phases of the study, covering the experimental design and the materials and methods used. All this, with the purpose of guaranteeing clarity and facilitating replication by other researchers (Fig. 1).

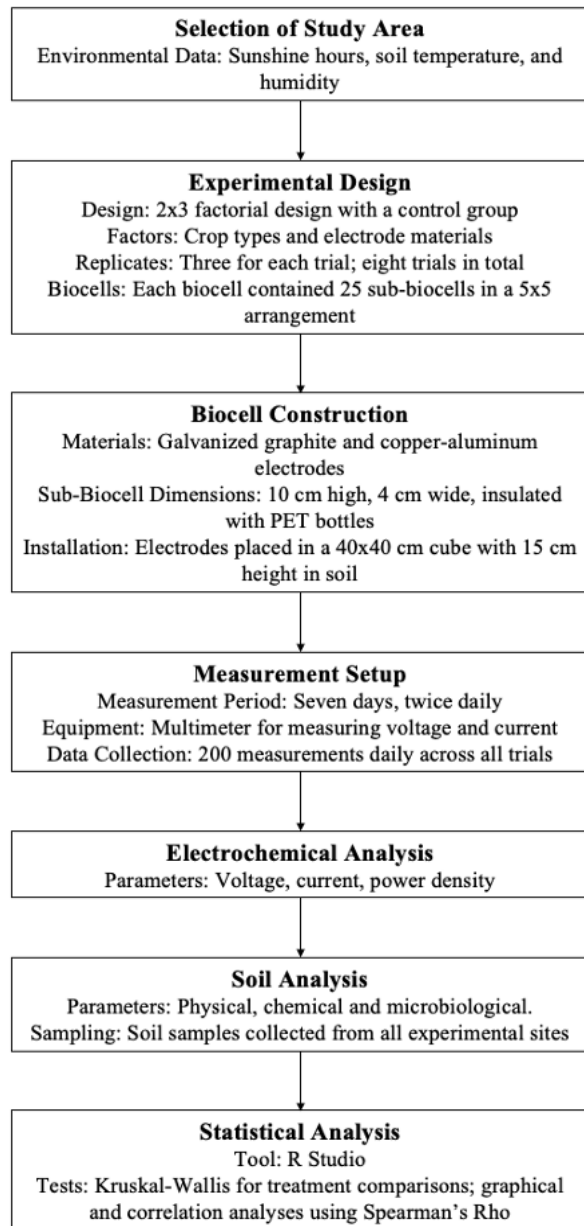


Figure 1. Flowchart of the experimental design and methodology
Source: Authors

Selection of the study area

The experiments were conducted at Universidad Nacional de Ucayali (UNU), located in the Calleria district, Coronel Portillo province, in the Ucayali region of Peru (Fig. 2). For the experiment, we selected three distinct sites within the university: the cacao botanical garden, the meteorological station where Torourco grass is cultivated, and the oil palm plantation. Additionally, a crop-less soil area was chosen to serve as the control. Data on sunshine hours, soil temperature, and environmental humidity were collected from the UNU weather station.

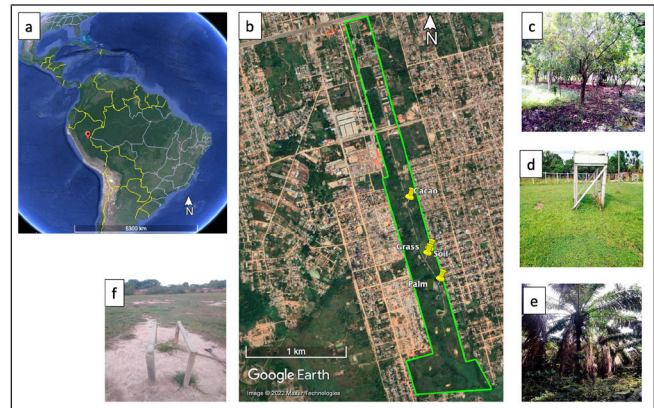


Figure 2. Location of the study area in a) America and b) the UNU campus. Specific location of biocells installed in soils with different vegetation, i.e., c) cacao, d) grass, e) oil palm, and f) non-cultivated soil
Source: Authors (images retrieved from Google Earth)

Oil palm, cacao, and grass were selected because they represent different vegetation strata – high, medium, and low elevation, respectively. Additionally, cacao and oil palm were chosen due to their agricultural significance in the study area and the extensive land area they occupy, which indicates their potential as energy sources beyond traditional agriculture. Torourco grass, widely cultivated in the study area for recreational centers, parks, and central berms, was selected due to the potential of using soil-generated electricity for public lighting purposes.

Care was taken to ensure that the experimental substrates were not affected by other experiments. Throughout the study period, the experimental sites remained free from any interference, ongoing work, or additional experiments conducted nearby. Moreover, no alternate uses of the area were permitted, thus preserving the integrity of the experimental conditions.

The experiments were intentionally conducted under natural field conditions, in order to accurately represent real-world scenarios involving bioelectricity generation systems. The frequency of voltage and current measurements, taken twice every day, was dictated by the study's objectives, specifically targeting the immediate response of the bioelectrochemical system to natural daily variations.

Despite the relatively short duration of our experiments (seven days), the standard deviation of the voltage measurements indicated stability (below 0.113 mV on most days), which was deemed sufficient for the preliminary scope of this work. The seven-day measurement period was strategically selected to balance the detailed observation of microbial consortia dynamics with logistical feasibility. Extending the duration of the study would have required significant additional resources, potentially complicating field logistics. Although a longer study could yield more comprehensive insights into external factors influencing energy generation, this preliminary research captures the immediate response under natural field conditions and sets a foundation for future extended studies.

Experimental design

A 2 x 3 factorial design, complemented with a control group, was employed to evaluate the influence of crop type and electrode material on the energy efficiency of the biocells, with a total of eight trials (Table I). The structure of the biocell is illustrated in Fig. 3. The control group consisted of uncultivated soil. As previously mentioned, the evaluated factors were crop type (Torourco grass, cacao, and oil palm) and electrode type (GG and CA). Each trial was conducted in triplicate. At each experimental site, two biocells were installed, each containing 25 sub-biocells with one electrode type, arranged in a 5 x 5 configuration.

Table I. Experiment design and coding

Electrode	Plant species			Control
	Cacao	Grass	Palm	Soil
Galvanized graphite	CGG	GGG	PGG	SGG
Copper and aluminum	CCA	GCA	PCA	SCA

Source: Authors



Figure 3. Biocell structure
Source: Authors

Biocell construction

The electron capture system was designed based on the fundamental principles of microbial galvanic cells [16]. Two types of electrodes were prepared: GG electrodes, which were constructed using galvanized steel (BCC Zincpro) (Fig. 4a) and graphite extracted from Panasonic AA batteries (Fig. 4b); and CA electrodes made from copper wire (Indeco) (Fig. 4c) and bare aluminum wires (Indeco) (Fig. 4d). After extracting the required materials, anodes and cathodes were fashioned to a length of 4.7 cm. The electrodes were soldered using 22-gauge tin wire and insulated with hot silicone to prevent oxidation and wire sulfation. Each electrode was placed within a sub-biocell structure measuring 10 cm high, 4 cm long, and 4 cm wide. A 650 mL PET water bottle, cut to a height of 10 cm, served as an insulator for each sub-biocell (Fig. 4e). Each biocell comprised 25 sub-biocells containing

either GG or CA electrodes in a 5 x 5 array (Fig. 4f). The complete experimental set up is depicted in Fig. 4.

To install the biocells, a cube measuring 40 cm per side and 15 cm deep was excavated in the soil. Some of the removed soil was poured into the PET bottles before installing the electrodes. Each biocell was covered with a felt cloth layer and topped with an additional 5 cm of soil, leaving 5 cm of wire exposed at the soil surface for voltage and current measurements (Fig. 4g). To allow for sufficient root growth around the biocells, measurements began three months after installation.

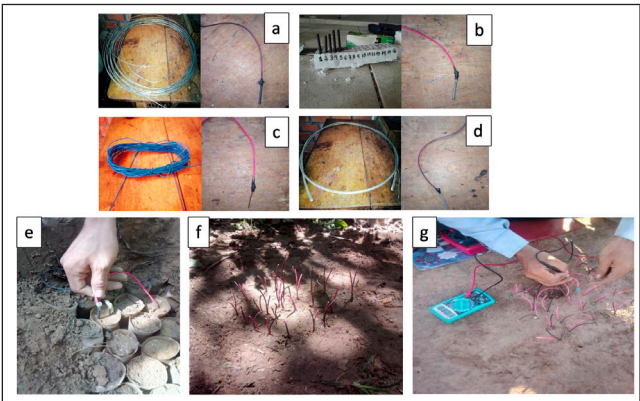


Figure 4. Experimental setup: a) galvanized steel cathode, b) AA battery graphite anode, c) copper wire anode, d) aluminum wire cathode, e) sub-biocell with anode and cathode, f) biocell composed of 25 sub-biocells, and g) voltage and current measurement setup
Source: Authors

Electrochemical analysis

Voltage and current were measured using a multimeter over seven days, twice every day, from 7 am to 12 pm and again from 2 pm to 5 pm. In each turn, 25 current and 25 voltage measurements were recorded for each of the two electrode types. Consequently, a total of 100 measurements per electrode (50 for current and 50 for voltage) were collected daily, amounting to 200 daily measurements for both electrodes.

The energy performance of each biocell was evaluated based on voltage, current, and power density. Voltage and current readings were obtained directly from the multimeter, while the power was calculated by multiplying the voltage (in mV) by the current (in mA). Subsequently, the power density (mW/ m²) was calculated by dividing the power by the biocell's surface area.

The seven-day measurement period was chosen based on the specific objectives and logistical constraints of the study. This duration was considered sufficient to capture the immediate responses and initial patterns of bioelectricity generation under the given environmental conditions. While extending the measurement period could potentially yield more comprehensive data, seven days provided a preliminary exploration sufficient to observe notable trends

within a feasible timeframe. It should also be noted that all measurements were conducted during September.

Soil analysis

Soil samples were collected from each of the four experimental sites in order to analyze their physicochemical and microbiological parameters by means of established methodologies. The physical properties assessed included sand, clay, and silt percentages, determined via the hydrometer method; as well as the wilting point, which was evaluated using the gravimetric method. The bulk density and field capacity were measured using the measuring cylinder method. The chemical analyses included electrical conductivity (measured using a conductometer in an aqueous extract), pH (determined with a potentiometer using a water-soil ratio of 1:1), and the organic matter content (assessed using the Walkley and Black method). The total nitrogen was measured using the Micro Kjeldahl method, while the available phosphorus was determined via the modified Olsen method (NHCO_3 0.5M extract at pH 8.5). The available potassium was extracted using ammonium acetate (1N) at pH 7, and cadmium levels were analyzed through Tessier sequential extraction. The cation exchange capacity was measured using ammonium acetate (1N) at pH 7, while the effective cation exchange capacity and exchangeable acids were determined via displacement with KCl 1N (for soils at $\text{pH} < 5.6$). The calcium, magnesium, potassium, and sodium concentrations were measured using an atomic absorption spectrophotometer, and the aluminum and hydrogen contents were analyzed using the Yuan method.

For the microbiological analysis, aerobic heterotrophic bacteria were quantified using the standard plate count method with nutrient agar, while anaerobic heterotrophic bacteria were counted using the most probable number (MPN) method on anaerobic agar. Actinomycetes were assessed via the spread plate method using actinomycete isolation agar, and fungi (mold and yeast) were identified through the dilution plate method on potato dextrose agar. Lactobacillus enumeration was conducted using the de Man, Rogosa, and Sharpe (MRS) agar plate count method, and the presence of mycorrhizae was evaluated through root staining and microscopic observation.

Statistical analysis

The statistical analysis was performed using R. The differences between the treatments and the control were assessed via the Kruskal-Wallis test. All treatments showed significant differences at an alpha level of 0.01. The variations in daily and hourly averages were examined through graphical representations, and the correlation between voltage and current was evaluated using Spearman's Rho.

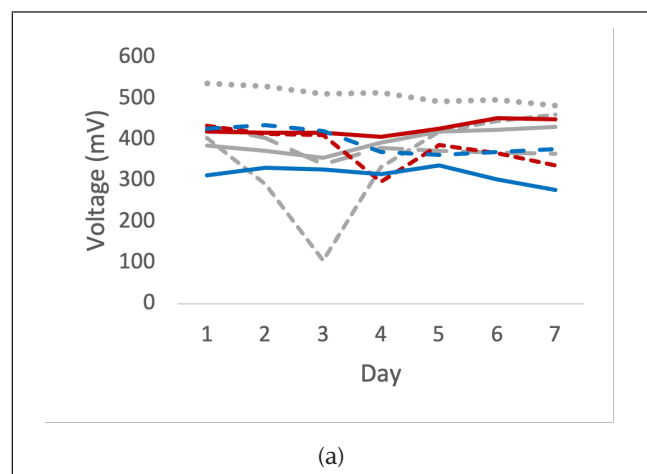
Results and discussions

Under uncontrolled field conditions, the daily variability in energy generation was analyzed and found to be stable. The data, collected twice in a day throughout the experiments, exhibited minimal fluctuations, with modest standard deviations for the voltage measurements on most days. The standard deviations ranged from 0.052 to 0.113 mV on all but one day, indicating a negligible impact of external climatic and solar variations on bioelectricity generation. These observations support the validity of the measurement approach and highlight the resilience of the bioelectric system under varying environmental conditions.

The only exception occurred on day 2, for which a higher voltage fluctuation was observed, possibly due to specific environmental events. This outlier underscores the importance of careful data interpretation and the consideration of unquantified variables, which are inherent in field-based studies. Nonetheless, the overall stability of voltage measurements throughout the experiment suggests that the chosen measurement frequency was sufficient to capture variations in bioelectricity generation under field conditions.

Efficiency of the biocells

The area of non-cultivated soil exhibits voltage emissions comparable to those from soils with crops (Figs. 5a, 5b). This could be explained by cattle frequently passing through and depositing manure, thereby promoting the growth of pioneer grasses. The roots of these grasses, along with the organic matter content from manure, could contribute to voltage generation. This hypothesis was supported by the soil characterization results (Table III), according to which the soil without cultivated crops had similar levels of organic matter, nutrients, and microorganisms, albeit slightly lower than those in the cropped soils.



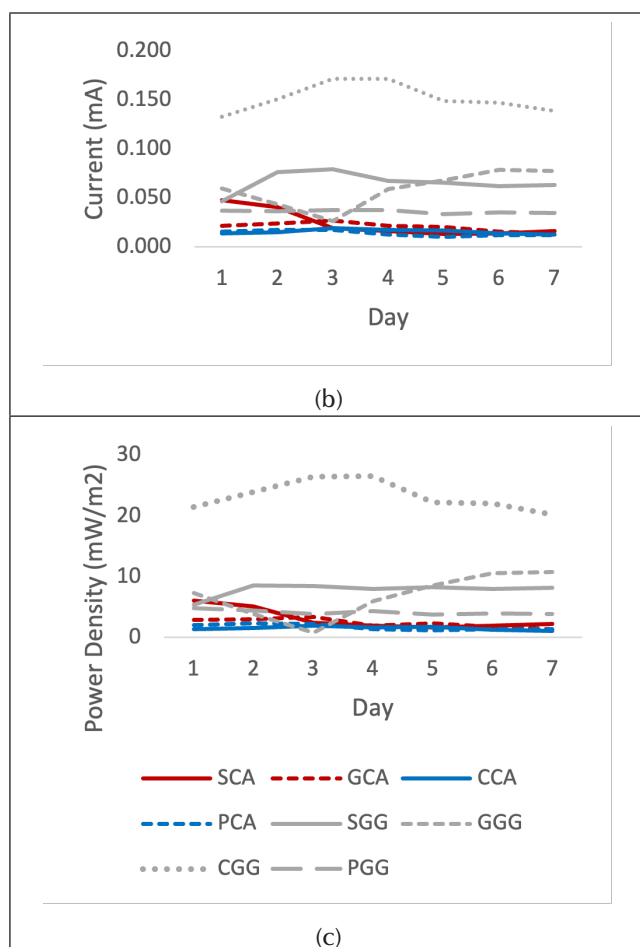


Figure 5. Performance of biocells with CA and GG in terms of a) voltage, b) current density, and c) power density
Source: Authors

Energy generation measurements began after root colonization had occurred, resulting in initial recorded voltage values of 355-434 mV. Throughout the seven-day period, no significant fluctuations in voltage, current, or power density were observed across the experiments, except for a slight dip in voltage on the third day of the GGG trial. All other trials maintained consistent voltage levels throughout the measurement period. This stability may relate to the acclimatization period before electrode installation. The generated energy is associated with the characteristics of microbial communities in the rhizosphere around the electrodes.

Generally, the microbial growth cycle comprises four phases: dormancy, exponential growth, stationary, and death. The experimental conditions and results suggest that electrical current measurements were taken during the stationary phase [17], with the dormancy and exponential growth likely occurring during the initial three-month period prior to measurement. Therefore, our measurements remained relatively stable during the seven-day experiment. A steady voltage output is associated with the maturity of the biofilms formed on electrodes [18], indicating that, by the time measurements commenced, plant roots had already

colonized the biocells and had begun releasing the enzymes necessary for electron transfer [19].

However, electricity generation is not expected to remain unchanged indefinitely. The voltage showed a slight downward trend on day 7, while current and power density peaked between days 3 and 4. In these systems, fluctuations in daily power generation measurements are common and can be attributed to microbial consumption and the subsequent depletion of the available organic matter around the electrodes [20].

Throughout the biocells' operation, both the voltage and power density values remained relatively stable but displayed distinct trends. For instance, the CGG trial achieved the highest power density: 26.2 W/m², with a maximum voltage of 537 mV. Meanwhile, the GGG trial reached a maximum voltage of 444 mV and a peak power density of 10.8 W/m². These differing outcomes can be explained by several hypotheses [21]: (i) biocells undergo complex variations under natural conditions; (ii) plant species influence microbial communities, altering the root exudate composition; and (iii) root exudates, primarily products of photosynthesis, nourish the rhizosphere microorganisms. Several studies have demonstrated that the composition and quality of microbial communities in the rhizosphere vary according to plant type, genotype, and photosynthesis strategy [22]. Although exudate production generally reflects photosynthetic activity, numerous factors influence its composition, which can vary even within different root zones of the same plant [23].

A comparison between the maximum power density values obtained in this study and those reported in previous research (Table II) indicates that plant species tested under field conditions generate electricity at levels comparable to those of laboratory studies. Furthermore, identifying electroactive microorganisms responsible for electrical activity could allow their introduction into crops in order to enhance electricity generation.

Table II. Maximum power density values in biocell experiments

Plant species	Maximum power density (mW/m ²)	Reference
<i>Chlorophytum inornatum</i> with copper plate electrodes	10	[24]
<i>Pipremnum aureum</i> with carbon fiber electrodes	15.38	[25]
<i>Agapanthus africanus</i> with carbon fiber electrodes	15.55	[26]
<i>Chlorophytum</i> with graphite electrodes	18	[21]
Rice with carbon felt anode	41.41	[27]
<i>Eichhornia crassipes</i> with graphite and zinc electrodes	100.2	[17]
<i>Theobroma cacao</i> with graphite and galvanized electrodes	26.6	This study

Source: Authors

The duration of the measurements (seven days) was selected as a practical compromise to achieve meaningful data collection within the feasibility constraints of this preliminary study. Recognizing the limitations of a shorter observation period, this timeframe was sufficient to observe initial trends and variations in bioelectricity generation under natural environmental conditions. Despite its limited duration, the study successfully identified key trends and patterns, providing foundational insights into bioelectrochemical systems in natural settings. Although extended measurement periods could offer more comprehensive insights, the selected seven-day window effectively captured the system's immediate response to environmental fluctuations, establishing groundwork for future extended studies.

Biocell performance and atmospheric parameters

The atmospheric conditions recorded over the seven-day study period are shown in Fig. 6. We monitored environmental and soil temperatures, which are considered to be among the most influential abiotic factors affecting photosynthesis [28], [29]. Despite this, Spearman's Rho correlation analysis indicated a very weak relation between electrical parameters (current and voltage) and atmospheric conditions. It is generally expected that increased temperatures enhance photosynthesis, thus promoting plant growth and rhizosphere biomass. However, the anticipated temperature-voltage relationship was not evident during our experiment, nor was there a clear association between environmental humidity and biocell efficiency. This suggests that environmental moisture might not significantly influence plant growth, and that, consequently, it has a limited impact on microbial activity within the rhizosphere. Nonetheless, it is possible that an extended observation period could reveal more pronounced correlation between these parameters [21].

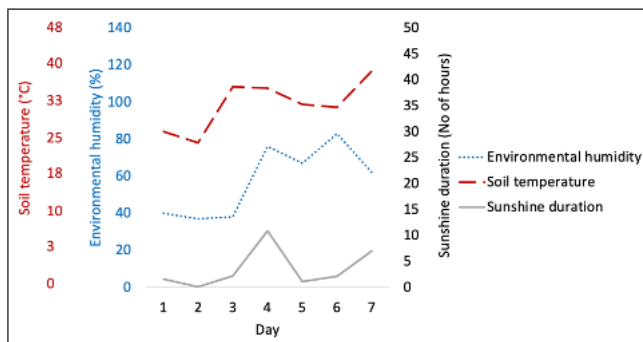


Figure 6. Atmospheric data for the seven-day study period
Source: Authors

Effect of electrode type

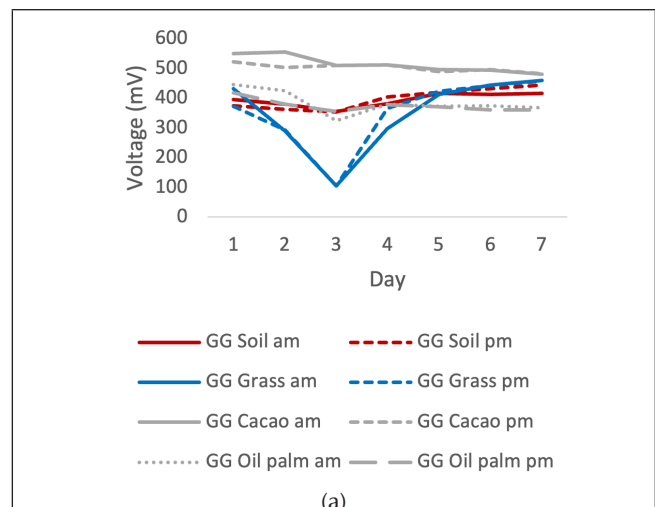
In our experiments, the GG electrodes outperformed the CA ones, particularly in cacao soils. Throughout the experiment, the CCG trial consistently achieved the highest voltage (527 mV), current (0.17 mA), and power density (26.2 mW/m²). This trial exhibited superior performance from the outset, beginning with a notable power density of 21.5 W/m², while other trials started with values ranging between 1.3 and 7.3 W/m². The

trials using CA electrodes maintained stable power densities between 1 and 3 W/m², and they even recorded lower values than the control during the first three days (Fig. 5c).

These findings underscore the significant role that electrode materials play in electricity generation. The GG electrode conducted electricity more effectively than the CA one, likely due to factors such as standard reduction potential, electrical conductivity, and resistivity. The standard reduction potential of zinc (used to coat galvanized steel) ranges between -0.76 and -1.25 V at 25 °C, whereas those of copper and aluminum are 0.52 V and -1.66 V, respectively [30]. Metals with lower standard reduction potentials are generally harder to reduce [31]. Consequently, although aluminum has a more favorable reduction potential compared to the zinc in galvanized steel, the GG electrodes demonstrated greater efficiency. Regarding electrical conductivity, zinc and copper provide values of 1.60×10^7 and 5.96×10^7 ds/M, respectively. In terms of resistivity, copper and aluminum exhibit lower values (1.72×10^{-8} and 2.74×10^{-8} Ω•m, respectively) than graphite (3.5×10^{-5} Ω•m) [32]. A lower electrical resistivity implies a lower resistance to electric current. Despite graphite's higher resistivity, the presence of graphite in the GG electrodes significantly improved energy performance, potentially even more so than the zinc coating itself.

Effect of light variation

As shown in Fig. 7, solar radiation appeared to influence the experiments. Voltage readings in the morning were slightly higher than those recorded in the afternoon. Peak light intensity hours coincide with higher temperatures, potentially facilitating the release of root exudates and enhancing the activity of soil microorganisms [21]. Nevertheless, plant-generated electricity is expected to persist into nighttime, since root exudate production from photosynthesis includes processes independent of immediate light availability. This characteristic allows biocells to function similarly to energy storage batteries [33]. Interestingly, during the GG electrode trials in this study, the afternoon's voltage generation occasionally exceeded morning values, a phenomenon previously observed in algal biocell studies [34].



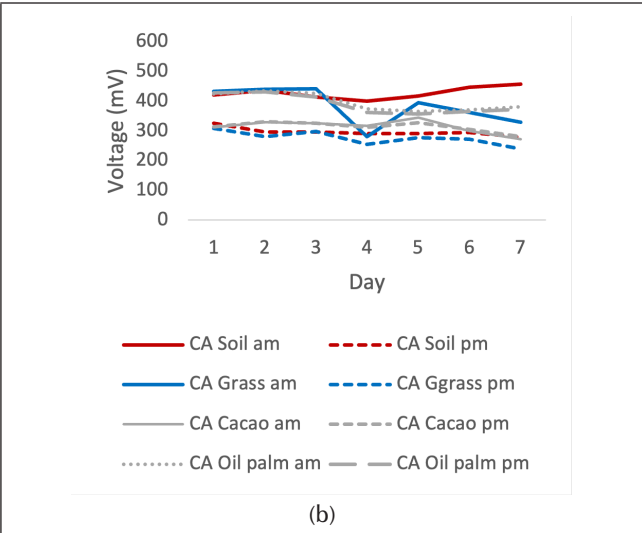


Figure 7. Variation during the seven-day experiment, illustrating the effects of morning and afternoon light phases on a) GG and b) CA electrodes

Source: Authors

Effect of soil type

Table III shows that physical soil parameters were fairly consistent across all four experimental sites. However, the chemical properties were notably superior in the soil cultivated with cacao, as evidenced by higher concentrations of organic matter, nutrients (N, P, K, Ca, Mg), and the percentage of exchangeable bases (associated with substrate fertility). The cacao-cultivated soil also had a higher concentration of heterotrophic microorganisms, fungi, and yeasts. These findings highlight a direct correlation between electricity generation and soil organic matter, nutrients, and microbial content. Specifically, the cacao crop trial with the GG electrodes produced the highest energy output (Fig. 5). It can therefore be inferred that cacao had better access to carbon sources and displayed increased microbial activity, both contributing to enhanced electricity generation. This pattern aligns with previous research studies that observed greater bioelectricity generation in soils richer in organic matter and nutrients [35]. Although this study did not explicitly measure the impact of fertilization on energy production, the observed trends suggest a hypothesis for future investigation: increased fertilization could correlate positively with enhanced bioelectricity generation.

Additionally, the robust electricity generation observed in cacao-soil biocells may be attributed to the unique interactions between the cacao crop's root system and the soil microbial community. Cacao plants, known for their dense and extensive root network [36], significantly influence the biological and chemical properties of the soil, creating favorable conditions for microbial electricity generation. As described by [37], plant roots provide substantial biomass, serving as substrates for cellulose-degrading microorganisms, which are often associated with electroactive anode-respiring bacteria. This interaction between cacao roots and soil microorganisms likely

facilitates efficient electron transfer processes, resulting in higher electricity production. Furthermore, organic matter from dead plant roots and manure, as indicated by the soil characterization results, may stimulate microbial consortia involved in bioelectricity generation, thereby enhancing the overall biocell performance in cacao-cultivated soils.

Table III. Physical, chemical and microbiological soil parameters

Parameters	Soil	Grass	Cacao	Oil palm
<i>Physical</i>				
Texture	Loam	Loam	Loam	Loam
Bulk density (g/mc ³)	1.4	1.4	1.4	1.4
Field capacity (%)	25	24	22	23
Field capacity (%)	8	7.8	7.8	7.7
Soil temperature (°C)	32	28	32	28
<i>Chemicals</i>				
Electrical conductivity (mS/cm)	0.1	0.05	0.16	0.04
pH	4.1	5.11	5.33	4.64
Organic matter (%)	0.9	0.9	1.19	0.75
Nitrogen (%)	0.04	0.04	0.05	0.03
Phosphorus (ppm)	3.4	5.99	6.39	4.4
Potassium (ppm)	51.98	72.47	88.46	55.98
Calcium Cmol(+)/kg	1.78	2.29	2.99	1.83
Magnesium Cmol(+)/kg	0.8	0.87	1.16	0.81
Aluminum Cmol(+)/kg	5.7	0.69	0.4	1.3
Hydrogen Cmol(+)/kg	0.3	0.01	0.1	0.1
Cation exchange capacity (meq/100g)	8.58	3.85	4.65	4.04
Exchangeable bases (%)	30.04	81.83	89.24	65.33
Exchangeable acids (%)	69.96	18.17	10.76	36.67
Aluminum saturation (%)	65.47	17.91	8.61	32.19
<i>Microbiological</i>				
Aerobic heterotrophic bacteria counting (CFU/g)	56x10 ³	48x10 ³	63x10 ³	8x10 ³
Anaerobic heterotrophic bacteria counting (CFU/g)	31x10 ³	25x10 ³	88x10 ³	10x10 ³
Actinomycetes counting (CFU/g)	9x10 ³	86x10 ³	88x10 ³	53x10 ³
Fungi (Moha and yeasts) counting (CFU/g)	1x10 ³	1x10 ³	5x10 ³	2x10 ³
Enum. of Lactobacillus (mo/g)	Absence	Absence	Absence	Absence
Mycorrhizae	Presence	Presence	Presence	Presence

Note: CFU: Colony-forming unit, mo: microorganisms

Source: Authors

Conclusions

Bioelectricity generation through interactions between the plant rhizosphere and soil microorganisms presents a promising pathway for renewable energy production. This study evaluated bioenergy performance in three soil types cultivated with cacao, oil palm, and Torourco grass. Biocells were installed in these soils and in a non-cultivated control, with voltage and current measurements taken over seven days, both in the morning and afternoon. Power density results demonstrated that the evaluated crops generate

electricity at levels comparable with previous studies, suggesting potential for even higher energy outputs with increased fertilization. Notably, biocells with GG electrodes consistently outperformed those with CA electrodes. An analysis of the standard reduction potential, electrical conductivity, and resistivity values indicated that graphite likely played a crucial role in enhancing electricity generation.

Interestingly, variations in atmospheric parameters and daily light intensity had minimal impact on electricity production. The most effective setup involved cacao cultivation with a GG electrode, achieving peak voltage, current, and power density values of 537 mV, 0.17 mA, and 26.6 mW/m², respectively. The superior energy performance observed in cacao soils coincided with their enhanced chemical and microbiological properties, including higher organic matter, nutrient content, and microbial populations.

Despite its limited duration, this study successfully identified significant trends and patterns in bioelectricity generation, validating the utility of monitoring bioelectrochemical systems under natural field conditions. These results address a critical gap in the existing research by quantifying the energy potential in soils cultivated with agro-industrial crops and ornamental grass. The demonstrated potential for electricity generation in soils with established oil palm, cacao, and grass crops suggests these agricultural soils can serve as renewable energy sources.

As the first exploration of electricity generation in Amazonian soils cultivated with agro-industrial crops and ornamental grass, this research provides novel insights into the renewable energy potential of these under-researched soils. Consequently, this study represents a significant advancement in the field of bioelectricity, laying important groundwork for future research and sustainable energy applications while highlighting the potential of soil-based bioelectricity as an environmentally friendly energy source that can contribute to a sustainable future.

CRediT author statement

Edwar Edinson Rubina-Arana: conceptualization, methodology, writing – review & editing (original draft), supervision, funding acquisition. *Letty Leonor Sandoval-Mendoza* and *Glendy Sánchez-Sunci6n*: formal analysis, investigation. *Dalia Carbonel*: formal analysis, writing – review & editing (original draft), visualization. *Grober Panduro-Pisco*: investigation, supervision, resources and validation.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

References








- [1] UNCTAD, "Handbook of statistics 2023," 2024. Accessed: Feb. 16, 2025. [Online]. Available: <https://unctad.org/publication/handbook-statistics-2023>
- [2] T. K. Ghose, M. R. Islam, K. Aruga, A. Jannat, and Md. M. Islam, "Disaggregated impact of non-renewable energy consumption on the environmental sustainability of the United States: A novel dynamic ARDL approach," *Sustainability*, vol. 16, no. 19, art. 8434, Sep. 2024. <https://doi.org/10.3390/su16198434>
- [3] M. Á. Caraballo Pou and M. J. García Sim6n, "Energías renovables y desarrollo económico. Un análisis para España y las grandes economías europeas," *Trimest. Econ.*, vol. 84, no. 335, art. 571, Jul. 2017. <https://doi.org/10.20430/ete.v84i335.508>
- [4] A. T. Hoang *et al.*, "Microbial fuel cells for bioelectricity production from waste as sustainable prospect of future energy sector," *Chemosphere*, vol. 287, art. 132285, Jan. 2022. <https://doi.org/10.1016/j.chemosphere.2021.132285>
- [5] P. Jalili, A. Ala, P. Nazari, B. Jalili, and D. D. Ganji, "A comprehensive review of microbial fuel cells considering materials, methods, structures, and microorganisms," *Heliyon*, vol. 10, no. 3, art. e25439, Feb. 2024. <https://doi.org/10.1016/j.heliyon.2024.e25439>
- [6] K. Obileke, H. Onyeaka, E. L. Meyer, and N. Nwokolo, "Microbial fuel cells, a renewable energy technology for bio-electricity generation: A mini-review," *Electrochem. Commun.*, vol. 125, art. 107003, Apr. 2021. <https://doi.org/10.1016/j.elecom.2021.107003>
- [7] Z. Du, H. Li, and T. Gu, "A state of the art review on microbial fuel cells: A promising technology for wastewater treatment and bioenergy," *Biotechnol. Adv.*, vol. 25, no. 5, pp. 464–482, Sep. 2007. <https://doi.org/10.1016/j.biotechadv.2007.05.004>
- [8] D. Pant, G. van Bogaert, L. Diels, and K. Vanbroekhoven, "A review of the substrates used in microbial fuel cells (MFCs) for sustainable energy production," *Bioresour. Technol.*, vol. 101, no. 6, pp. 1533–1543, Mar. 2010. <https://doi.org/10.1016/j.biortech.2009.10.017>
- [9] H. R. González-Moreno *et al.*, "Bioelectricity generation and production of ornamental plants in vertical partially saturated constructed wetlands," *Water (Basel)*, vol. 13, no. 2, art. 143, Jan. 2021. <https://doi.org/10.3390/w13020143>
- [10] K. R. S. Pamintuan, M. A. L. Calma, K. A. D. Feliciano, and K. J. P. D. Lariba, "Potential of bioelectricity generation in plant-microbial fuel cells growing house plants," *IOP Conf. Ser. Earth. Environ. Sci.*, vol. 505, no. 1, art. 012043, Jul. 2020. <https://doi.org/10.1088/1755-1315/505/1/012043>

- [11] T. Kuleshova *et al.*, "Plant microbial fuel cells as an innovative, versatile agro-technology for green energy generation combined with wastewater treatment and food production," *Biomass Bioener.*, vol. 167, art. 106629, Dec. 2022. <https://doi.org/10.1016/j.biombioe.2022.106629>
- [12] F. T. Kabutey *et al.*, "An overview of plant microbial fuel cells (PMFCs): Configurations and applications," *Renew. Sust. Ener. Rev.*, vol. 110, pp. 402–414, Aug. 2019. <https://doi.org/10.1016/j.rser.2019.05.016>
- [13] R. Agüero-Quinones *et al.*, "Activated carbon electrodes for bioenergy production in microbial fuel cells using synthetic wastewater as substrate," *Sustainability*, vol. 15, no. 18, art. 13767, Sep. 2023. <https://doi.org/10.3390/su151813767>
- [14] M. M. Haque, M. K. Mosharaf, M. A. Haque, M. Z. H. Tanvir, and M. K. Alam, "Biofilm formation, production of matrix compounds and biosorption of copper, nickel and lead by different bacterial strains," *Front. Microbiol.*, vol. 12, art. 615113, Jun. 2021. <https://doi.org/10.3389/fmicb.2021.615113>
- [15] J. V. Boas, V. B. Oliveira, M. Simões, and A. M. F. R. Pinto, "Review on microbial fuel cells applications, developments and costs," *J. Environ. Manage.*, vol. 307, art. 114525, Apr. 2022. <https://doi.org/10.1016/j.jenvman.2022.114525>
- [16] F. J. Rodríguez Varela, O. Solorza Fera, and Z. Hernández, *Celdas de Combustible*. Canada: Y1d books, 2010. <https://es.scribd.com/document/250218936/Celdas-Combustibles>
- [17] I. D. Widharyanti, M. A. Hendrawan, and M. Christwardana, "Membraneless plant microbial fuel cell using water hyacinth (*Eichhornia crassipes*) for green energy generation and biomass production," *Int. J. Renew. Ener. Dev.*, vol. 10, no. 1, pp. 71–78, Feb. 2021. <https://doi.org/10.14710/ijred.2021.32403>
- [18] M. Christwardana, S. W. A. Suedy, U. Harmoko, and K. S. S. Buanawangsa, "Exploring and evaluating the relationship between *Saccharomyces cerevisiae* biofilm maturation on carbon felt anodes and microbial fuel cell performance," *J. Electrochem. Sci. Eng.*, vol. 14, no. 5, art. 653–669, Sep. 2024. <https://doi.org/10.5599/jese.2383>
- [19] B. E. Logan *et al.*, "Microbial fuel cells: Methodology and technology," *Environ. Sci. Technol.*, vol. 40, no. 17, pp. 5181–5192, Sep. 2006. <https://doi.org/10.1021/es0605016>
- [20] J. Dziegielowski, M. Mascia, B. Metcalfe, and M. Di Lorenzo, "Voltage evolution and electrochemical behaviour of soil microbial fuel cells operated in different quality soils," *Sust. Ener. Tech. Assess.*, vol. 56, art. 103071, Mar. 2023. <https://doi.org/10.1016/j.seta.2023.103071>
- [21] Y. M. Azri, I. Tou, M. Sadi, and L. Benhabyles, "Bioelectricity generation from three ornamental plants: *Chlorophytum comosum*, *Chasmanthe floribunda* and *Papyrus diffusus*," *Int. J. Green Ener.*, vol. 15, no. 4, pp. 254–263, Mar. 2018. <https://doi.org/10.1080/15435075.2018.1432487>
- [22] Y. Zhang, J. Lin, S. Chen, H. Lu, and C. Liao, "The influence of the genotype and planting density on the structure and composition of root and rhizosphere microbial communities in maize," *Microorganisms*, vol. 11, no. 10, art. 2443, Sep. 2023. <https://doi.org/10.3390/microorganisms11102443>
- [23] V. Vives-Peris, C. de Ollas, A. Gómez-Cadenas, and R. M. Pérez-Clemente, "Root exudates: from plant to rhizosphere and beyond," *Plant Cell Rep.*, vol. 39, no. 1, pp. 3–17, Jan. 2020. <https://doi.org/10.1007/s00299-019-02447-5>
- [24] K. Manohar, A. A. Shinde, and S. Supriya, "Green electricity production from living plant and microbial fuel cell," *Int. J. Adv. Res. Sci. Eng.*, vol. 6, no. 9, pp. 459–466, Sep. 2019. https://ijarse.com/images/fullpdf/1505454223_506.pdf
- [25] P. J. Sarma and K. Mohanty, "*Epipremnum aureum* and *Dracaena braunii* as indoor plants for enhanced bio-electricity generation in a plant microbial fuel cell with electrochemically modified carbon fiber brush anode," *J. Biosci. Bioeng.*, vol. 126, no. 3, pp. 404–410, Sep. 2018. <https://doi.org/10.1016/j.jbiosc.2018.03.009>
- [26] J. C. Gómora-Hernández, J. H. Serment-Guerrero, M. C. Carreño-de-León, and N. Flores-Alamo, "Voltage production in a plant-microbial fuel cell using *Agapanthus africanus*," *Rev. Mex. Ing. Quím.*, vol. 19, no. 1, pp. 227–237, Aug. 2019. <https://doi.org/10.24275/rmiq/IA542>
- [27] J. Md Khudzari, Y. Gariépy, J. Kurian, B. Tartakovsky, and G. S. V. Raghavan, "Effects of biochar anodes in rice plant microbial fuel cells on the production of bioelectricity, biomass, and methane," *Biochem. Eng. J.*, vol. 141, pp. 190–199, Jan. 2019. <https://doi.org/10.1016/j.bej.2018.10.012>
- [28] A. Sharma *et al.*, "Photosynthetic response of plants under different abiotic stresses: A review," *J. Plant Growth Regul.*, vol. 39, no. 2, pp. 509–531, Jun. 2020. <https://doi.org/10.1007/s00344-019-10018-x>
- [29] S. F. Bucher, K. Auerswald, C. Grün-Wenzel, S. I. Higgins, and C. Römermann, "Abiotic site conditions affect photosynthesis rates by changing leaf functional traits," *Basic Appl. Ecol.*, vol. 57, pp. 54–64, Dec. 2021. <https://doi.org/10.1016/j.baae.2021.09.003>
- [30] D. Linden and T. B. Reddy, "Basic concepts," in *Linden's Handbook of Batteries*, 5th ed., T. B. Reddy and D. Linden, Eds. New York, NY, USA: McGraw Hill, 2011, ch. 1.
- [31] A. Singh, A. Contractor, R. D. Kale, and V. A. Juvekar, "Underpotential electroless deposition of metals on polyaniline," *App. Phys.*, vol. 20, no. 06, pp. 1–18, 2020. <https://doi.org/10.48550/arXiv.2006.03635>
- [32] V. Bhatt, *Essentials of coordination chemistry: A simplified approach with 3D visuals*. London, UK: Academic Press, 2016.
- [33] M. M. Ghangrekar, S. Das, and S. Das, "Microbial electrochemical technologies for CO₂ sequestration," in *Biomass Biofuels Biochemicals*, A. Pandey, R. D. Tyagi, and S. Varjani, Eds., India: Elsevier, 2021, pp. 413–443. <https://doi.org/10.1016/B978-0-12-821878-5.00016-7>
- [34] A. K. Yadav, P. Srivastava, N. Kumar, R. Abbassi, and B. K. Mishra, "Constructed wetland-microbial fuel cell: an emerging integrated technology for potential industrial wastewater treatment and bio-electricity generation," in *Constructed Wetlands for Industrial Wastewater Treatment*, Chichester, UK: John Wiley & Sons, 2018, pp. 493–510. <https://doi.org/10.1002/9781119268376.ch22>
- [35] S. R. B. Arulmani *et al.*, "Sustainable bioelectricity production from *Amaranthus viridis* and *Triticum aestivum* mediated plant microbial fuel cells with efficient electrogenic bacteria selections," *Process Biochemistry*, vol. 107, pp. 27–37, Aug. 2021. <https://doi.org/10.1016/j.procbio.2021.04.015>

- [36] A. Ishida, I. Ogiwara, and S. Suzuki, "Elevated CO₂ influences the growth, root morphology, and leaf photosynthesis of cacao (*Theobroma cacao* L.) seedlings," *Agronomy*, vol. 13, no. 9, p. 2264, Aug. 2023. <https://doi.org/10.3390/agronomy13092264>
- [37] I. Rusyn, "Role of microbial community and plant species in performance of plant microbial fuel cells," *Renew. Sust. Ener. Rev.*, vol. 152, art. 111697, Dec. 2021. <https://doi.org/10.1016/j.rser.2021.111697>

Spatial and Temporal Analysis of Road Transport Emissions in Colombia

Análisis espacial y temporal de las emisiones del transporte carretero en Colombia

Julieth V. Alfonso A ¹, Dayron Bermúdez ², Sonia C. Mangones ³, Néstor Y. Rojas ⁴, Carlos Rivera ⁵, Aquiles Darghan ⁶, and Mauricio Osses ⁷

ABSTRACT

Road transport is a major contributor to climate change and the degradation of air quality, posing risks to human health. To effectively manage air quality, it is crucial to quantify emissions and understand their spatial and temporal distribution. In this study, we focused on estimating the annual exhaust emissions of NO_2 , $\text{PM}_{2.5}$, SO_2 , CO , CO_2 , and BC from road transport in Colombia during 2019. To derive emission estimates, we employed a comprehensive approach. Vehicle activity data was obtained from car sales websites, and emission factors were obtained from COPERT, using vehicle registration data and other local variables. We distributed the total annual emissions using the road network and highway traffic flows to break down the emissions into a 10×10 km resolution. Similarly, we analyzed the temporal dimension, examining hourly emissions patterns. Our findings revealed that cargo vehicles are the primary contributors to emissions, significantly impacting BC , NO_2 , and $\text{PM}_{2.5}$ levels. Meanwhile, motorcycles were identified as the main source of CO emissions. In terms of CO_2 and SO_2 , passenger cars emerged as the most substantial mobile source. Concerning spatial and temporal disaggregation, the emissions were concentrated in metropolitan areas and exhibited a daily dispersion pattern, with three distinct peak hours. This study sheds light on the emissions profile of road transport in Colombia during 2019. By identifying the main culprits and understanding their spatial and temporal dynamics, policymakers can devise targeted strategies to mitigate the adverse impacts on air quality and human health, ultimately fostering a cleaner and healthier environment for all.

Keywords: road transport, air quality, spatial and temporal disaggregation, exhaust emissions, pollution source

RESUMEN

El transporte carretero contribuye en gran medida al cambio climático y a la degradación de la calidad del aire, lo que supone riesgos para la salud humana. Para gestionar eficazmente la calidad del aire, es crucial cuantificar las emisiones y comprender su distribución espacial y temporal. En este estudio nos centramos en estimar las emisiones anuales de escape de NO_2 , $\text{PM}_{2.5}$, SO_2 , CO , CO_2 y BC provenientes del transporte por carretera en Colombia durante el 2019. Para derivar las estimaciones de emisiones, empleamos un enfoque integral. Los datos de actividad de los vehículos se obtuvieron de sitios web de ventas de automóviles, y los factores de emisión se obtuvieron de COPERT utilizando datos de registro de vehículos y otras variables locales. Distribuimos las emisiones anuales totales utilizando la red de carreteras y los flujos de tráfico por carretera para desglosar las emisiones en una resolución de 10×10 km. Del mismo modo, analizamos la dimensión temporal, examinando los patrones de emisiones por hora. Nuestros resultados revelaron que los vehículos de carga son los principales contribuyentes a las emisiones, afectando significativamente los niveles de BC , NO_2 y $\text{PM}_{2.5}$. Entretanto, las motocicletas surgieron como la principal fuente de emisiones de CO . En cuanto al CO_2 y el SO_2 , los automóviles resultaron ser la fuente móvil más importante. En cuanto a la desagregación espacial y temporal, las emisiones se concentraron en las áreas metropolitanas y presentaron un patrón de dispersión diaria, con tres picos horarios distintos. Este estudio arroja luz sobre el perfil de emisiones del transporte por carretera en Colombia durante 2019. Mediante la identificación de los principales culpables y la comprensión de su dinámica espacial y temporal, los formuladores de políticas pueden diseñar estrategias específicas para mitigar los impactos adversos sobre la calidad del aire y la salud humana, fomentando en última instancia un medio ambiente más limpio y saludable para todos.

Palabras clave: transporte carretero, calidad del aire, desagregación espacial y temporal, emisiones de escape, fuente de contaminación

Received: February 22th, 2024

Accepted: April 25th, 2025

¹ MSc in Transport Engineering, Universidad Nacional de Colombia, Colombia. Affiliation: Department of Civil and Agricultural Engineering, Universidad Nacional de Colombia, Colombia. Email: jalfonsoa@unal.edu.co

² Civil engineer, Universidad Nacional de Colombia, Colombia. Affiliation: MSc student, Department of Civil and Agricultural Engineering, Universidad Nacional de Colombia, Colombia. Email: dbermudezm@unal.edu.co

³ PhD in Engineering and Public Policy, Carnegie Mellon University, United States. Affiliation: Researcher-professor, Universidad Nacional de Colombia, Colombia. Email: scmangonesm@unal.edu.co

⁴ PhD in Fuel and Energy, University of Leeds, United Kingdom. Affiliation: Researcher-professor, Universidad Nacional de Colombia, Colombia. Email: nyrojasr@unal.edu.co

⁵ Agronomics engineer, Universidad Nacional de Colombia, Colombia. Affiliation: Department of Agronomy, Universidad Nacional de Colombia, Colombia. Email: caariveramo@unal.edu.co

⁶ PhD in statistics, Universidad de Los Andes, Venezuela. Affiliation: Researcher-professor, Universidad Nacional de Colombia, Colombia. Email: aqedarghanco@unal.edu.co

⁷ PhD in Mechanical Engineering, University of Leeds, United Kingdom. Affiliation: Researcher-professor, Universidad Técnica Federico Santa María, Chile. Email: mauricio.osses@usm.cl



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Highlights:

- Road transport-related emissions were assessed using spatial and temporal disaggregation.
- The study revealed cargo vehicles as major contributors in BC, NO₂, and PM_{2.5} emissions.
- The findings highlight motorcycles as a CO pollution source in Colombia.
- Passenger cars are the primary mobile source of CO₂ and SO₂ emissions in Colombia.
- Concentrated emissions and daily dispersion patterns pose challenges to air quality management in metropolitan areas.

Introduction

Poor air quality has extensive detrimental effects. For example, it endangers community sustainability, compromises environmental preservation, and reduces residents' quality of life [1]–[7]. Surprisingly, almost 92% of the world's population lives in regions where pollution levels surpass the tolerable thresholds, as reported by the World Health Organization in 2005 [8]. According to the National Health Observatory in 2019 [9], poor air quality is responsible for 7% of Colombia's yearly mortality rate.

In order to successfully tackle and control air quality problems, it is crucial to compile a comprehensive record of air pollutant emissions. The Environmental Protection Agency (EPA) [10] states that it offers essential quantitative data regarding the amount of pollutants released by multiple sources in a certain area over a specified period of time. Emission inventories are crucial for the formulation and oversight of public programs that target emissions reduction. They allow researchers to perform comparative evaluations of pollutant emissions across time, aiding environmental authorities and decision-makers in formulating air pollution management plans.

When creating an emissions inventory, it is crucial to include both climate forcers and criteria air pollutants. Climate forcers affect the Earth's climate system by emitting or absorbing radiation, thus contributing to climate change, with greenhouse gases (GHGs) being a major component. Criteria air pollutants are common and pose risks to the environment and human health [11]. Emission inventories also play a vital role in raising public awareness about air pollution in specific areas [12].

A clear grasp of atmospheric emissions is essential to prevent misunderstandings. When creating an emissions inventory, prioritizing data accuracy is also crucial. This is achieved by comparing projected emissions values against documented energy consumption statistics from the assessed sources. For instance, in the context of road transport, projected emissions can be verified against gasoline sales data from official organizations and emissions inventory obtained through various methods. The goal is to adjust transport activity values (e.g., annual kilometers traveled) in the inventory by consistently matching them with relevant statistical data for the region [13].

Spatial and temporal disaggregation of emissions is widely used in emissions inventories. This practice consists of dividing emissions into smaller areas (grids) and shorter time

periods (weekly and hourly records). This information is crucial for environmental management models that evaluate pollutant concentrations and atmospheric deposition. The spatial distribution of emissions plays a crucial role in determining how they spread and the magnitude of their effects in a specific region [14]. Several authors have suggested several approaches and guidelines to help break down emissions into spatial and temporal components [15]–[18].

Colombia regularly participates in emissions estimation exercises and is included in global inventories such as EDGAR [19]. EDGAR is a global database that tracks greenhouse gas emissions and air pollution using international statistics and the methodology of the Intergovernmental Panel on Climate Change (IPCC). Two distinct methodologies have been employed to calculate the national emissions inventory from mobile sources in Colombia: [20] assessed the emissions of fine particulate matter (PM_{2.5}), black carbon (BC), carbon monoxide (CO), and sulfur dioxide (SO₂) between 2010 and 2014, and [21] evaluated carbon dioxide (CO₂), methane (CH₄), nitrogen oxides (NO_x), volatile organic compounds (VOCs), PM_{2.5}, and BC emissions from 1990 to 2020 using 0.01 × 0.01° cells for high-resolution representation. These authors compared their results against national and global inventories to highlight emission trends and discrepancies.

Colombia is geographically, politically, and administratively separated into 32 departments, along with its capital city, Bogotá. These departments exhibit significant economic and demographic variety and demonstrate diverse patterns with regard to transportation [22]. Colombia's transportation system comprises 18 million vehicles, with automobiles accounting for nearly seven million and motorcycles making up the remaining 11 million [23]. The country possesses a vast road network that spans a significant portion of its land, measuring approximately 205 000 km in length. This network is categorized by the Ministry of Transportation into three distinct types: around 18 000 km are classified as *primary roads*, 45 000 km as *secondary roads*, and 142 000 km as *tertiary roads* [24].

The road transport sector is Colombia's primary source of pollution among several transportation subsectors, surpassing civil aviation, sea and river navigation, and railways [20]. In 2016, the transportation industry was responsible for 11% of the country's greenhouse gas emissions, with road transport accounting for 91% of that total [25]. In relation to criteria air pollutants, the transportation sector accounted for substantial portions of the yearly emissions,

specifically 15.5, 4.5, 27.2, 70.7, and 9.7% of the annual emissions of BC, PM_{2.5}, CO, nitrogen dioxide (NO₂), and SO₂, respectively [20].

Although road transport emissions in Colombia significantly impact air quality and public health, there is a lack of studies developing robust methodologies to estimate and analyze them based on their spatial and temporal distribution. Prior research has predominantly concentrated on estimating pollution levels at the national level and has only examined a few types of pollutants. As a result, there is no comprehensive understanding of the unique contributions of different types of vehicles and their geographical distribution throughout the country. Moreover, the application of cutting-edge data sources and sophisticated techniques for estimating and breaking down emissions has been restricted. In light of the above, this study seeks to fill the existing research gap by conducting a thorough and meticulous examination of road transport emissions in Colombia. This analysis encompasses various criteria air pollutants and climate forcers and utilizes cutting-edge data sources and advanced spatial-temporal disaggregation techniques.

The aim of this study is to calculate and break down the yearly emissions of specific air pollutants (CO, NO₂, PM_{2.5}, SO₂) and climate-altering agents (BC, CO₂) produced by road transport in Colombia throughout 2019 while considering their spatial and temporal distribution. To this effect, a methodology is employed which integrates various transportation data sources to determine the average mileage by vehicle type, such as vehicles listed for sale on websites and data acquired from mechanical and technical inspections. Additionally, this study incorporates traffic volume data from vehicle counts and average operational speed data from Google's Roads API, among other sources.

Method

The inventory estimation of exhaust emissions from road transport employs a top-down technique, where emissions estimates for smaller geographical areas are derived from broader or macro-scale data. Eq. (1) [13] presents the primary formula for calculating the overall annual emissions originating from mobile sources.

$$E_{ij} = AF_j \cdot EF_{ij} \cdot N_j \quad (1)$$

Here, the variable E_{ij} represents the overall amount of pollutant i emitted by vehicles in category j , which is measured in grams per year. The activity factor (AF_j) represents the annual distance traveled by vehicles of type j , measured in kilometers per year per vehicle. The emissions factor (EF) is a measure of the amount of pollutants emitted by a certain source. In the case of road transport, EF_{ij} is the emissions factor of pollutant i for vehicle category j , measured in grams per kilometer per vehicle. N_j represents the quantity of vehicles of type j inside the region under study.

Our research methodology is organized into six distinct stages aimed at outlining the sequential methods employed to obtain full estimates of road transport emissions. These stages cover important aspects such as vehicle activity, fleet composition, EFs, fuel sales validations, and consumption statistics. Fig. 1 illustrates the input, process, and output of each of the following stages:

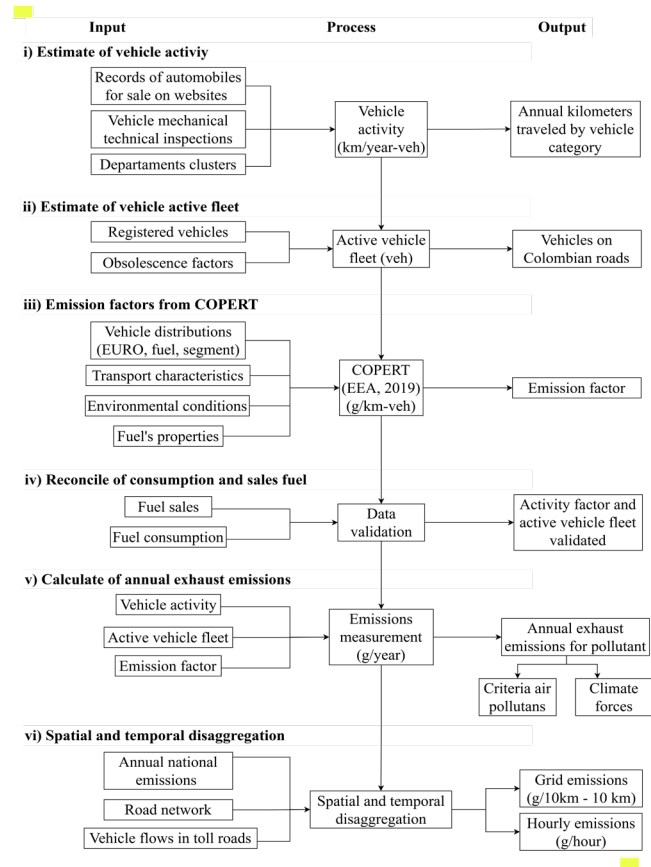


Figure 1. Method structure of the research

Source: Authors

Activity factor (AF)

The AF for road transport measures emissions intensity based on the total distance traveled by the active vehicles in a category over a year. This study used a clustering methodology to categorize departments with similar attributes, following [21]. We employed spatial cluster analysis to define six geographic clusters for 2019 while using data on the road network, registered vehicles, population, and gasoline sales. The annual mileage for each cluster was calculated using data from websites specializing in used vehicle sales (specifically www.tucarro.com.co), which included vehicle type and fuel information for passenger cars (excluding taxis) and motorbikes. Additionally, we incorporated data from the Ministry of Transport [24] and standardized inspections covering mechanical, environmental, and safety aspects from local environmental organizations and automobile diagnostic centers [26].

Regarding vehicle activity, motorbikes typically cover approximately 8000 km per year, whereas buses, tractor-trailers, and taxis often cover around 50 000 km. Table I presents the transportation activities for different vehicle classifications within each cluster.

Vehicle fleet

Data regarding the vehicle fleet in Colombia were obtained from the database of vehicles registered in the platform of the Single National Traffic Registry [27]. This information was not used for the direct estimation of emissions due to its substantial level of uncertainty, as the original database encompasses all vehicles that have been registered, regardless of their operational status.

To determine Colombia's vehicle fleet for 2019, several measures were implemented, including the exclusion of vehicles registered before 1990, the elimination of emissions-free vehicles (electric, hydrogen, ethanol, and LPG), and the application of an obsolescence factor for vehicles taken out of circulation due to accidents or poor conditions.

The obsolescence variables, unique to each vehicle category, were created via an iterative procedure aimed at attaining an optimal alignment between fuel consumption and sales data. The obsolescence factors used for the different vehicle categories in Colombia were as follows:

- buses: 3%
- heavy-duty vehicles: 1%
- L-Category vehicles: 3%

- passenger cars: 2%
- light commercial vehicles: 2%

The aforementioned parameters were employed to consider the proportion of vehicles that were no longer in active operation within each specific category, thereby yielding a more precise and detailed depiction of the active automobile fleet [21].

After conducting the necessary procedures, the initial count of 15 428 671 registered vehicles was adjusted to a refined estimate of 10 675 062, providing a more representative foundation for calculating emissions.

Emissions factor (EF)

Local EFs from fieldwork in Colombia are currently unavailable. The Colombian guide for estimating atmospheric emissions [12] recommends using reference EFs from international agencies like the EPA, the European Environment Agency (EEA), or the IPCC when representative EFs are not available for the study area. This study used the COPERT model [26] to calculate EFs based on vehicle segment, fuel type, EURO technology, and department.

COPERT requires multiple inputs to estimate an EF. The grouping of the fleet by EURO technology (Table A – supplementary information), fuel type, and vehicle segment (Table B – supplementary information) was based on data from the vehicle registration database [27], and the classification of the fleet by EURO standards presented

Table I. Annual activity by vehicle type and cluster (km)

Vehicle Category	Fuel	Cluster F*	Cluster E*	Cluster D*	Cluster C*	Cluster B*	Cluster A*
Motorcycle	Petrol	7 888	8 299	7 468	8 880	9 650	8 084
Passenger Cars	Diesel	17 708	16 227	15 995	18 382	15 957	17 181
Passenger Cars	Petrol / CNG	12 435	11 422	12 827	14 164	12 371	12 698
Taxi	Diesel	51 240	52 346	41 258	42 650	52 379	53 744
Taxi	Petrol / CNG	48 575	50 208	39 091	39 494	40 193	33 132
Bus	Diesel	46 305	53 240	45 890	43 568	45 201	43 521
Bus	Petrol / CNG	40 032	50 320	42 950	40 000	35 916	30 609
Truck	Diesel	38 588	47 667	38 588	39 231	45 210	42 130
Truck	Petrol / CNG	30 935	28 448	26 819	28 319	26 880	26 352
Rigid Truck	Diesel	40 396	42 500	40 396	40 396	38 210	35 301
Tractor-Trailer	Diesel	48 752	50 000	48 752	48 752	48 752	47 503

*Departments by cluster

Cluster A: Caldas, Cauca, Casanare, Huila, La Guajira, Magdalena, Sucre, Nariño

Cluster B: Cundinamarca, Valle del Cauca

Cluster C: Amazonas, Arauca, Archipelago of San Andres, Caqueta, Choco, Guainia, Guaviare, Putumayo, Quindio, Risaralda, Vaupes, Vichada

Cluster D: Atlantico, Bolivar, Cesar, North of Santander

Cluster E: Antioquia, Bogota

Cluster F: Boyaca, Cordoba, Meta, Tolima, Santander

Source: Authors

in Table 2 was based on official vehicle registration data. However, studies such as that by [28] and reports from the ICCT (2023) indicate that gasoline-powered light-duty vehicles in Colombia were primarily meeting EURO 2/Tier 1 standards in practice, particularly in major cities. While EURO 4 standards were expected to be applied to new vehicles from 2023 onwards, the 2019 fleet comprised a mix of technologies, including older vehicles that remained in operation. This distinction is important when interpreting emissions estimates, as real-world values may differ from regulatory classifications.

The analysis considered various transport attributes, such as the mean operational velocity from historical data via Google Big Data, the total distance traveled by each vehicle category, the incline (estimated using accurate techniques), and the carrying capacity (for buses and freight vehicles) from BID data [29]. Environmental conditions, including temperature, relative humidity, and air conditioning consumption, were sourced from the Information System for the Management of Hydrological and Meteorological Data (DHIME-IDEAM) for each department and month of the reference year. Fuel characteristics were also taken into account.

Using the inputs in COPERT, we derived the EFs for each department, vehicle category, segment, motor technology, fuel type, and European emissions standard for four criteria air pollutants (CO , NO_2 , $\text{PM}_{2.5}$, SO_2) and two climate forcers (BC , CO_2). Consequently, a total of 15 357 emission variables were acquired and utilized for estimating emissions.

Data validation via the reconciliation of fuel consumption and sales

To verify the estimated activity factor and the active vehicle fleet data, a comparison was made between the estimated fuel consumption and sales to road transport. Validating this data was crucial, as national statistics may not distinguish between on-road and off-road fuel use, affecting the accuracy of emissions estimation [30,31].

According to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, it is good practice to estimate fuel consumption using data on distances traveled and vehicle and fuel type.

The fuel sales were determined using the Colombian energy balance for 2019, as reported by UPME in 2021. According to [33], the road transport sector sold a total of 1 986 370 113 gallons of petrol and 1 584 403 079 gallons of diesel in 2019. The approach outlined in the fourth stage of the methodology, which involved reconciling fuel consumption and sales, yielded a total consumption of 1 872 107 705 gallons of petrol and 1 521 986 611 gallons of diesel. Therefore, there is a predetermined agreement between the amount of gasoline used and the amount of gasoline sold, with a difference of 5.8%. Similarly, for diesel, there is a predetermined agreement between consumption and

sales, with a difference of 3.9%. This validates the projected activity of road transport and the approximation to the number of active vehicles, which in turn reduces uncertainty in quantifying emissions.

Estimated emissions

After validating data consistency, we estimated the exhaust emissions for each department using the methodological approximation outlined in Eq. (1). This computation considered the current number of vehicles in use, the yearly usage per kind of vehicle, and the emission rates specified by the category, segment, engine technology, and fuel type of each vehicle. The application of these elements allowed determining the departmental emissions related to criteria air pollutants and climate forcers, which in turn led to the estimation of national emissions associated with road transport.

Spatial and temporal disaggregation

The spatial disaggregation of emissions enhances the local-scale resolution, aiding air quality models in assessing pollutant dispersion and deposition [14]. Geospatial tools effectively manage and visualize emissions, often using grids to standardize datasets and integrate emission sources [33]. The literature suggests a $0.1 \times 0.1^\circ$ grid (longitude/latitude) for national reporting. This is approximately comparable to cells that are 11×11 km in size [34]. This study employed a cell size of 10×10 km, as well as a more detailed resolution of 1×1 km to examine emissions in urban areas and local government units.

Different approaches have been suggested for breaking down emissions into smaller geographical units, typically relying on parameters like road hierarchy, vehicle flows, population density, or urbanized regions [35]. Building on the suggestions provided by earlier research works conducted by [36], [17], and [37], this study utilized a computational tool called *DROVE* [16]. This tool employs an algorithm to calculate disaggregation factors (DFs) for each grid cell. DFs are determined based on specific data, e.g., road segment length, road types, and traffic flow. It should be noted that the accuracy of the disaggregated emissions inventory is contingent upon the quality of these data. The emission fluxes per grid cell are calculated by multiplying the DFs by the total emissions of each pollutant while ensuring the conservation of the total mass of emissions, as described by [16].

The *DROVE* algorithm provides three levels of disaggregation based on the given information. The primary level solely necessitates the road network, while the second level encompasses supplementary data such as road length, hierarchy types, and average flow based on road type. The highest level, which is the most sophisticated, employs a transportation model of the research region. In this analysis, we employed level two, as indicated by Eq. (2) [16].

$$DF_{I,J,K} = \left(\frac{L_j}{\sum_0^n L_j} \right) * fm_{i,k} \quad (2)$$

The variables in this equation are defined as follows:

- DF represents the disaggregation factor
- L denotes the total length of the road segments within each grid cell
- j corresponds to the grid cell
- n represents the total number of cells in the domain
- k signifies the type of road
- the subscript I denotes the vehicle category
- fm represents the weight factor of the type of road

The data on the road network inside urban and interurban areas was acquired from OpenStreetMap's publicly available data. This data was then categorized into six distinct hierarchical groups, based on their functional features. The classification of roads includes interurban roads categorized as primary, secondary, and tertiary, and urban roads categorized as primary, local roads categorized as secondary, and residential roads categorized as tertiary. Fig. 2 depicts the layout of the road network, while Fig. 3 illustrates the process of geographical disaggregation for one of the 32 departments in Colombia.

We generated two temporal distribution curves to break down the annual emissions into hourly levels, using transportation demand as the basis. These curves distinguished between metropolitan regions and interurban roadways, offering

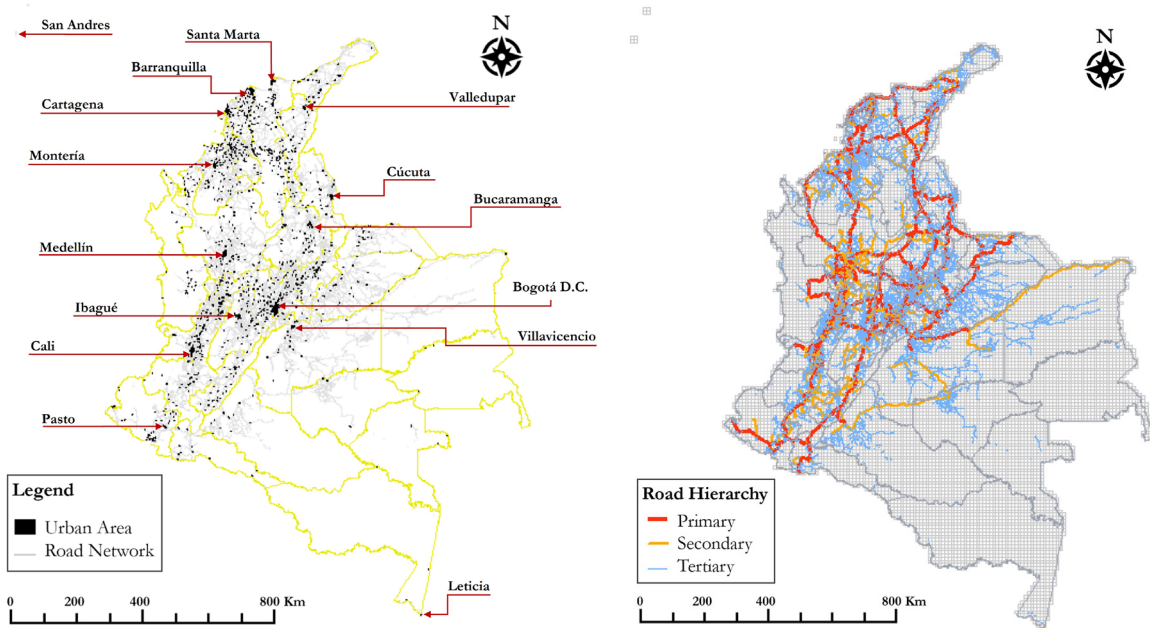


Figure 2. Urban area and road network configuration

Source: Authors

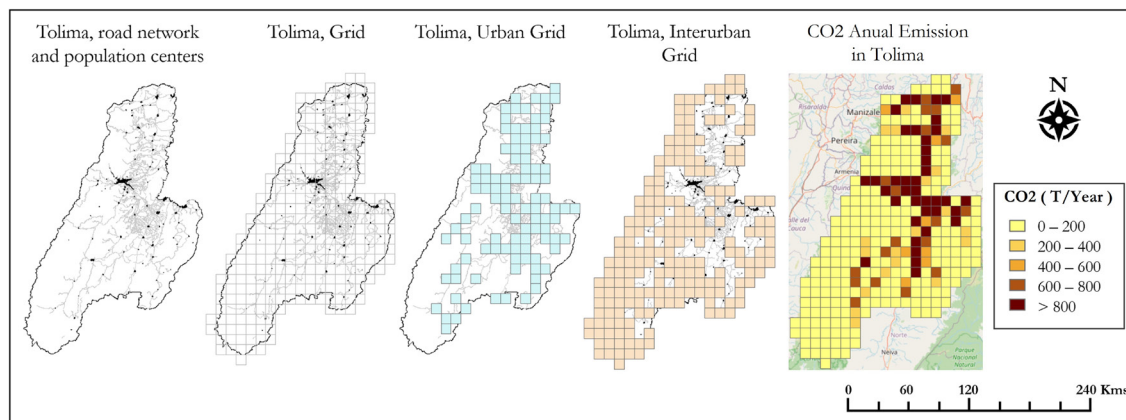


Figure 3. Tolima's emissions disaggregation using DROVE at a 10 x 10 km resolution

Source: Authors

valuable information about emission patterns throughout the day. The estimation of hourly profiles is limited by its use of a uniform factor. However, it was clearly established that the selected reference cities provide a representative sample. The urban curve was constructed using authoritative data obtained from mobility surveys carried out in the cities and regions included in our analysis, such as Bogotá [38], Valle de Aburrá [39,40], Cali [41], and Manizales [42].

In contrast, the interurban curve was built using data obtained from national highway toll booths provided by the National Infrastructure Agency [43]. These data sources ensured the dependability and inclusiveness of the analysis of the temporal distribution in urban and interurban areas.

For a typical day, the three peaks observed in the urban environment can be explained by daily commuting patterns: morning trips to work, midday travel associated with work breaks, and evening returns home. Additionally, the selection of reference cities was justified by the fact that the four departments they belong to (Bogotá DC, Antioquia, Valle del Cauca, and Atlántico) account for more than half of the total vehicle fleet in the country. This substantial representation enhances the robustness and representativeness of the temporal distribution analysis. Fig. 4 depicts the curves obtained.

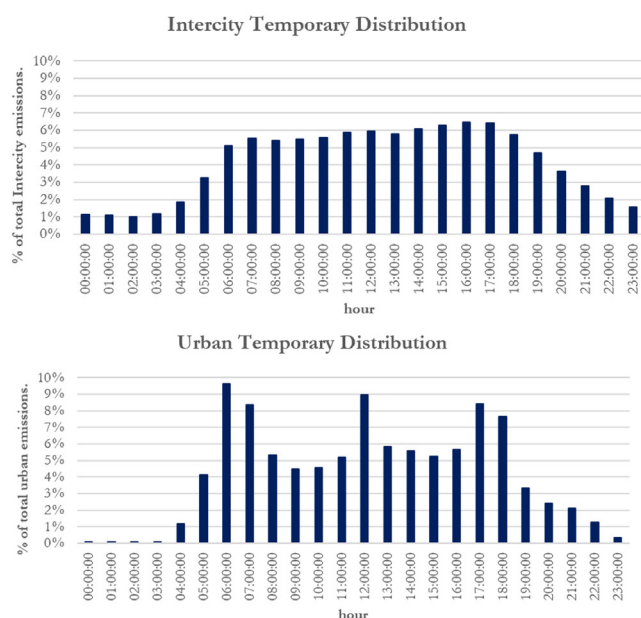


Figure 4. Urban and interurban profiles of the hourly vehicle flow intensities in Colombia

Source: Authors

Results and discussion

Our analysis of exhaust emissions from road transport in Colombia provides insights into the distribution of air pollutants and climate forcers. In 2019, road transport emitted approximately 34 184 853 metric tons (T) of CO₂ and 2 352 T of BC. The Colombian emissions inventory

for 2018 reported 28 894 520 T of CO₂ emitted by road transport, accounting for 13.6% of the total national CO₂ emissions, 34.7% of the CO₂ energy emissions, and 78.6% of the CO₂ emissions in the transportation sector.

Our results for 2019 evidence an 18% increase in CO₂ emissions from road transport in comparison with the 2018 national emissions inventories. The annual emissions of criteria air pollutants show a significant coherence between our results and those of [20] for the 2010-2014 period (Fig. 5). Compared to 2010, SO₂ emissions decreased by 17.8%, PM_{2.5} emissions by 51.5%, and BC emissions by 13.9%. These reductions can be attributed to the conversion of the private vehicle fleet to cleaner technologies, as demonstrated in Table A (supplementary information). Conversely, CO emissions increased by 14.5%, primarily due to motorcycle emissions, which have a high EF. This rise is linked to a significant increase in motorcycle registrations, making up 58% of all vehicles in 2019, and Colombia's lenient regulations, which only required EURO 2 technology in 2019, despite the advancement to EURO 5.

In terms of CO₂, we found that passenger cars represent the highest emissions, accounting for 33.3% of the total (Fig. 6). On the other hand, with regard to BC, freight vehicles make the largest contribution (54.8% of the emissions) (Fig. 6). Although BC is not a greenhouse gas, it can absorb light, leading to localized warming effects both in the atmosphere and on the Earth's surface. This warming effect becomes particularly significant in regions marked by high BC emissions, such as densely populated urban areas and transportation corridors.

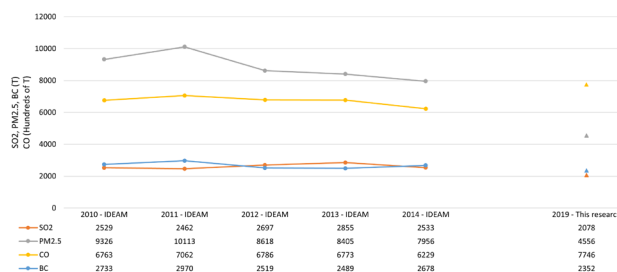


Figure 5. National emissions from road transport

Source: Authors

The emissions of criteria air pollutants resulting from road transport in Colombia for 2019 were estimated as follows: 2078 T of SO₂, 4556 T of PM_{2.5}, 774 637 T of CO, and 16 446 T of NO₂. The amount of SO₂ emitted is directly proportional to the sulfur content present in the fuel. In Colombia, gasoline contains a higher sulfur content compared to diesel fuel, leading to passenger cars and motorcycles contributing 54.9 and 25.9% of the SO₂ emissions, respectively. As for PM_{2.5}, which is known to be significantly associated with respiratory and cardiovascular diseases, it was found that freight vehicles and buses are the largest emitters, accounting for 49.3 and 25.3% of the emissions. Motorcycles, representing approximately 58% of the registered vehicle fleet, contribute the most to CO emissions, constituting 69.1% of the total.

As for NO_2 , freight vehicles and buses emit 53.1 and 35.8% of the total emissions, respectively.

In 2019, freight vehicles contributed significantly to the national emissions in Colombia; they accounted for 54.8% of BC, 53.1% of NO_2 , and 49.3% of $\text{PM}_{2.5}$. This finding aligns with the fact that Colombia has one of the oldest cargo transport fleets in the region, which translates into a higher proportion of less efficient vehicles [29]. The second most polluting vehicle category was passenger cars, accounting for 33.3% of CO_2 and 54.9% of SO_2 . Lastly, motorcycles were the third most polluting category, with 69.1% of the CO emissions.

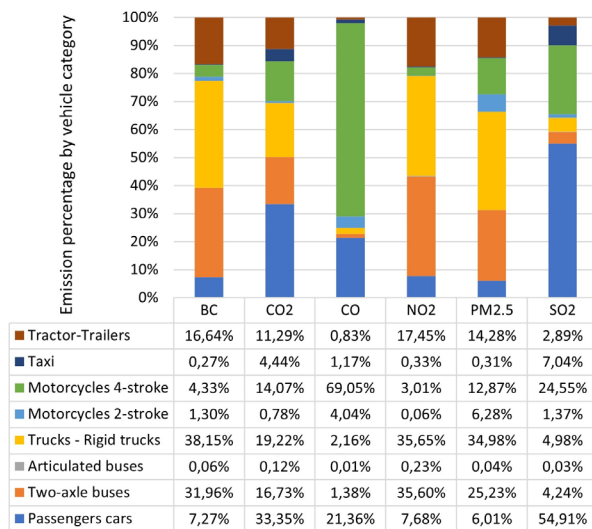


Figure 6. Emissions of each pollutant by vehicle category
Source: Authors

Spatial distribution of pollutant emissions in the regions of Colombia

Upon completion of the annual emission calculations, it became evident how various pollutants are distributed across different regions. Notably, the departments characterized by larger populations, extensive road networks, and substantial registered vehicle fleets, such as Bogotá, Cundinamarca, Antioquia, and Valle del Cauca, exhibit the highest emission loads.

Fig. 7 provides insights into the spatial distribution of emissions across departments and highlights the relative magnitudes of each pollutant. Notably, the annual CO_2 emissions exhibit a significantly higher load compared to other pollutants. Nationally, road transport reports a total annual CO_2 emission of 34.18 million T, while the emissions of SO_2 amount to only 2.08 T.

Our analysis of the spatial distribution of pollutant emissions from mobile sources revealed a strong correlation between economic activity and pollutant emissions at the departmental level. The disaggregation into 10×10 and 1×1 km grids allowed for an accurate representation at both the national and urban levels (Figs. 8 and 9). Consistent with the study by [44] in India, the statistical analysis shows high correlation coefficients between emissions and the departmental GDP, reaching values of 0.98 for SO_2 , 0.9 for CO, 0.87 for CO_2 , 0.67 for $\text{PM}_{2.5}$, 0.66 for BC, and 0.64 for NO_2 , indicating a significant relationship. This supports the trend observed in previous studies, such as that by [45], which highlighted the link between economic growth and increased emissions from the transport sector. However,

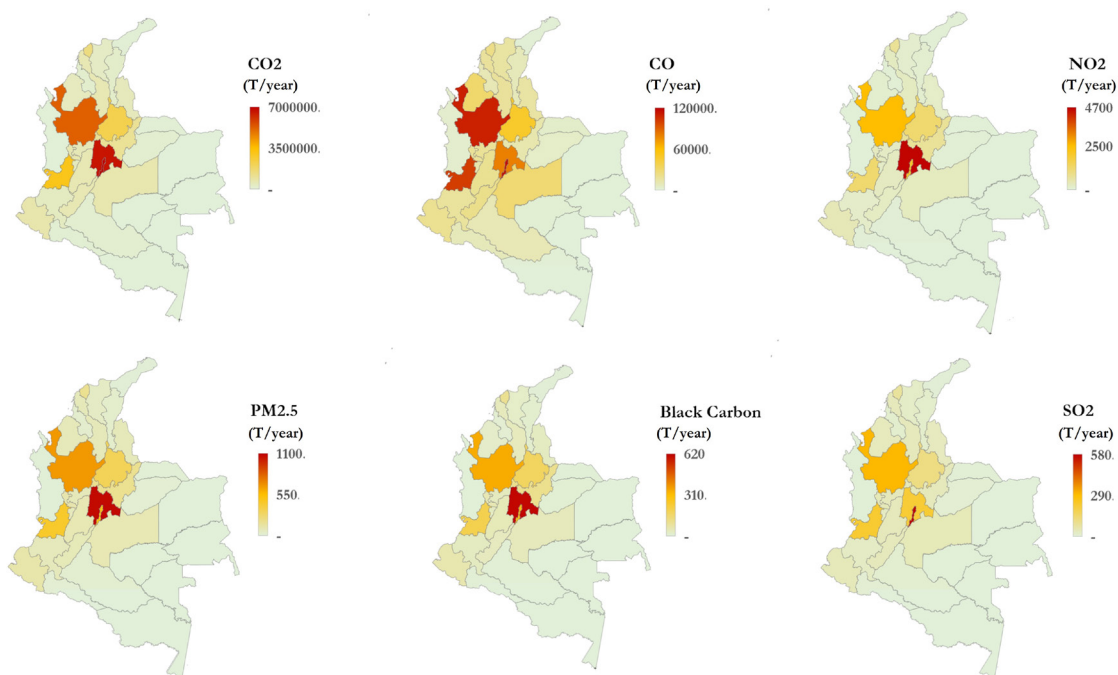


Figure 7. Annual road exhaust emissions by department
Source: Authors

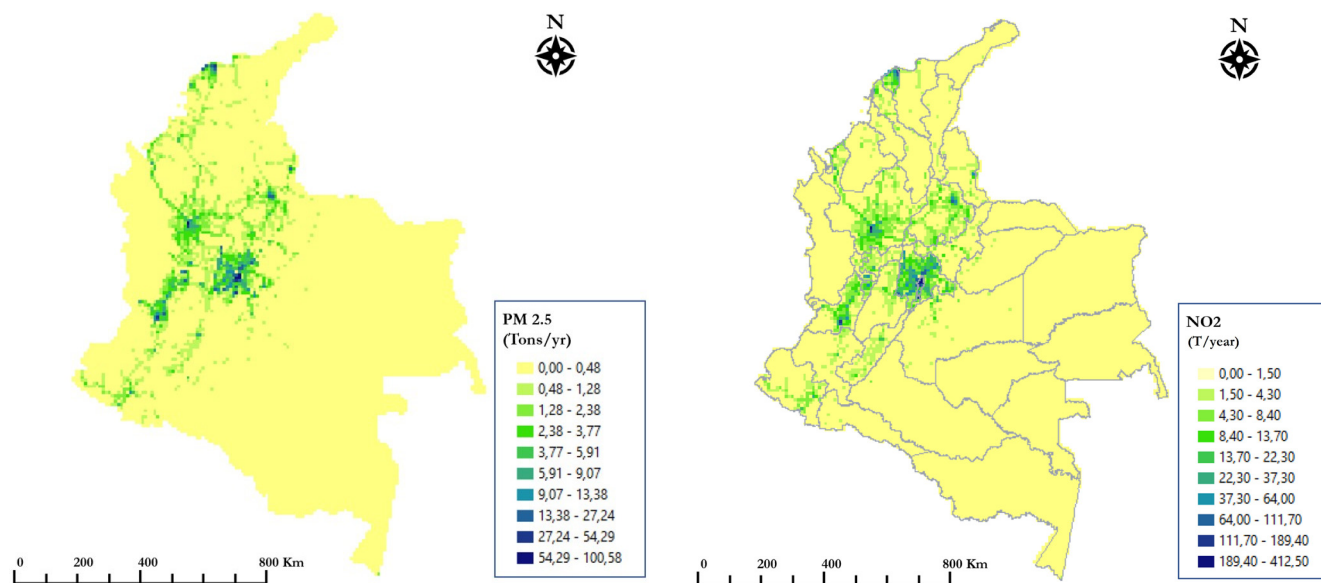


Figure 8. Spatial distribution of mobile source emissions by pollutant ($PM_{2.5}$ and CO_2) at the national level (10×10 km grid)
Source: Authors

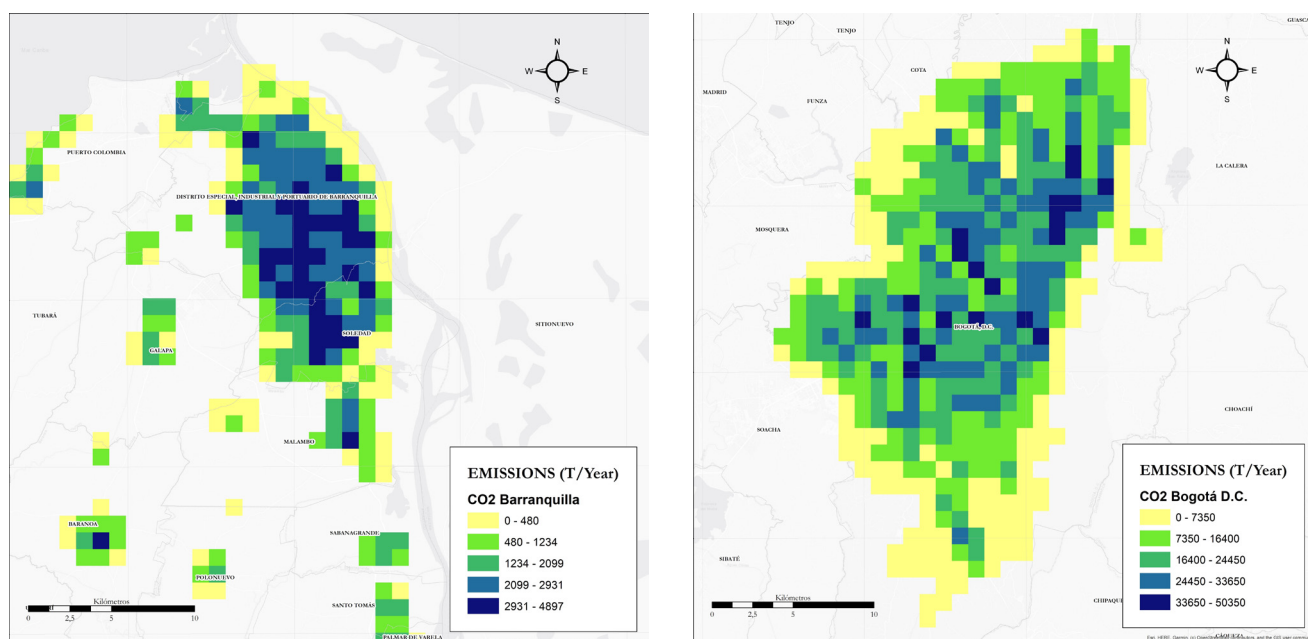


Figure 9. Spatial distribution of CO_2 emissions from mobile sources at the urban level in Barranquilla and Bogotá (1×1 km grid)
Source: Authors

a particular case is Cundinamarca, where emissions are disproportionately high compared to its GDP, potentially due to the relocation of economic and logistical activities from Bogotá to this department.

Conclusions

Utilizing various tools and data sources, this study estimated and spatially and temporally disaggregated the exhaust emissions from road transport in Colombia for the year 2019. With the purpose of enhancing the coherence and

representativeness of the estimates, alternative techniques were employed to determine activity factors for mobile sources. These techniques included data from mechanical and technical inspections provided by local environmental agencies and automotive diagnostic centers, as well as records of vehicles for sale on online platforms. Additionally, vehicle obsolescence factors were identified, and the active fleet was assessed, thereby reducing uncertainty, reconciling fuel consumption and sales, and achieving a more accurate emissions estimation. The analysis indicated that road transport significantly contributes to climate forcing and air pollution by criteria pollutants.

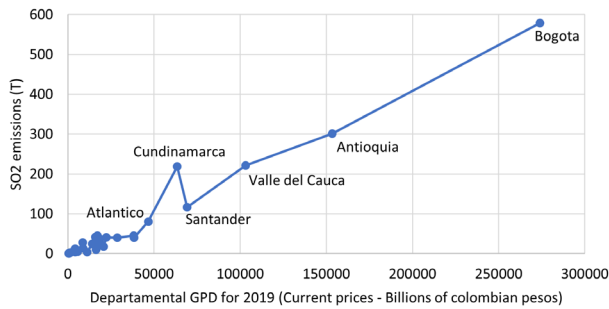


Figure 10. Relationship between SO₂ emissions and departamental GDP in 2019 (correlation coefficient of 0.98)

Source: Authors

The spatial distribution of the annual emissions was analyzed using a weighting method based on the road network. The areas with the highest emissions were identified at a resolution of 10×10 km while considering temporal variability. It was established that the highest emissions occur in urban regions and are distributed throughout the day, with three peak hours identified.

The estimates of this study confirm that passenger cars and freight vehicles were the primary sources of emissions in 2019, with the former contributing 33.3% of CO₂ emissions and the latter accounting for 54.8% of BC emissions. Additionally, the main sources of criteria pollutant emissions were identified: passenger cars and motorcycles were responsible for the majority of SO₂ emissions, while freight vehicles and buses were the leading emitters of PM_{2.5} and NO₂. Motorcycles represented the largest share of CO emissions.

Departmental estimates indicate that Bogotá, Antioquia, Cundinamarca, and Valle del Cauca account for the majority of road transport emissions in Colombia, with impacts ranging from 52 to 64% depending on the pollutant. A direct relationship was identified between pollutant emissions and economic activity in these regions, demonstrating that GDP growth drives increased transport demand and, consequently, elevated emissions. Without effective regulations, the impact of transport on air quality and the environment will likely continue to rise.

Cundinamarca's case highlights the need to consider the relocation of productive activities in emissions reduction policies. Its emissions are disproportionately high relative to its GDP, likely due to operations shifting from Bogotá, which has concentrated economic activity and transport. These findings underscore the necessity of targeted measures to ensure sustainable development in areas where economic growth leads to excessive emissions.

While studies like that by [21] offer valuable insights into long-term trends and the effects of environmental policies, our analysis delivers a comprehensive assessment of the current situation. Additionally, the identification of daily dispersion patterns with specific hourly peaks underscores the necessity of implementing control and regulatory

measures tailored to each vehicle type and its operational context. This approach will enable the formulation of effective strategies aimed at mitigating road transport emissions in Colombia. We conclude that the country should mandate the exclusive commercialization of new vehicles equipped with the most advanced EURO technology available for each segment. Furthermore, it is essential to establish a maximum lifespan for freight vehicles in order to prevent the circulation of obsolete trucks with poor environmental performance, which adversely affects transportation efficiency and air quality. Lastly, stricter regulations on engine technology for motorcycles are recommended, as its rapid growth has significantly increased CO emissions.

One of the key uncertainties in emissions estimation is the real-world application of EURO standards in the Colombian fleet. While official records classify new gasoline vehicles under EURO 4, studies (e.g., [28], ICCT, 2023) suggest that a significant portion of the vehicles in circulation in major cities were still operating under EURO 2 or Tier 1 standards in 2019. This discrepancy may affect emissions estimates, particularly for pollutants sensitive to technological advancements in emissions control. Based on the results obtained in this study, the team's future work will focus on projecting road transport emissions under different technological advancement scenarios and evaluating compliance with national emissions reduction targets.

The scientific community is encouraged to utilize local EFs in future studies (UPME is already working on this) and to estimate all vehicle emissions, as this study only calculated exhaust emissions.

CRediT author statement

Julieth V. Alfonso A.: Conceptualization, methodology, data analysis, formal analysis, writing (original draft).

Sonia C. Mangones, Néstor Y. Rojas: Conceptualization, methodology, validation, writing (review and editing).

Mauricio Osses: Validation, writing (review and editing).

Dayron Bermúdez: Data curation, writing (original draft).

Carlos Rivera, Aquiles Darghan: Data curation.

All authors have read and agreed to the published version of the text.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could appear to influence the work reported in this paper.

Data availability

The data that support the findings of this study are available upon request from the corresponding author.

Funding

This study was supported by Universidad Nacional de Colombia (grant no. 51326).

Additionally, we acknowledge the support from Center of Climate and Resilience Research, CR2, FONDAP no. 1523A0002, and Fondecyt Regular no. 1241477.

References

- [1] M. R. Miller and D. Newby, "Air pollution and cardiovascular disease: Car sick," *Cardiovasc. Res.*, vol. 116, no. 2, pp. 279-294, 2019. <https://academic.oup.com/cardiiovascres/article/116/2/279/5579822>
- [2] C. Knittel, D. L. Mille, and N. J. Sanders, "Caution, drivers! Children present: Traffic, pollution, and infant health," *Nat. Bureau Econ. Res.*, working paper no. 17222, 2011. <https://www.nber.org/papers/w17222>
- [3] J. Currie and R. Walker, "Traffic congestion and infant health: Evidence from E-ZPass.," *Amer. Econ. J. App. Econ.*, vol. 3, no. 1, pp. 65-90, 2011. <https://www.aeaweb.org/articles?id=10.1257/app.3.1.65>
- [4] R. B. Smith *et al.*, "Impact of London's road traffic air and noise pollution on birth weight: Retrospective population based cohort study," *BMJ (Clin. Res. Ed.)*, vol. 359, art. j5299, 2017. <https://www.bmj.com/content/359/bmj.j5299>
- [5] W.-J. Guan, X.-Y. Zheng, K. F. Chung, and N.-S. Zhong, "Impact of air pollution on the burden of chronic respiratory diseases in China: Time for urgent action," *The Lancet*, vol. 388, pp. 1939-1951, 2016. <https://www.sciencedirect.com/science/article/abs/pii/S0140673616315975>
- [6] C. Song *et al.*, "Air pollution in China: Status and spatiotemporal variations.," *Environ. Poll.*, vol. 227, pp. 334-347, 2017. <https://www.sciencedirect.com/science/article/abs/pii/S0269749116322850>
- [7] Q. Xiao *et al.*, "Pediatric emergency department visits and ambient Air pollution in the U.S. State of Georgia: A case-crossover study," *Environ. Health Glob. Access Sci. Source*, vol. 15, no. 1, art. 115, 2016. <https://ehjournal.biomedcentral.com/articles/10.1186/s12940-016-0196-y>
- [8] World Health Organization, "Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. Actualización mundial 2005," 2005. [Online]. Available: <https://apps.who.int/iris/handle/10665/69478>
- [9] National Health Observatory and National Institute of Health, "Carga de enfermedad ambiental en Colombia – Informe técnico especial 10," 2019. [Online]. Available: <https://www.ins.gov.co/Direcciones/ONS/Informes/10%20Carga%20de%20enfermedad%20ambiental%20en%20Co>
- [10] Environmental Protection Agency, "Managing air quality – Emissions Inventories," 2022. [Online]. Available: <https://www.epa.gov/air-quality-management-process/managing-air-quality-emissions-inventories>
- [11] IDEAM, "Informe del estado de la calidad del aire en Colombia 2018" 2019. [Online]. Available: <https://www.andi.com.co/Uploads/Informe%20estado%20calidad%20del%20aire%202018.pdf>
- [12] Ministerio de Ambiente y Desarrollo Sostenible, "Guía para la elaboración de inventarios de emisiones atmosféricas." 2017. [Online]. Available: <https://www.minambiente.gov.co/index.php/asuntos-ambientales-sectorial-y-urbana/gestion-del-aire/emisiones-contaminantes>
- [13] European Monitoring and Evaluation Programme and European Environment Agency, *EMEP/EEA air pollutant emissions inventory guidebook 2019: Technical guidance to prepare national emission inventories*. Luxembourg: Publications Office of the European Union, 2019. <https://www.eea.europa.eu/en/analysis/publications/emep-eea-guidebook-2019>
- [14] N. Veldeman and W. van der Maas, "Spatial mapping of emissions, Guidebook 2013," European Environment Agency, 2013. [Online]. Available: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2013/part-a-general-guidance-chapters/7-spatial-mapping-of-emissions>
- [15] J. Londoño, M. A. Correa, and C. A. Palacio, "Estimación de las emisiones de contaminantes atmosféricos provenientes de fuentes móviles en el área urbana de Envigado, Colombia," *Rev. EIA*, vol. 16, pp. 149-162, 2011. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-12372011000200012&lng=en&tlng=es
- [16] C. M. González, C. D. Gómez, and B. H. Aristizábal, "DROVE: An algorithm for spatial and temporal disaggregation of on-road vehicle emission inventories," *Aerosol Air Qual. Res.*, vol. 20, no. 12, pp. 2765-2779, 2020. <https://doi.org/10.4209/aaqr.2020.04.0184>
- [17] C. D. Gómez, C. M. González, M. Osses, and B. H. Aristizábal, "Spatial and temporal disaggregation of the on-road vehicle emission inventory in a medium-sized Andean city. Comparison of GIS-based top-down methodologies," *Atmos. Environ.*, vol. 179, pp. 142-155, 2018. <https://doi.org/10.1016/j.atmosenv.2018.01.049>

- [18] M. Osses *et al.*, "High-resolution spatial-distribution maps of road transport exhaust emissions in Chile, 1990-2020," *Earth Sys. Sci. Data*, vol. 14, no. 3, pp. 1359-1376, 2022. <https://doi.org/10.5194/essd-14-1359-2022>
- [19] European Commission, "EDGAR – Emissions database for global atmospheric research," 2022. [Online]. Available: <https://edgar.jrc.ec.europa.eu/>
- [20] IDEAM, "Primer inventario nacional de emisiones de contaminantes criterio y carbono negro para el periodo 2010-2014," 2020. [Online]. Available: <http://documentacion.ideam.gov.co/openbiblio/bvirtual/023893/1InventarioBLACK.pdf>
- [21] N. Y. Rojas *et al.*, "Road transport exhaust emissions in Colombia. 1990-2020 trends and spatial disaggregation," *Trans. Res. Part D Trans. Environ.*, vol. 121, art. 103780, 2023. <https://doi.org/10.1016/j.trd.2023.103780>
- [22] J. C. Ramírez and J. de Aguas, "Configuración territorial de las provincias de Colombia. CEPAL – Serie estudios y perspectivas," 2022. [Online]. Available: <https://www.cepal.org/es/publicaciones/48021-configuracion-territorial-provincias-colombia>
- [23] Ministerio de Transporte, "RUNT en cifras: parque automotor (2023)" 2023. [Online]. Available: <https://www.runt.com.co/runt-en-cifras/parque-automotor>
- [24] Ministerio de Transporte, "Transporte en cifras vigencia," 2020. [Online]. Available: <https://plc.mintransporte.gov.co/Estad%C3%ADsticas/Transporte-en-Cifras>
- [25] IDEAM, "Tercera comunicación nacional de Colombia a la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC): Inventario nacional y departamental de gases efecto invernadero – Colombia," 2016. [Online]. Available: <http://www.ideam.gov.co/web/siac/infonalcliclimatico>
- [26] Ministerio de Transporte, Ministerio de Ambiente and Ministerio de Vivienda y Desarrollo Territorial, "Resolución 3500 – 2005," 2005. [Online]. Available: <https://corponarino.gov.co/expedientes/tramites/2005res3500.pdf>
- [27] Ministerio de Transporte, "Registro Nacional de Automotores. Registro Único Nacional de Tránsito," 2020. [Online]. Available: <https://www.runt.com.co/registros-runt/rna>
- [28] J. Ramírez, J. E. Pachón, O. Casas, and S. González, "A new database of on-road vehicle emission factors for Colombia: A case study of Bogotá," *CT&F Cien. Tecnol. Fut.*, vol. 9, pp. 73-82, 2019. <https://doi.org/10.29047/01225383.154>
- [29] BID, "Logística en América Latina y el Caribe: oportunidades, desafíos y líneas de acción," 2021. [Online]. Available: <https://publications.iadb.org/es/logistica-en-america-latina-y-el-caribe-oportunidades-desafios-y-lineas-de-accion>
- [30] Intergovernmental Panel on Climate Change, "2006 IPCC guidelines for national greenhouse gas inventories, prepared by the National Greenhouse Gas Inventories Programme," 2006. [Online]. Available: https://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/2_Volume2/V2_3_Ch3_Mobile_Combustion.pdf
- [31] A. Hoen, L. Leestemaker, P. Scholten, G. Geilenkirchen, and M. Traa, "Verkoop en verbruik wegbrandstoffen: Onderzoek naar verschillen verkoop en verbruik," 2022. [Online]. Available: <https://cedelft.eu/publications/sale-and-consumption-of-road-transport-fuels/>
- [32] Unidad de Planeación Minero Energética, "Colombian energy balance (BECO)," 2021. [Online]. Available: <https://www1.upme.gov.co/DemandayEficiencia/Paginas/BECO-Consulta.aspx>
- [33] N. Veldeman and W. van der Maas, "Spatial mapping of emissions, Guidebook 2019," European Environment Agency, 2019. <https://www.eea.europa.eu/publications/emep-eea-guidebook-2019/part-a-general-guidance-chapters/7-spatial-mapping-of-emissions/view>
- [34] N. Veldeman *et al.*, "Spatial mapping of emissions," in *EMEP/EEA Air Pollutant Emission Inventory Guidebook - 2016*. Luxembourg: Publications Office of the European Union, 2017. <https://www.eea.europa.eu/publications/emep-eea-guidebook-2016>
- [35] D. Tuia, M. Ossés de Eicker, R. Zah, M. Osses, E. Zarate, and A. Clappier, "Evaluation of a simplified top-down model for the spatial assessment of hot traffic emissions in mid-sized cities," *Atmos. Environ.*, vol. 41, no. 17, pp. 3658-3671, 2007. <https://doi.org/10.1016/j.atmosenv.2006.12.045>
- [36] P. Saide, R. Zah, M. Osses, and M. Ossés de Eicker, "Spatial disaggregation of traffic emission inventories in large cities using simplified top-down methods," *Atmospheric Environment*, vol. 43, no. 32, pp. 4914-4923, 2009. <https://doi.org/10.1016/j.atmosenv.2009.07.013>
- [37] M. Osses *et al.*, "High-definition spatial distribution maps of on-road transport exhaust emissions in Chile, 1990–2020," *Earth Syst. Sci. Data Disc.*, vol. 14, no. 3, 1359-1376, 2021. <https://doi.org/10.5194/essd-2021-218>
- [38] Secretaría Distrital de Movilidad, "Encuesta de movilidad 2019," 2019. [Online]. Available: https://www.movilidadbogota.gov.co/web/sites/default/files/Paginas/22-04-2020/20191216_presentacion_encuesta_v2.pdf
- [39] Área Metropolitana del Valle de Aburrá, "Encuesta de origen destino 2017," 2017. [Online]. Available: <https://www.metropol.gov.co/observatorio/Paginas/encuestaorigendestino.aspx>

- [40] Aburra Valley Metropolitan Area and Pontificia Universidad Bolivariana, "Actualización inventario de emisiones atmosféricas del Valle de Aburrá - Año 2018," 2018. [Online]. Available: <https://www.metropol.gov.co/ambiental/calidad-del-aire/Paginas/Herramientas-de-gestion/Inventario-de-emisiones-atmosfericas.aspx>
- [41] Metro Cali , "Encuesta de movilidad hogares Cali 2015 – Producto 4: Indicadores Encuesta de Movilidad," 2015. [Online]. Available: <https://metrocali.gov.co/wp/wp-content/uploads/2019/02/Encuesta-de-movilidad-2015.pdf>
- [42] Findeter, "Plan Maestro de Movilidad de Manizales," 2018. [Online]. Available: <https://www.infimanizales.com/wp-content/uploads/2018/06/Plan-Maestro-de-Movilidad-Manizales-PDF.pdf>
- [43] Agencia Nacional de Infraestructura, "SIG Peajes. ANISCOPIO," 2023. [Online]. Available: <https://aniscopio.ani.gov.co/datos-abiertos>
- [44] R. Singh, C. Sharma, and M. Agrawal, "Emission inventory of trace gases from road transport in India," *Trans. Res. Part D*, vol. 52, pp. 64-72, 2017. <https://www.sciencedirect.com/science/article/abs/pii/S1361920915300122>
- [45] R. Jiménez and D. Pulido-Guío, "Disparidad y deslocalización de emisiones de gases de efecto invernadero en la Región Cundinamarca-Bogotá," presented at *VI Col. Cong. Int. Conf. Air Quality Pub. Health (CASAP)*, 2017.

Supplementary Information

Table A. Mapping the Colombian automobile fleet to EURO vehicle technologies

Vehicle Category	Fuel / Service type	1996 and earlier	Since 1997	Since 1998	Since 2000	Since 2001	Since 2010	Since 2011	Since 2013	Since 2014	Since 2015	Since 2019
Cars and taxis	Diesel / Private	Conventional					Euro 2				Euro 4	
	Diesel / Public	Conventional					Euro 4					
	Petrol	ECE 15/04		Euro 1			Euro 2			Euro 3		Euro 4
	Vehicular natural gas	Euro 4										
	Gasoline electric	Euro 6										
Cargo transportation vehicles	Diesel / Private	Conventional				Euro I	Euro II				Euro IV	Euro V
	Diesel / Public	Conventional				Euro I	Euro V					
	Petrol	Conventional	Euro I				Euro II			Euro III		Euro IV
Buses	Diesel / Private	Conventional			Euro I		Euro II		Euro IV			Euro V
	Diesel / Public	Conventional			Euro I		Euro IV					Euro V
	Vehicular natural gas / Private	Euro I					Euro II				Euro IV	
	Vehicular natural gas / Public	Euro I					Euro IV					
	Electric diesel	Euro VI										
Motorcycle	Petrol	Conventional						Euro 2				

Source: Authors

Table B. Vehicle segment and category rules

RUNT Class	Category	Segment	Classification rules
Car, van, camper	Passenger cars	Mini Small Medium Large-SUV-Executive	Cylinder capacity $\leq 1200 \text{ cm}^3$ Cylinder capacity $\leq 1600 \text{ cm}^3$ Cylinder capacity $\leq 2000 \text{ cm}^3$ Cylinder capacity $> 2000 \text{ cm}^3$
Tricycle motorcycle, motorcar, moped, quadricycle, four-wheeler	L-Category	Quad & ATVs	–
Motorcycles	L-Category	Mopeds 2-stroke $< 50 \text{ cm}^3$	Cylinder capacity $\leq 50 \text{ cm}^3$ 2-stroke engine
		Mopeds 4-stroke $< 50 \text{ cm}^3$	Cylinder capacity $\leq 50 \text{ cm}^3$ 4-stroke engine
		Motorcycles 2-stroke $> 50 \text{ cm}^3$	Cylinder capacity $> 50 \text{ cm}^3$ 2-stroke engine
		Motorcycles 4-stroke $< 250 \text{ cm}^3$	50 cc $<$ Cylinder capacity $\leq 250 \text{ cm}^3$ 4-stroke engine
		Motorcycles 4-stroke 250 - 750 cm^3	250 cc $<$ Cylinder capacity $\leq 750 \text{ cm}^3$ 4-stroke engine
		Motorcycles 4-stroke $> 750 \text{ cm}^3$	Cylinder capacity $> 750 \text{ cm}^3$ 4-stroke engine
Minibus, bus	Buses	Urban CNG Buses	Vehicular natural gas
		Urban Buses Diesel Hybrid	Diesel electric
		Urban Buses Midi $\leq 15 \text{ T}$	Diesel Cylinder capacity $< 3000 \text{ cm}^3$
		Urban Buses Standard 15–18 T	Diesel 3,000 $\text{cm}^3 <$ Cylinder capacity $\leq 6000 \text{ cm}^3$
		Coaches Standard $\leq 18 \text{ T}$	Diesel Cylinder capacity $> 6000 \text{ cm}^3$ Less than or equal to 3 axes
		Coaches Articulated $> 18 \text{ T}$	Diesel Cylinder capacity $> 6000 \text{ cm}^3$ More than 3 axes
Trucks	Light commercial vehicles	N1-III	Petrol Cylinder capacity $< 2200 \text{ cm}^3$
Trucks, Rigid Truck	Heavy Duty Vehicles (Diesel)	Rigid $\leq 7.5 \text{ T}$	2 200 $\text{cm}^3 \leq$ Cylinder capacity $\leq 5000 \text{ cm}^3$
		Rigid 7.5 - 12 T	5 001 $\text{cm}^3 \leq$ Cylinder capacity $\leq 6400 \text{ cm}^3$
		Rigid 12 - 14 T	6 401 $\text{cm}^3 \leq$ Cylinder capacity $\leq 7300 \text{ cm}^3$
		Rigid 14 - 20 T	7 301 $\text{cm}^3 \leq$ Cylinder capacity $\leq 8200 \text{ cm}^3$
		Rigid 20 - 26 T	8 201 $\text{cm}^3 \leq$ Cylinder capacity $\leq 11000 \text{ cm}^3$
		Rigid 26 - 28 T	Cylinder capacity $\geq 11001 \text{ cm}^3$
Tractor-Trailer	Heavy Duty Vehicles (Diesel)	Articulated 28 - 34 T	Cylinder capacity $\leq 14000 \text{ cm}^3$
		Articulated 50 - 60 T	Cylinder capacity $\geq 14001 \text{ cm}^3$

Source: Authors

Application of Graphene Oxide Nanoparticles to Cementitious Composites to Mitigate the Effects of Attacks by Aggressive Agents

Aplicación de nanopartículas de óxido de grafeno en compuestos cementosos para mitigar los efectos del ataque de agentes agresivos

S. Castro-Lopes ¹, Barbara Simões ², Sergio Peres ³, V. D. Rodrigues ⁴, T. F. A. Santos ⁵, and Romildo Berenguer ⁶

ABSTRACT

This study investigates the potential of adding 0.03% graphene oxide (GO) nanoparticles into cementitious composites, assessing their mechanical performance and resistance to carbonation. The results indicate a 26% increase in compressive strength, from 23.92 to 32.42 MPa, with tensile strength increasing by an average of 1.14 MPa. Furthermore, the composite exhibits 14% lower capillary water absorption, enhancing resistance to moisture ingress. In terms of carbonation resistance, the addition of GO reduces the carbonation front by approximately 46% compared to the reference samples. Service life estimations suggest that, under equivalent exposure conditions, a structure incorporating GO would experience a degradation equivalent to five years, whereas a conventional structure would degrade over 20 years. These findings highlight the effectiveness of GO nanoparticles in enhancing both the mechanical properties and durability of cementitious materials.

Keywords: nanotechnology, performance, mechanical properties, additions, graphene oxide

RESUMEN

Este estudio investiga el potencial de incluir una concentración de 0.03 % de nanopartículas de óxido de grafeno (GO) en compuestos cementosos, evaluando su rendimiento mecánico y resistencia a la carbonatación. Los resultados indican un aumento del 26 % en la resistencia a la compresión, pasando de 23.92 a 32.42 MPa, con un aumento promedio de 1.14 MPa en la resistencia a la tracción. Además, el compuesto mostró una absorción de agua capilar un 14 % menor, mejorando la resistencia a la penetración de humedad. En términos de resistencia a la carbonatación, la adición de GO redujo el frente de carbonatación en aproximadamente un 46 % en comparación con las muestras de referencia. Las estimaciones de vida útil sugieren que, bajo condiciones de exposición equivalentes, una estructura que incorpore GO experimentaría una degradación equivalente a cinco años, mientras que una estructura convencional se degradaría en más de 20 años. Estos resultados destacan la efectividad de las nanopartículas de GO para mejorar tanto las propiedades mecánicas como la durabilidad de los materiales cementosos.

Palabras clave: nanotecnología, desempeño, propiedades mecánicas, adiciones, óxido de grafeno

Received: November 17th, 2024

Accepted: March 6th, 2025

Introduction

The field of civil construction is one of the most dynamic sectors in Brazil's economy, playing a fundamental role in providing services and materials for the industry [1]. However, despite its positive contributions, this sector also has a significant environmental impact. [2] state that the construction industry consumes over 30% of the natural resources, and that 75% of the waste it generates is neither recycled nor reused. In light of these challenges, there has been increasing interest from both the industry and government entities in developing construction materials with higher durability and better performance to minimize their environmental footprint [3]–[5]. Among the strategies for sustainable development, recent studies have investigated alternative binders, e.g., partially replacing cement with gypsum [6], as well as supplementary cementitious materials like slag cement and rice husk ash to reduce the demand

¹ Civil engineer, Santo Agostinho University Center, UNIFSA, Brazil. PhD in Materials Science, Universidade Federal de Pernambuco, Recife (PE), Brazil. Affiliation: Full professor, Instituto Federal do Pará, IFPA, Brazil. Email: samuel.jonatas@ufpe.br

² Civil engineer, Universidade de Pernambuco, Escola Politécnica de Pernambuco, Recife (PE), Brazil. Affiliation: Young researcher, Universidade de Pernambuco, Brazil. E-mail: bls1@poli.br

³ Mechanical engineer, Universidade Federal de Pernambuco, Recife (PE), Brazil. PhD in Mechanical Engineering, University of Florida, United States of America. Affiliation: Associate professor, Universidade de Pernambuco, Brazil. Email: speres@poli.br

⁴ Biotecnologia, Universidade Presidente Antônio Carlos, Brazil. PhD in Genetics and Molecular Biology, Universidade Estadual de Campinas, Brazil. Affiliation: Researcher, Universidade Federal de Pernambuco, Recife (PE), Brazil. Email: vivibiotec@gmail.com / viviane.drumonnd@ufpe.br

⁵ Graduate in Physics, Universidade Federal de Ouro Preto, UFOP, Brazil. PhD in Mechanical Engineering, Universidade Estadual de Campinas, Brazil. Affiliation: Full professor, Universidade Federal de Pernambuco, Recife (PE), Brazil. Email: tiago.felipe@ufpe.br

⁶ Civil engineer, Universidade Federal de Pernambuco, Recife (PE), Brazil. PhD in Materials Science, Universidade Federal de Pernambuco, Recife (PE), Brazil. Affiliation: Full professor, Universidade de Pernambuco, Escola Politécnica de Pernambuco, Recife (PE), Brazil. Email: romildo.berenguer@upe.br



Attribution 4.0 International (CC BY 4.0) Share - Adapt

for Portland cement [7]. Additionally, the incorporation of recycled concrete aggregate (RCA) in precast concrete has been highlighted as a potential approach to enhancing sustainability in the construction sector [8].

Nanotechnology, introduced as a scientific concept by Norio Taniguchi in 1974, has brought significant advancements across multiple fields, including civil engineering [9]. A simplified definition of nanotechnology involves the exploration of systems at the nanoscale, down to atomic precision, enabling the creation of materials with tailored structures that exhibit unique properties [10]. Over time, the integration of nanotechnology in civil engineering has proven beneficial for enhancing material performance, particularly in improving mechanical strength and durability while reducing the exploitation of natural resources [11]. Among the innovative materials being explored, shape memory alloy fibers such as Ni-Ti have demonstrated the ability to enhance crack recovery in engineered cementitious composites, improving mechanical performance and extending service life [12]. Furthermore, the optimization of cementitious systems with alumina-rich materials has been investigated to improve workability, hydration kinetics, and early-age mechanical performance [13]. Concrete durability is a crucial factor in ensuring that structures perform as intended throughout their service life. The reduction in durability can be attributed to both external environmental factors and intrinsic material deficiencies [14]. Additionally, permeability plays a critical role in determining the extent to which fluids and gases can penetrate concrete, leading to material degradation over time [15].

In response to these challenges, research has focused on applying nanomaterials to concrete and mortar in order to enhance their durability, with carbon-based nanomaterials emerging as promising additives [16]. Among these, graphene oxide (GO) has gained considerable attention in civil engineering, given its high specific surface area, excellent dispersion in water [17], [18], and superior mechanical properties, including its high tensile strength and elastic modulus [19]. Some studies have also indicated that GO may improve the resistance of cementitious materials to aggressive agents, although further research is needed to fully understand these effects.

In this vein, this study investigates the effects of GO nanoparticles on cementitious composites. To this effect, the synthesis of GO nanoparticles is detailed in the supplementary material. Therefore, the research focuses on three main objectives: (i) to assess GO's enhancement of tensile and compressive strength; (ii) to evaluate its reduction in capillary absorption; and (iii) to examine the improved resistance to carbonation attack that it provides. Based on the existing literature, it is hypothesized that the incorporation of GO leads to improved mechanical properties and increased durability of cementitious composites, making them more resistant to environmental factors such as moisture ingress and carbonation.

The document is organized as follows. The next section outlines the materials and methods used to prepare the cementitious composites and conducting the experiments. Subsequently, the results of the mechanical tests, permeability assessments, and carbonation resistance studies are presented, and the final section discusses the conclusions of this study, which demonstrate the significant improvements in the mechanical performance and durability of cementitious composites through the incorporation of GO nanoparticles at a 0.03% concentration.

Materials and methods

This section characterizes the fine aggregate (sand) and Portland CPV-ARI cement used in this research. The fine aggregate was characterized according to [20]. For the chemical and physical characterization of the cement, a literature review was conducted to obtain the necessary data. The physical characterization tests for CPV-ARI cement included unit mass, specific gravity, specific surface area (Blaine), particle size distribution, and loss on ignition. To quantitatively analyze the cement's chemical composition, X-ray fluorescence spectroscopy (FRX) was employed. The physical characterization tests were conducted following established standards, and the results are presented in Tables I and II. The loss on ignition value, expressed as a mass percentage, complies with the provisions of [21], as it remains below the 6.5% limit [22].

Table I. Physical characterization, particle distribution, and loss on ignition of CPV-ARI cement

Quantity	CPV-ARI cement
Unit mass (kg/dm ³)	1.00
Specific gravity (kg/dm ³)	3.17
Specific surface area (m ² /g) (electronic Blaine)	0.452
Particle distribution D10 µm	0.94
Particle distribution D50 µm	5.9
Particle distribution D90 µm	29.15
Loss ignition % (LI)	6.37

Source: Authors

Table II. Properties of the reduced graphene oxide (GO) used

Quantity	Graphene oxide (GO)
pH	2.0-5.0
Tap density (g/cm ³)	<0.1
Specific surface area (m ² /g)	180-280
H ₂ O	<4.0
Particle size D50 µm	<10.0
C (wt%)	73 ± 5
O (wt%)	15 ± 3
S (wt%)	<0.4

Source: Authors

Nevertheless, the results presented in Table III confirm that the Portland cement complies with the percentage limits for magnesium oxide ($\text{MgO} \leq 6.5\%$) and sulfur trioxide ($\text{SO}_3 \leq 4.5\%$) established by [21].

Table III. Chemical composition, expressed as a percentage, as obtained via X-ray fluorescence (FRX)

Quantity	CPV ARI Cement
SiO_2	19.40
Al_2O_3	3.80
CaO	60.64
Fe_2O_3	2.23
TiO_2	0.19
MgO	5.19
K_2O	0.90
SO_3	4.23
Na_2O	0.45
P_2O_5	0.21
MnO	0.05
Other oxides	0.03

Source: Authors

A sand granulometry analysis was conducted, as it is a crucial indicator of the aggregate's physical and chemical properties, influencing both compaction and resistance to mechanical stresses [20]. The sand used in this study was oven-dried at $105 \pm 5^\circ\text{C}$ and divided into two 500 g samples, as specified in [20]. Two tests (A and B) were performed. A set of standard sieves (4.8, 2.38, 1.19, 0.59, 0.3, and 0.15 mm) was used to carry out the analysis. The sieve set was subjected to agitation for 10 minutes in a mechanical sieving machine, as shown in Fig. S1 (supplementary material), which allowed determining the maximum characteristic dimension and the sand fineness modulus. According to [20], these correspond, respectively, to (i) the particle size at which the accumulated retained percentage of the aggregate is equal to or less than 5%, and (ii) the sum of the accumulated retained percentages in the standard sieves, divided by 100. We observed that the variation limit of $\pm 4\%$ in the retained percentage between tests A and B was respected, as established by [20]. Furthermore, the maximum characteristic dimension (MD) and the sand fineness modulus (FM) were found to be 4.75 and 2.00 mm, respectively. The granulometric composition of the sand used in this research is presented in Fig. S2 (supplementary material) and Table IV.

GO is primarily synthesized through the oxidation of graphite, with the modified Hummers method being the most widely used approach [23], [24]. Initially, the Hummers method employed high proportions of reactive substances. Over time, modifications to this process were introduced, leading to what is now referred to as the *modified Hummers method* [25]. This approach involves the oxidation of graphite using potassium permanganate and sodium nitrate in the presence of concentrated sulfuric acid [23], [26]. According to [27],

[28], this is the most stable and reliable technique for producing two-dimensional GO sheets. It also offers two key advantages: the ability to produce large quantities of GO and the relatively low cost of equipment and reagents [29].

Table IV. Granulometric composition of the sand

Sieve (mm)	Retained mass (g) Test A	Retained mass (g) Test B	Retained (%) Test A	Retained (%) Test B	Variation ($\pm 4\%$)	Mean (%)	Accumulated (%)
4.75	3.31	2.72	0.7	0.5	0.2	0.6	0.6
2.38	22.66	21.76	4.5	4.4	0.2	4.4	5.1
1.19	45.09	44.04	9.0	8.8	0.2	8.9	14.0
0.59	97.40	94.06	19.5	18.8	0.7	19.2	33.1
0.3	149.09	144.30	29.9	28.9	1.0	29.4	62.5
0.15	111.80	112.54	22.4	22.5	0.1	22.5	85.0
Pan	70.07	80.04	14.0	16.0	2.0	15.0	100.0
Total	499.42	499.46					

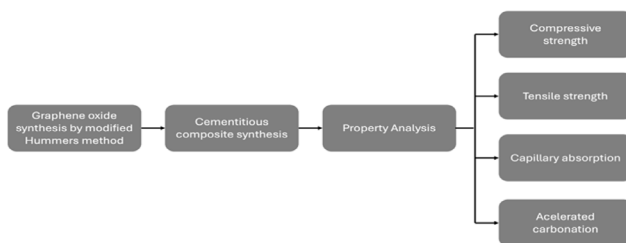
Source: Authors

In this study, GO nanoparticles were synthesized using the modified Hummers method, as illustrated in Figs. S3a to S3f (supplementary material) [30]. The process began with the dispersion of 1 g of graphite in a solution of NaNO_3 and H_2SO_4 , followed by continuous stirring for one hour. Subsequently, 3 g of KMnO_4 were gradually added at a rate of 1 g every 15 minutes while maintaining the solution in an ice bath to prevent the temperature from exceeding 20°C , thereby mitigating the risk of overheating and potential explosions. The mixture was then vigorously stirred before the slow addition of 100 mL of heated distilled water to dilute the solution. Afterwards, 3 mL of a 30% H_2O_2 solution were introduced to ensure the complete reaction of the KMnO_4 , and the mixture was left to settle. Finally, the suspension was washed with HCl and deionized water, followed by filtration and drying to obtain the final GO product [26].

For mortar preparation, 0.1872 g of GO were manually pre-diluted, followed by dispersion using an electromagnetic stirrer for one minute to ensure a uniform distribution before application. The execution of this process, as illustrated in Fig. S4 (supplementary material), follows the procedure described by [28], which has been recognized as an effective method for achieving homogeneous nanoparticle dispersion.

For the preparation of test specimens, samples were produced with and without the addition of GO nanoparticles. These were designated as GO 0.03% and reference samples, respectively, for subsequent analysis and performance comparison. The specimens were molded while following the guidelines of [31], with a mix proportion of 1:3:0.48 (cement:sand:water) by mass. This corresponds to 624 g of cement, 1872 g of sand, and 300 mL of water, resulting in a water-to-cement ratio (w/c) of 0.48. If needed, the water/cement ratio can be determined by measuring the concrete's dielectric constant [32], [33].

The quantity of GO nanoparticles used in each test specimen was 0.03% (0.1872 g) by weight of cement. To facilitate the incorporation of GO into the mortar, a portion of the mix water was removed and used for diluting the GO, which was agitated for a few minutes. The diluted GO was then combined with the remaining water, followed by the addition of cement. The mixing procedures were carried out using a mechanical mortar mixer, as specified in [31]. The mortar was prepared in four layers of approximately equal height, with manual compaction achieved by applying 30 blows per layer using a tamper after applying a release agent to the molds. After molding, the test specimens were left in the molds for 24 hours. A total of 12 cylindrical test specimens were produced, each measuring 5 x 10 cm, six with GO and six without, as shown in Fig. S5 (supplementary material). Flowchart 1 shows the steps of this research.



Flowchart 1. Research stages and development

Source: Authors

Compressive test

After a wet curing period of 28 days, six cylindrical test specimens (three with the addition of GO nanoparticles and three without) were subjected to an axial compression test according to [34]. Before testing, the top surface of the specimens was smoothed using a machine called a *rectifier*. The purpose of the rectifier is to create a smooth and even surface, ensuring that the load applied by the testing machine is distributed uniformly across the material.

Capillary water absorption test

The capillary water absorption test was conducted in accordance with the procedures established by [35]. Water absorption tests were performed on samples both with and without GO in their matrix to analyze potential differences in water absorption capacity between the groups. After 28 days of wet curing, the masses of the test specimens were measured, and they were then placed in an oven at a temperature of 105 ± 5 °C until a constant mass was achieved. Once the steady state had been reached, the dry mass (m_d) of each specimen was determined.

In the second step, the test specimens were positioned on a support and submerged in a container of water, with the water level maintained at 5 ± 1 mm above their lower face. The saturated mass (m_s) of each specimen was recorded at specific time intervals following initial water contact (3, 6, 14, 48, and 72 h).

Tensile strength and accelerated carbonation resistance tests

Tensile strength tests were conducted using the Brazilian method, adhering to all the conditions outlined in [36]. During the diametral compression test for tensile strength, the load required to induce failure in the test specimen was evaluated based on the force applied in the direction of the diameter.

For the accelerated carbonation resistance test, six test specimens without additions (reference samples) and six specimens with GO were prepared. The samples were placed in an oven until a constant mass was achieved and then transferred to a carbonation chamber. The chamber used in this research was a CARON refrigerated CO₂ incubator featuring digital automatic temperature control within ± 1 °C, relative humidity control within $\pm 2\%$, and regulated CO₂ levels.

At predetermined analysis intervals, 30 days after the specimens were placed in the carbonation chamber, each specimen was fractured into three parts. At each evaluation, the carbonation front was verified in slices measuring approximately 3.3 cm of the test specimen. An EMIC press with a maximum capacity of 100 kN was used for specimen rupture, as shown in Fig. 1.



a



b

Figure 1. a) Carbonation tests with reference samples and samples with GO in the first 30 days, b) carbonation test specimens after 90 days
Source: Authors

Results and discussion

Axial compressive strength

For the compressive strength test, 12 specimens were analyzed, six with GO nanoparticles and six without, serving as reference specimens. All specimens were tested after 28 days. The average compressive strength results are presented in Fig. 2.

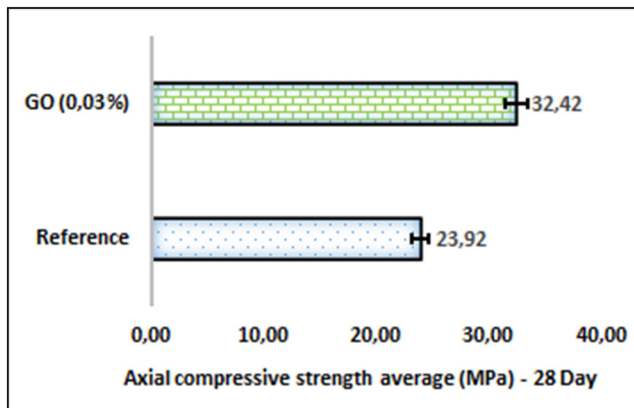


Figure 2. Average compressive strength of both mortar types after 28 days

Source: Authors

An increase in compressive strength was observed in the mortar with the addition of GO in its matrix. The average strength, when compared to the reference mortar, showed an increase of approximately 26%, rising from 23.92 to 32.42 MPa. These results corroborate previous research works that also demonstrated improvements in compressive strength with the addition of 0.03% GO in the mortar. This can be attributed to the excellent dispersion of GO in the cement matrix, the formation of covalent bonds between GO and the cement hydration products, the degree of hydration, and the effect of GO sheets filling the pores of the mortar [37]–[39].

Similarly, machine learning-based approaches have been employed to predict the compressive strength of concrete incorporating alternative materials. [40] proposed a data-driven hybrid machine learning framework to estimate the compressive strength of concrete with rice husk ash (RHA), contributing to mitigating the environmental impact of the cement industry. Additionally, experimental studies reinforce the effectiveness of material modifications in enhancing mechanical properties. [41] demonstrated that adding low concentrations of steel fiber to recycled aggregate concrete can improve its shear strength by more than 51%.

Capillary water absorption

The water absorption capacity of a material is one of the most important transport mechanisms that can affect the

durability of cementitious composites [42]. Fig. 3 presents the comparative results regarding the average water absorption by capillary, expressed as a percentage, for the reference mortar vs. the one with GO addition in each weighing time established by the standard.

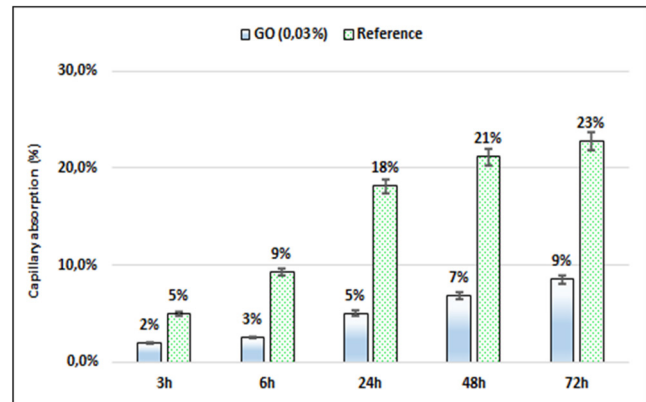


Figure 3. Comparison of results regarding the average water absorption by capillary

Source: Authors

Fig. 3 shows that the test specimens with the addition of GO exhibited superior performance, characterized by lower capillary absorption compared to the reference samples. In all measurements, the reference samples absorbed more water than those with GO. When analyzing the average absorption for both mortar types, it became evident that, at the final measurement time, the reference samples had a final absorption rate 14% higher than that of the samples containing GO. This behavior may be related to the hydrophobicity of GO, which stems from the presence of oxygenated functional groups on its surface [43]. The hydrophobicity of GO can ensure lower water absorption in samples with nanoparticles, which occasionally ensures superior performance during capillary absorption testing. The easy dispersion of GO in water, due to its oxygenated functional groups, facilitates its incorporation into the cement matrix [16]. Despite this, nanoparticle agglomeration remains a challenge, as van der Waals forces can lead to clustering [44]. However, in this study, it was observed that GO, due to its hydrophobicity, was able to reduce water absorption. [45] demonstrated that the incorporation of colloidal nanosilica (CNS) into concrete can reduce its water absorption. This effect is attributed to enhanced pore filling and the pozzolanic reaction facilitated by nanosilica. Furthermore, the presence of CNS improved cement hydration and produced a greater quantity of CSH gel, thereby enhancing the mechanical properties of the concrete.

Tensile strength and accelerated carbonation resistance

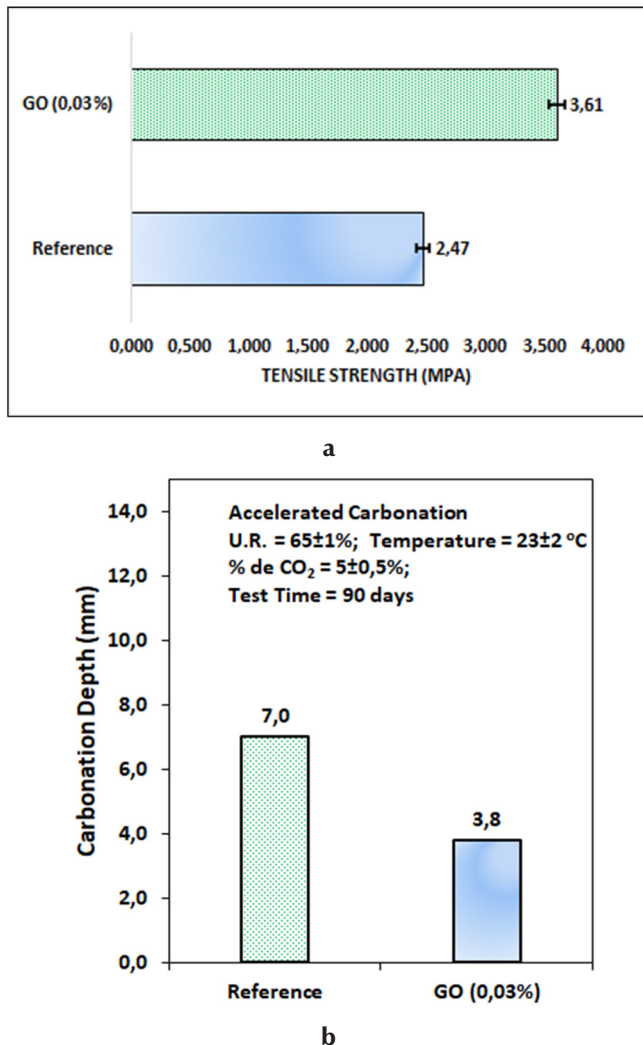


Figure 4. a) Average tensile strength of both mortar types, b) accelerated carbonation (90 days)
Source: Authors

Fig. 4a illustrates the tensile strength performance of both the reference mortar and the GO-enhanced sample. The reference mortar exhibited a tensile strength of approximately 2.47 MPa, whereas the GO-modified sample reached 3.61 MPa, representing a 46% increase (1.14 MPa). This enhancement can be attributed to improvements in the microstructure of the cementitious matrix; GO facilitates stronger bonding between cement hydration products, resulting in a denser structure with fewer internal defects. Additionally, GO layers act as bridging agents, effectively distributing the applied stresses. Similar findings have been reported for mortars incorporating GO [18], [26], [37]–[39]. [46] demonstrated that integrating hybrid fibers into self-compacting concrete enhances mechanical properties and durability after exposure to high temperatures and freeze-thaw cycles by mitigating crack formation and propagation. Conversely, the incorporation of expanded polystyrene (EPS) spheres reduces the

mechanical strength of concrete – the low density of EPS leads to an overall decline in mechanical performance [47]. In another study, [48] demonstrated that incorporating EPS spheres, silica fume, and superplasticizer into concrete can effectively reduce its weight while significantly enhancing the shear capacity of two-way slabs. [49] proposed a novel cement-based anchoring material and demonstrated that an optimal combination of fly ash (15%), blowing agent (3%), water-reducing agent (0.2%), accelerator (1%), and Huaqiansu (1%) enhances bond strength and improves tensile strength.

Fig. 4b presents the results of the carbonation resistance test. The addition of GO reduced the carbonation front by approximately 46% compared to the reference samples, providing a basis for further analysis regarding service life calculations. These findings suggest that the addition of GO nanoparticles improves the durability of the structure by mitigating carbonation.

The addition of minerals generally improves concrete properties, including durability. However, carbonation can have a negative impact, leading to unfavorable outcomes [50]. Research has shown that the use of fly ash, a medium reactivity pozzolan, significantly increases the rate of carbonation front progression [50]–[53].

On the other hand, the incorporation of supplementary pozzolans like active silica, metakaolin, and rice husk ash, as well as unconventional by-products such as submerged arc welding (SAW) slag, has been shown to improve resistance to carbonation, helping to slow down its advancement [54]–[56]. These pozzolanic reactions refine the pore structure and reduce CO₂ permeability, a result consistent with the findings of [22], [57], [58].

Through Eq. (1), it is possible to evaluate the concrete structure's time of degradation for the two types analyzed, which would correspond to the carbonation front shown in Fig. 4b [58].

$$x = k\sqrt{t} \quad (1)$$

In the equation, x represents the depth of the carbonation front, k is the constant ranging from 0.1 to 1.0 cm²/year, and t denotes the degradation time in years. In this research, with a constant of 0.5 cm²/year, we found that, regarding the reference samples, the structure would experience a degradation equivalent to that of a two-year-old structure. In contrast, the structure with 0.03% GO would only degrade as if it were six months old under the same conditions. Thus, considering a degradation lifespan of 20 years for the reference samples, structures containing GO nanoparticles (0.03%) would only experience a degradation equivalent to five years.

The results suggest that the addition of GO nanoparticles can extend the service life and durability of structures by approximately four times compared to conventional

structures. This indicates a potential for more sustainable construction practices and environmentally friendly building solutions.

[59] conducted a study simulating the incorporation of cellulose nanofibers from agricultural waste as a sustainable cement replacement for producing high-performance concrete. The results indicate that high-performance concrete enhanced with cellulose nanofibers has superior mechanical properties in comparison with conventional high-performance concrete, including improved stress, strain, and equivalent deformation under load. In addition, [60] developed a silica-based superhydrophobic coating to enhance the durability and safety of cementitious surfaces in road construction.

Conclusions

This study demonstrates the significant improvements in both mechanical performance and durability of cementitious composites through the incorporation of graphene oxide nanoparticles at a 0.03% concentration. The results indicate a 26% increase in compressive strength, accompanied by enhanced tensile strength and a reduction in water absorption. Moreover, the inclusion of GO notably improved the resistance to carbonation, with a reduction of approximately 46% in carbonation front progression compared to the reference samples. These findings underscore the potential of GO to enhance the mechanical properties and long-term durability of concrete, offering promising implications for its use in construction materials exposed to aggressive environmental conditions.

Acknowledgements

The authors are grateful for the financial support of the Brazilian Agencies FACEPE, CAPES, and National Council for Scientific and Technological Development (CNPq). The authors also thank Universidade de Pernambuco (POLI/UPE, Brazil), Universidade Federal da Paraíba (UFPB, Brazil), and Universidade Federal de Pernambuco (UFPE, Brazil) for making the equipment available.

CRedit author statement

BS and SCL were in charge of conceptualization, and SP collected the data, developed the workflow, and performed assessments. VDR and TFSA provided critical feedback, led the drafting process, and wrote the main part of the manuscript. RB contributed with supervision and validation, to which all authors contributed.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] T. A. Frej and L. H. Alencar, "Fatores de sucesso no gerenciamento de múltiplos projetos na construção civil em Recife," *Production*, vol. 20, no. 3, pp. 322-334, Sept. 2010. <https://doi.org/10.1590/S0103-65132010005000043>
- [2] C. K. Purchase *et al.*, "Circular economy of construction and demolition waste: A literature review on lessons, challenges, and benefits," *Materials*, vol. 15, no. 1, art. 76, 2021. <https://doi.org/10.3390/ma15010076>
- [3] A. A. Firoozi, A. A. Firoozi, D. O. Oyejobi, S. Avudaiappan, and E. S. Flores, "Emerging trends in sustainable building materials: Technological innovations, enhanced performance, and future directions," *Results Eng.*, vol. 24, art. 103521, Dec. 2024. <https://doi.org/10.1016/j.rineng.2024.103521>
- [4] N. Makul and G. Suaiam, "New insights into the early age time dependent dielectric evolution of pozzolan modified eco efficient cement pastes within a frequency range of 200 MHz to 6500 MHz: Experiments and statistical modeling," *Eng. Sci.*, vol. 31, art. 1170, 2024. <https://doi.org/10.30919/es1170>
- [5] M. P. Rosa *et al.*, "Viabilidade na reutilização da ardósia em substituição do agregado graúdo na dosagem do concreto," *Braz. J. Dev.*, vol. 6, pp. 936-948, 2020. <https://doi.org/10.34117/bjdv6n1-065>
- [6] N. Gerges *et al.*, "The novelty of partially replacing cement with Gypsum: Optimum mix design and structural applications," *ES Mater. Manuf.*, 2024. [Online]. Available: <https://doi.org/10.30919/esmm1336>
- [7] P. Hiremath *et al.*, "Investigating the mechanical properties, durability, and environmental impact of partial cement substitution with slag cement and rice husk ash for sustainable concrete production," *ES Food Agrofor.*, vol. 18, art. 1267, 2024. <https://doi.org/10.30919/esfaf1267>
- [8] G. Sua-iam and N. Makul, "Potential future direction of the sustainable production of precast concrete with recycled concrete aggregate: A critical review," *Eng. Sci.*, vol. 28, art. 1075, 2024. <https://doi.org/10.30919/es1075>
- [9] N. Taniguchi, "On the basic concept of 'nano-technology'," in *Proc. Intl. Conf. Prod. Eng.*, Tokyo, Japan, 1974, pp. 18-23.
- [10] S.A. Filho and B.P. Backx, "Nanotecnologia e seus impactos na sociedade," *R. Tecnol. Soc.*, vol. 16, no. 40, pp. 1-15, Apr./Jun. 2020. <https://doi.org/10.3895/rt.s.v16n40.9870>
- [11] M. Feizbahr and P. Pourzanjani, "Nanotechnology in construction: Innovations, applications, and impacts authors," *J. Civ. Eng. Res.*, vol. 6, no. 1, pp. 35-4, 2024. <https://doi.org/10.61186/JCER.6.1.35>
- [12] M. Song, J. Wang, L. Yuan, C. Luan, and Z. Zhou, "Investigation on crack recovery behavior of engineered cementitious composite (ECC) incorporated memory alloy fiber at low temperature," *ES Mater. Manuf.*, vol. 17, pp. 23-33, 2022. <https://doi.org/10.30919/esmm5f662>
- [13] S. Al-Shereiqi, H. A. Abdel-Gawwad, and M. S. Meddah, "Unveiling the influence of varied alumina sources on fresh properties of ordinary Portland cement mortar and concrete: A comprehensive review," *Eng. Sci.*, vol. 30, art. 1118, 2024. <https://doi.org/10.30919/es1118>

- [14] R. M. Ferreira, "Avaliação dos ensaios de durabilidade do betão," Master's thesis, Esc. Eng., Univ. do Minho, Braga, Portugal, 2000.
- [15] L. A. Araujo et al., "Concrete gas permeability: Implications for hydrogen storage applications," *Appl. Sci.*, vol. 14, no. 15, art. 6408, 2024. <https://doi.org/10.3390/app14156408>
- [16] A. Bagheri, E. Negahban, A. Asad, H.A. Abbasi, S.M. Raza, "Graphene oxide-incorporated cementitious composites: a thorough investigation," *Mater. Adv.*, vol. 3, pp. 9040-9051, 2022. <https://doi.org/10.1039/D2MA00169A>
- [17] S. Chuah, Z. Pan, J. G. Sanjayan, C. M. Wang, and W. H. Duan, "Nano reinforced cement and concrete composites and new perspective from graphene oxide," *Constr. Build. Mater.*, vol. 73, pp. 113-124, Dec. 2014. <https://doi.org/10.1016/j.conbuildmat.2014.09.040>
- [18] Y. Y. Wu, L. Que, Z. Cui, and P. Lambert, "Physical properties of concrete containing graphene oxide nanosheets," *Materials*, vol. 12, no. 10, art. 1707, May. 2019. <https://doi.org/10.3390/ma12101707>
- [19] M. Frąc, W. Pichór, and P. Szołdra, "Cement composites with expanded graphite as resistance heating elements," *J. Compos. Mater.*, vol. 54, no. 25, pp. 3821-2831, 2020. <https://doi.org/10.1177/0021998320921510>
- [20] *Aggregates - Sieve analysis of fine and coarse aggregates*, NBR NM 248, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2003.
- [21] *Portland cement - Requirements*, NBR 16697, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2018.
- [22] R. Berenguer, A. P. B. Capraro, M. H. F. Medeiros, A. M. P. Carneiro, R. A. Oliveira, "Sugar cane bagasse ash as a partial substitute of Portland cement: Effect on mechanical properties and emission of carbon dioxide," *J. Environ. Chem. Eng.*, vol. 8, no. 2, art. 103655, Apr. 2020. <https://doi.org/10.1016/j.jece.2020.103655>
- [23] G. Dias, T. Cellet, M. Santos, C. Carvalho, and L. Malmonge, "A caracterização morfológica de óxido de grafeno preparados pelo método de Hummers modificado," *Rev. Tecnol.*, vol. 29, no. 1, pp. 199-216, 2020. <https://doi.org/10.4025/revtecnol.v29i1.51286>
- [24] S. Park and R. S. Ruoff, "Chemical methods for the production of graphenes," *Nature Nanotech.*, vol. 4, pp. 217-224, Apr. 2009. <https://doi.org/10.1038/nnano.2009.58>
- [25] J. E. D. Vieira Segundo and E. O. Vilar, "Grafeno: Uma revisão sobre propriedades, mecanismos de produção e potenciais aplicações em sistemas energéticos," *REMAP*, vol. 11, no. 2, pp. 54-57, 2016. <https://remap.revistas.ufcg.edu.br/index.php/remap/article/viewFile/493/387>
- [26] M. M. Mokhtar, S. A. Abo-El-Enein, M. Y. Hassaan, M. S. Morsy, and M. H. Khalil, "Mechanical performance, pore structure and micro-structural characteristics of graphene oxide nanoplatelets reinforced cement," *Constr. Build. Mater.*, vol. 138, pp. 333-339, May. 2017. <https://doi.org/10.1016/j.conbuildmat.2017.02.021>
- [27] S. Stankovich et al., "Graphene-based composite materials," *Nature*, vol. 442, pp. 282-286, Jul. 2006. <https://doi.org/10.1038/nature04969>
- [28] C. S. R. Indukuri and R. Nerella, "Enhanced transport properties of graphene oxide-based cement composite material," *J. Build. Eng.*, vol. 37, art. 102174, May. 2021. <https://doi.org/10.1016/j.jobbe.2021.102174>
- [29] R. Hack, C. Correia, R. Zanon, and S. Pezzin, "Characterization of graphene nanosheets obtained by a modified Hummer's method," *Matéria (Rio J.)*, vol. 23, no. 1, 2018. <https://doi.org/10.1590/S1517-707620170001.0324>
- [30] A. Alazmi, S. Rasul, S. P. Patole, and P. M. F. J. Costa, "Comparative study of synthesis and reduction methods for graphene oxide," *Polyhedron*, vol. 116, pp. 153-161, Sep. 2016. <https://doi.org/10.1016/j.poly.2016.04.044>
- [31] *Portland cement - Determination of compressive strength of cylindrical test specimens*, NBR 7215, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2019.
- [32] R. He, T. Nantung, J. Olek, and N. Lu, "Use of dielectric constant for determination of water-to-cement ratio (W/C) in plastic concrete: part 2: comparison determined W/C values by ground penetrating radar (GPR) and microwave oven drying measurements," *ES Mater. Manuf.*, vol. 22, art. 874, 2023. <https://doi.org/10.30919/esmm5f874>
- [33] R. He, T. Nantung, and J. Olek, N. Lu, "Use of dielectric constant for determination of water-to-cement ratio (W/C) in plastic concrete: part 1. Volumetric water content modeling," *ES Mater. Manuf.*, vol. 21, art. 866, 2023. <https://doi.org/10.30919/esmm5f866>
- [34] *Concrete - Compression test of cylindrical specimens*, NBR 5739, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2018.
- [35] *Mortar and hardened concrete - Determination of water absorption by capillarity*, NBR 9779, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2012.
- [36] *Concrete and mortar - Determination of the tension strength by diametrical compression of cylindrical test specimens*, NBR 7222, Brazilian Association of Technical Standards, Rio de Janeiro, Brazil, 2011.
- [37] K. Gong et al., "Reinforcing effects of graphene oxide on Portland cement paste," *J. Mater. Civ. Eng.*, vol. 27, no. 2, Jul. 2015. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0001125](https://doi.org/10.1061/(ASCE)MT.1943-5533.0001125)
- [38] S. Lv, Y. Ma, C. Qiu, T. Sun, J. Liu, and Q. Zhou, "Effect of graphene oxide nanosheets of microstructure and mechanical properties of cement composites," *Constr. Build. Mater.*, vol. 49, pp. 121-127, Dec. 2013. <https://doi.org/10.1016/j.conbuildmat.2013.08.022>
- [39] Y. Wang, J. Yang, and D. Ouyang, "Effect of graphene oxide on mechanical properties of cement mortar and its strengthening mechanism," *Materials*, vol. 12, no. 22, art. 3753, 2019. <https://doi.org/10.3390/ma12223753>
- [40] R. Odeh, R. Alawadi, A. Tarawneh, A. Alghossoon, R. Al-Mazaidh, and H. Amerah, "Estimating rice husk ash concrete compressive strength using hybrid machine learning methodology," *Eng. Sci.*, vol. 29, art. 1111, 2024. <https://doi.org/10.30919/es1111>
- [41] S. Leelatanon, T. Imjai, M. Setkit, R. Garcia, and C. K. Ma, "Shear strength of stirrup-free recycled aggregate concrete beams reinforced with steel fibers," *Eng. Sci.*, vol. 31, art. 1249, 2024. <https://doi.org/10.30919/es1249>

- [42] G. L. Golewski, "Assessing of water absorption on concrete composites containing fly ash up to 30 % in regards to structures completely immersed in water," *Case Stud. Const. Mater.*, vol. 19, art. e02337, 2023. <https://doi.org/10.1016/j.cscm.2023.e02337>
- [43] L. Qiu *et al.*, "Dispersing carbon nanotubes with graphene oxide in water and synergistic effects between graphene derivatives," *Chem. - Eur. J.*, vol. 16, no. 35, art. 10653-10658, Sep. 2010. <https://doi.org/10.1002/chem.201001771>
- [44] S. Parveen, S. Rana, and R. Fanguiero. "A review on nano-material dispersion, microstructure and mechanical properties of carbon nanotube and nanofiber-based cement composites," *J. Nanomater.*, vol. 2013, art. 710175, 2013. <https://doi.org/10.1155/2013/710175>
- [45] C. Huang, Y. F. Su, R. Tokpatayeva, T. Nantung, and N. Lu, "Investigation of internal curing efficacy of Portland cement concrete incorporated with colloidal nano silica," *ES Mater. Manuf.*, vol. 20, art. 798, 2022. <https://doi.org/10.30919/esmm5f798>
- [46] A. Ashteyat, A. Obaidat, and R. Qerba, "The rheological, mechanical, and durability behavior of self-compacted concrete (SCC) mixed with hybrid fibers after exposure to high temperatures and cycles of freezing and thawing," *Eng. Sci.*, vol. 27, art. 1042, 2023. <https://doi.org/10.30919/es1042>
- [47] M. Abdel-Jaber, R. Al-Nsour, N. Shatarat, H. Katkhuda, and H. Al-zu'bi, "Thermal effect on the flexural performance of lightweight reinforced concrete beams using expanded polystyrene beads and pozzolana aggregate," *Eng. Sci.*, vol. 27, art. 1029, 2023. <https://doi.org/10.30919/es1029>
- [48] M. Abdel-Jaber, N. Shatarat, H. Katkhuda, and M. Al-Najjar, "Punching shear capacity of polystyrene lightweight concrete two-way slabs," *Eng. Sci.*, vol. 31, art. 1216, 2024. <https://doi.org/10.30919/es1216>
- [49] S. Huang, X. Meng, G. Zhao, and Z. Liu, "Development and engineering application of cement-based full length anchoring material," *ES Mater. Manuf.*, vol. 23, art. 1051, 2023. <https://doi.org/10.30919/esmm1051>
- [50] J. Khunthongkeaw, S. Tangtermsirikul, and T. Leelawat, "A study on carbonation depth prediction for fly ash concrete," *Constr. Build. Mater.*, vol. 20, no. 9, pp. 744-753, Nov. 2006. <https://doi.org/10.1016/j.conbuildmat.2005.01.052>
- [51] S. Rukzon and P. Chindaprasirt, "Strength and carbonation model of rice husk ash cement mortar with different fineness," *J. Mater. Civ. Eng.*, vol. 22, no. 3, pp. 253-259, Feb. 2010. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2010\)22:3\(253\)](https://doi.org/10.1061/(ASCE)0899-1561(2010)22:3(253))
- [52] É. Sousa, A. Santana, M. Moura, J. M. P. Q. Delgado, and R. Berenguer, "Thermography applied to the adhesion phenomenon of mortars with additions of submerged arc welding (SAW) slag," *Buildings*, vol. 14, no. 9, art. 2960, 2024. <https://doi.org/10.3390/buildings14092960>
- [53] A. Younsi, P. Turcry, A. Ait-Mokhtar, and S. Staquet, "Accelerated carbonation of concrete with high content of mineral additions: Effect of interactions between hydration and drying," *Cem. Concr. Res.*, vol. 43, pp. 25-33, Jan. 2013. <https://doi.org/10.1016/j.cemconres.2012.10.008>
- [54] K. Turk, M. Karatas, and T. Gonen, "Effect of fly ash and silica fume on compressive strength, sorptivity and carbonation of SCC," *KSCE J. Civ. Eng.*, vol. 17, no. 1, pp. 202-209, Jan. 2013. <https://doi.org/10.1007/s12205-013-1680-3>
- [55] M. V. Moura, S. A. Paiva, V. D. Rodrigues, S. J. Lopes, T. F. Santos, and R. A. Berenguer, "Analysis of the influence of mechanical properties of cementitious composites using submerged arc welding (SAW) slag recycled aggregate," *GESEC*, vol. 15, no. 13, art. e3623, 2024. <https://doi.org/10.7769/gesec.v15i3.3623>
- [56] R. A. Berenguer, *et al.*, "Cement-based materials: Pozzolanic activities of mineral additions are compromised by the presence of reactive oxides," *J. Build. Eng.*, vol. 41, art. 102358, Sep. 2021. <https://doi.org/10.1016/j.jobbe.2021.102358>
- [57] R. A. Berenguer *et al.*, "Durability of concrete structures with sugar cane bagasse ash," *Adv. Mater. Sci. Eng.*, vol. 2020, art. 6907834, 2020. <https://doi.org/10.1155/2020/6907834>
- [58] M. H. F. Medeiros, J. J. O. Andrade, and P. Helene, "Durabilidade e vida útil das estruturas de concreto," in *Concreto: Ciência e Tecnologia*, ed. 1. São Paulo, Brazil: IBRA-CON, 2011, ch. 22.
- [59] M. D. Sreeja and N. Nalanth, "Exploring environmentally sustainable concrete: An analytical investigation on high performance concrete using cellulose nanofibers," *ES Ener. Environ.*, vol. 27, art. 1382, 2025. <https://doi.org/10.30919/esee1382>
- [60] N. Kydyrbay *et al.*, "Enhancing road durability and safety: a study on silica-based superhydrophobic coating for cement surfaces in road construction," *Eng. Sci.*, vol. 30, art. 1221, 2024. <https://doi.org/10.30919/es1221>

Supplementary Material

1 Materials and Methods

The sand granulometry analysis was conducted since it is a key indicator of the aggregate's physical and chemical properties, influencing compaction and mechanical resistance. The sand was oven-dried at $105 \pm 5^\circ\text{C}$ and divided into two 500 g samples. Two tests, A and B, were performed using a standard sieve set (4.8, 2.38, 1.19, 0.59, 0.3, and 0.15 mm).



Figure S1 - (a) Mechanical sieve and (b) - Granulometry test
Source: Authors

The sieve set was agitated for 10 minutes using a mechanical sieving machine, as shown in Fig. S1. This process determined the maximum particle size and the sand fineness modulus. The granulometric composition of the sand used in this study is presented in Fig. S2.

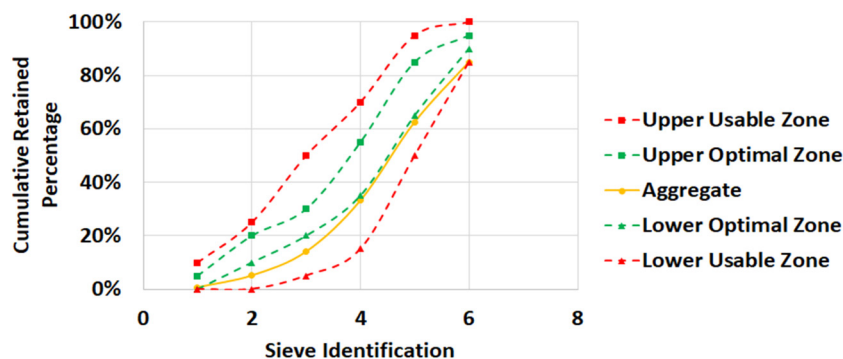


Figure S2 – Sand granulometric curve
Source: Authors

For this research, graphene oxide (GO) nanoparticles were synthesized using a modified Hummers method, as illustrated in Figs. S3(a)–S3(f) [1]. During the synthesis process, 1 g of graphite was dispersed in a jar with NaNO_3 and H_2SO_4 under continuous agitation for one hour. Then, 3 g of KMnO_4 were gradually added (1 g every 15 minutes) while maintaining an ice bath to keep the solution temperature below 20°C , preventing overheating and potential hazards. The mixture was then vigorously stirred before slowly adding 100 ml of heated distilled water for dilution. Subsequently, 3 ml of a 30% H_2O_2 solution were added to ensure the complete reaction of KMnO_4 , followed by a resting period. Finally, the mixture was washed with HCl and water, filtered, and dried to obtain the final GO product.

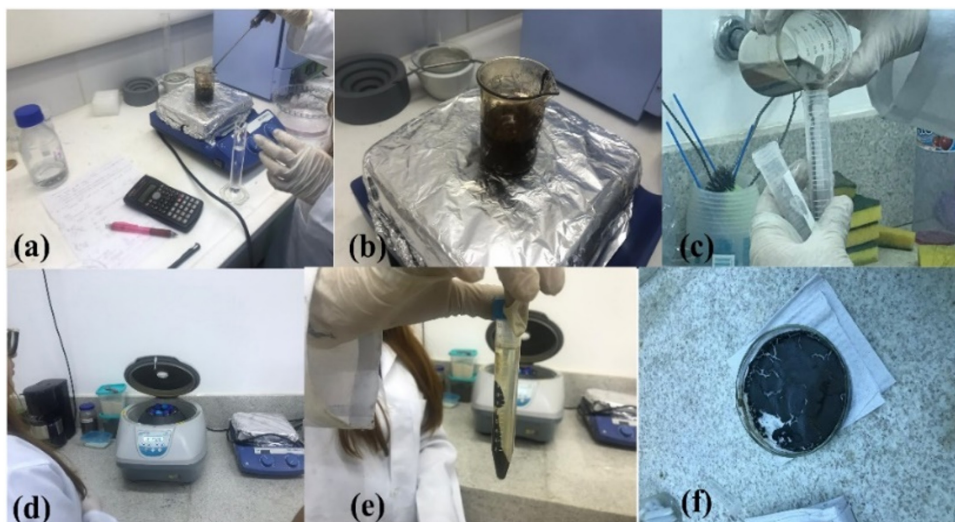


Figure S3 – GO nanoparticles obtained through the modified method of Hummers.

Source: Authors

The addition of 0.1872 g of GO to the mortar was carried out through manual pre-dilution, followed by uniform agitation for 1 minute using a magnetic stirrer before application. The execution of this process is shown in Fig. S4.



Figure S4 – Nanoparticles of GO dispersion

Source: Authors

To prepare the test specimens, reference samples and samples with 0.03% graphene oxide (GO) were produced. The mix ratio was 1:3:0.48 (cement:sand:water), with 624 g of cement, 1872 g of sand, and 300 ml of water, maintaining a water/cement ratio of 0.48. The GO nanoparticles (0.1872 g per specimen) were pre-dispersed in part of the mixing water before incorporation. Specimens were molded following [2] and mixed using a mechanical mixer.

The mortar molds were prepared in four layers of approximately equal height, with each layer compacted manually using a tamper with 30 blows after applying a release agent to the molds. Following molding, the test specimens remained in the molds for 24 hours. A total of 12 cylindrical specimens (5 × 10 cm) was produced, including six with GO and six without, as shown in Flg. S5.



Figure S5 – (a) Test specimens with GO addition, after demolding; and (b) Test specimens without GO addition, after demolding.

References

- [1] A. Alazmi, S. Rasul, S.P. Patole, P.M.F.J. Costa, "Comparative study of synthesis and reduction methods for graphene oxide," *Polyhedron*, vol. 116, pp. 153-161, Sep. 2016. <https://doi.org/10.1016/j.poly.2016.04.044>.
- [2] NBR 7215, /*Portland cement – Determination of compressive strength of cylindrical test specimens*/, NBR 7215, Brazilian Association of Technical Standards. Rio de Janeiro, Brazil, 2019.

Active Disturbance Rejection Control of a DC Brushed Motor Using Simulink and Raspberry Pi

Control por rechazo activo de perturbaciones de un motor DC con escobillas utilizando Simulink y Raspberry Pi

Mario E. González N. ¹, Oscar H. Sierra H. ², and Oscar I. Higuera M. ³

ABSTRACT

Active disturbance rejection control (ADRC) is a robust methodology that does not require precise knowledge of the plant. Developed in China by Professor Jingqing Han, it is based on PID control, state observers, and nonlinear functions. Brushed DC motors are known for their low cost and the noise they introduce into control circuits. This paper demonstrates that ADRC can effectively control low-power brushed DC motors using a general nonlinear model and Simulink for tuning. The model is simulated using parameters provided by the manufacturer. An ADRC developed and programmed by the authors in MATLAB is then integrated into the simulation. The controller is tuned, and its performance is verified. Subsequently, the ADRC is implemented on a Raspberry Pi 3 using MATLAB's support packages and methods developed by the authors. The controller is tested on a Faulhaber 2342L012CR DC motor (12 V/17 W). The results show that it is possible to control the position of the low-power brushed DC motor through simulation-based tuning. The interaction between Simulink and Raspberry Pi 3 enables an optimal control characterized by a fast response, a minimal steady-state error, and no perceptible overshoot. This implementation demonstrates that ADRC is a practical and efficient control method for brushed DC motors.

Keywords: ADRC, MATLAB, coreless brushed DC motor, automatic control, Raspberry Pi

RESUMEN

El control activo de rechazo de perturbaciones (ADRC) es un método robusto que no requiere un conocimiento preciso de la planta. Desarrollado en China por el profesor Jingqing Han, se basa en control PID, observadores de estado y funciones no lineales. Los motores DC con escobillas son conocidos por su bajo costo y el ruido que introducen en los circuitos de control. Este artículo demuestra que ADRC puede controlar eficazmente motores de DC de baja potencia con escobillas utilizando un modelo general no lineal y Simulink para su sintonización. Este modelo se simula utilizando parámetros proporcionados por el fabricante. Luego, se integra a la simulación un ADRC desarrollado y programado por los autores en MATLAB. Se sintoniza el controlador y se verifica su rendimiento. Posteriormente, el ADRC es implementado en una Raspberry Pi 3 utilizando los paquetes de apoyo de MATLAB y métodos desarrollados por los autores. El controlador es puesto a prueba en un motor de DC Faulhaber 2342L012CR (12 V/17 W). Los resultados muestran que es posible controlar la posición del motor de DC con escobillas de baja potencia mediante una sintonización basada en simulación. La interacción entre Simulink y Raspberry Pi 3 permite un sistema de control óptimo caracterizado por una respuesta rápida, un error mínimo en estado estacionario y ningún sobrepico perceptible. Esta implementación demuestra que el ADRC es un método de control práctico y eficiente para motores de DC con escobillas.

Palabras clave: ADRC, MATLAB, motor DC con escobillas sin núcleo, control automático, Raspberry Pi

Received: May 15th, 2024

Accepted: May 19th, 2025

Introduction

The concept and applications of feedback control have historically attracted significant interest, given the mathematical explanations to the ingenious mechanisms of control used on military applications during the second world war. Feedback control is one of the most used, and it is in constant change, getting better every day. There is always some kind of control device involved in every engineering process [1]. The literature on feedback control spans more than six decades of research. These developments were parallel to the mathematics applied in different academic disciplines as the formulation of

¹ Electronics engineer, Universidad Pedagógica y Tecnológica de Colombia, Colombia. MEng, Universidad Pedagógica y Tecnológica de Colombia, Colombia. Affiliation: Full profesor, Electromechanical Engineering, Universidad Pedagógica y Tecnológica de Colombia, Colombia. Email: marioeduardo.gonzalez@uptc.edu.co

² Electronics engineer, Universidad Pedagógica y Tecnológica de Colombia, Colombia. MEng, Universidad Pedagógica y Tecnológica de Colombia, Colombia. Affiliation: Full profesor, Electronics Engineering, Universidad Pedagógica y Tecnológica de Colombia, Colombia. Email: oscarhumberto.sierra@uptc.edu.co

³ Electronics engineer, Universidad Pedagógica y Tecnológica de Colombia, Colombia. MEng, Universidad Nacional de Colombia, PhD(c) Universidad Pedagógica y Tecnológica de Colombia, Colombia. Affiliation: Full profesor, Electronics Engineering, Universidad Pedagógica y Tecnológica de Colombia. Email: oscar.higuera@uptc.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

optimal control and the way to investigate it became more axiomatic and deductive [2].

Active disturbance rejection control (ADRC) allows controlling complex systems with almost zero knowledge of their dynamics. It represents a new paradigm in control theory since the concepts of *robustness* and *sensitivity to unmodeled dynamic perturbations* must be redefined. ADRC seeks to adapt the dynamic behavior of the system to the control strategy, rejecting external perturbations with the purpose of ensuring that the controller operates effectively in the presence of uncertainties [2]. ADRC has great potential for solving control problems in industrial processes characterized by elevated levels of uncertainty [3].

There is a considerable amount of bibliography related to ADRC, but there are few records of its implementation. This scarcity is due to its use of two nonlinear functions. A more popular variant called *linear active disturbance rejection control* (LADRC) is more commonly implemented and has a variety of tuning strategies available in the literature [4]. Despite this, the original ADRC remains a challenging controller to design and implement [5]. The design of an ADRC controller is particularly difficult due to its nonlinear nature, wherein empirical tuning through simulation before testing is the most widespread and recommended approach [6], [7].

The ADRC theory was developed in China by professor Jingqing Han. Based on proportional-integral derivative (PID) controllers [1], this theory has been simulated and implemented in power electronics [8], drone altitude control [9], robotics [10], tank level control [11], and permanent magnet synchronous motors [12], among others [13]. Since this is a relatively new theory, its use in practical applications has been very limited. Besides, it requires a considerable amount of time and resources for its implementation [14]. There are some implementations involving field programmable gate arrays (FPGAs) [15] and digital signal processors (DSPs) [16], but they are extremely complex and difficult to replicate.

Despite the complexity of ADRC implementations on platforms like DSPs and FPGAs [14]–[16], recent studies have shown that affordable embedded systems such as the Raspberry Pi offer a viable alternative for real-time control experiments [17]–[19]. The Raspberry Pi 3, with its quad-core ARM processor and compatibility with Simulink [20]–[22], enables the implementation of advanced controllers without the need for specialized hardware. Additionally, MATLAB/Simulink environments are widely used in control education and research due to their simulation capabilities and intuitive design tools [23].

The implementation of electronic control systems requires, at a minimum, the use of an analog-to-digital (A/D) converter, power systems, and processing systems, which include the controller, pulse width modulation

(PWM) signal generators, and other signal processing components [24]. These systems can be complex and hard to implement on platforms like FPGA, DAQ, DSPs, etc. [14].

Brushed DC motors are widely used in robotics and small appliances due to their low cost and high torque at low rpm. However, they are prone to noise from brushes and heat generation, which can alter their performance over time. Therefore, it is important to implement a dynamic controller that is capable of adapting to changing parameters in order to achieve more efficient systems [25].

In the literature on the implementation of controllers with DC brushed motors, significant works include [26], where a permanent-magnet brushed DC motor is controlled via an ADRC to compensate for load torque and variations in the rotor shaft, which do not require measuring the rotor shaft speed since they can be verified only through simulation; [27], where the speed of a brushless DC motor is controlled using a PID and artificial neural network-based controller; and [25], where a software is developed to control brushed DC motors using a PI controller tuned dynamically using fuzzy logic.

Works with more emphasis on theory or other types of motors include [6], which is characterized by its focus on controlling a moving mirror system and tuning ADRC parameters through improved snake optimization (I-SO). Moreover, [28] highlights the use of a genetic algorithm (GA) to optimize control in five-phase motors, enhancing the efficiency of a specific AC brushless motor. [29] proposes a set of rules for tuning ADRC in first-order plus delay time models, focusing on deep theoretical analysis without including practical implementation. [30] introduces a combined ADRC and iterative learning control (ILC) strategy to improve torque control in switched reluctance motors (SRM), offering innovative solutions in this specific field. [31] investigates the use of ADRC for speed control in DC brushless motors, comparing this technique against other control methods such as the PI and PID approaches.

Regarding ADRC and low-power motors, the authors of [32] focus on the angular position control of a servo motor, demonstrating the versatility of their method in various motor configurations. [33] provides a detailed analysis applied to a DC servo motor in real time, including the theory, practical implementation, and verification of linear ADRC in a real hardware environment. A related study is presented in [34], where a Raspberry Pi 4 is used together with Simulink (through external mode over Wi-Fi) to implement an ADRC controller for a DC motor. However, this work does not evaluate the real-world performance of the ADRC when tuned through simulation using a nonlinear DC motor model.

Brushed low-power DC motors are typically controlled using PID controllers. Nevertheless, there are no implementations

of ADRC for brushed low-power DC motors. The authors aim to address this gap by tuning an ADRC controller using a nonlinear DC motor model and Simulink. The tuned parameters will then be used to effectively control the motor.

The objectives of this paper are to tune an ADRC controller through simulation using a nonlinear DC motor model and Simulink, to demonstrate the control of a coreless brushed low-power DC motor, and to compare its simulated and actual performance.

This paper is organized as follows. First, the theoretical background of ADRC and the methodology used to design the control system are presented. Then, the results from both simulation and physical implementation are described. Finally, the main conclusions of this research are discussed.

Methodology

The methodology used in this research was empirical and analytical, which implies a scientific research model based on experimentation and experimental logic, as well as on the observation and application of the concepts around ADRC. To obtain the results described in this work, the control system design was simulated and implemented using a Raspberry Pi 3.

Active disturbance rejection control

To design and implement an ADRC, perturbations must be managed as time-variant signals. This can be done through extended state observers (ESOs). ESOs support the ADRC by estimating the total action of uncertain models and perturbations on the system. Knowing these data allows reacting to and canceling the perturbations and undesirable effects arising from the nonlinearities in the system, thereby enabling robust control [35], [36].

To implement an ADRC, the system under study must be regarded as nonlinear and dynamic. This is described in Eq. (1), where t is the time, ω are the external perturbations, u is the input, y is the output, and b is a real constant of the system, which can be obtained by estimating b_0 [37].

$$\ddot{y}(t) = f\left(t, y, \dot{y}, w\right) + bu \quad (1)$$

Eq. (1) expresses the unknown dynamics of a time-variant system. The principle governing the ADRC is the estimation and cancelation of (1) by means of derivation. The ADRC consists of three fundamental parts: a tracking differentiator (TD), an ESO, and nonlinear state error feedback (NLSEF) [38]-[39].

The general scheme of the ADRC is presented in Fig. 1.

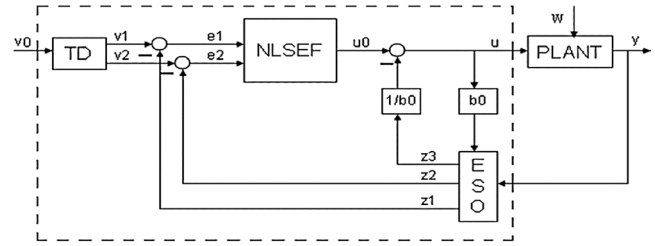


Figure 1. ADRC structure

Source: [40]

The structure of the signals that govern the ADRC is shown in Eqs. (2) to (10) [41].

$$v_1 = v_1 + hv_2 \quad (2)$$

$$v_2 = v_2 + hf \quad (3)$$

$$z_1 = z_1 + h(z_2 - \beta_{01}e) \quad (4)$$

$$z_2 = z_2 + h(z_3 - \beta_{02}fe + b_0u) \quad (5)$$

$$z_3 = z_1 + h(-\beta_{03}fe_1) \quad (6)$$

$$e_1 = v_1 - z_1 \quad (7)$$

$$e_2 = v_2 - z_2 \quad (8)$$

$$u_0 = k_p(r - z_1) + k_d(\dot{r} - z_2) \quad (9)$$

$$u = (u_0 - z_3) / b_0 \quad (10)$$

Where:

- h_1 : the coefficient of precision that determines the aggressivity of the control cycle and is often a multiple of the sampling time h
- $\beta_{01}, \beta_{02}, \beta_{03}$: gain vectors
- e_1, e_2 : differential error
- r_0 : tracking speed
- b_0 : a constant that ensures system robustness
- h : parameter used to track the range of the differential input signal
- z_1, z_2, z_3 : ESO outputs
- v_1, v_2 : TD outputs
- v_0 : TD input

In addition, β_{01} is the proportional gain parameter analogous to proportional gain of a PID controller, and β_{02} is the differential gain, also analogous to the differential gain of a PID controller. If the value of β_{02} is too high, high-frequency noise may occur. The control law relies on the differential errors e_1 and e_2 , so the design will behave like a PD controller. The parameter r defines the tracking speed – a larger value of r means a faster speed. Moreover, the parameter h tracks the range of the differential input signal, and b_0 is a gain constant used to tune the controller and adjust the control signal magnitude. This constant also helps the ESO with perturbation and system error estimation [40].

The system used to tune the controller was the nonlinear model of the DC motor described in Eqs. (11) and (12) [42].

$$L_A \dot{I}_A = -R_A I_A(t) - \Psi w(t) - k_B |w(t)| I_A(t) + U_A(t) \quad (11)$$

$$J \dot{w} = \Psi I_A(t) - M_{F1} w(t) - M_{F0} \text{sign}(w(t)) - M_L(t) \quad (12)$$

Where:

- U : armature voltage
- I_A : armature current
- R_A : armature resistance
- Ψ : magnetic flux
- k_B : voltage drop factor
- M_{F1} : viscous friction
- M_{F0} : dry friction
- M_L : input load

Eqs. (11) and (12) were implemented in MATLAB's Simulink (Fig. 2), considering the technical specifications of the Faulhaber 2342L012CR motor datasheet and assuming a 12 V input with a 100% PWM duty cycle. The input of the motor is a PWM signal, and the output is the position of the motor in radians.

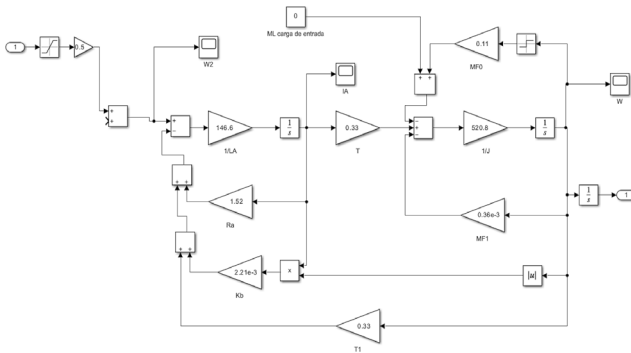


Figure 2. Nonlinear DC motor model implemented in Simulink
Source: Authors

Implementation of the fal and $fhan$ nonlinear functions

Our ADRC was based on nonlinear estimation algorithms and the feedback from nonlinear state variables. In this context, the fal and $fhan$ nonlinear functions were used. These functions are continuous and linear around the origin d , and they are described in terms of the constants d and α .

The parameters $\beta 01$, $\beta 02$, $\beta 03$ establish the dynamic output of the ESO block and depend proportionally on the magnitude of the perturbation. In addition, the delay relies proportionally on $\beta 03$. The fal and $fhan$ functions are defined in Eqs. (13) and (14) [40], [43], [44].

$$fal(x, \alpha, d) = \begin{cases} x / d^{1-\alpha} & , |x| \leq d \\ |x|^\alpha \text{sign}(x) & , |x| > d \end{cases} \quad (13)$$

$$fhan(e_1, e_2, r, h1) = \begin{cases} d = rh^2, a_0 = hx_2, y = x_1 + a_0 \\ a_1 = \sqrt{d(d + 8|y|)} \\ a_2 = a_0 + \frac{\text{sign}(y)(a_1 - d)}{2} \\ S_y = \frac{\text{sign}(y + d) - \text{sign}(y - d)}{2} \\ a = (a_0 + y - a_2)S_y + a_2 \\ S_a = \frac{\text{sign}(a + d) - \text{sign}(a - d)}{2} \\ fhan = -r \left(\frac{a}{d} - \text{sign}(a) \right) S_a - r \text{sign}(a) \end{cases} \quad (14)$$

fal and $fhan$ were programmed through a MATLAB function. The code used to this effect is presented below.

```
function out = fhan(v1,v2,ro,ho)
    d = ho*ho*ro;
    ao = ho*v2;
    y = v1 + ao;
    a1 = sqrt(d*(d + 8*abs(y)));
    a2 = ao + sign(y)*(a1-d)/2;
    sy = (sign(y+d) - sign(y-d))/2;
    a = (ao + y - a2)*sy + a2;
    sa = (sign(a+d) - sign(a-d))/2;
    out = -ro*(a/d - sign(a))*sa - ro*sign(a);
end
```

```
function fe = fal(err,alfa,d)
    if abs(err) <= d
        fe = err/d^(1-alfa);
    else
        fe = (abs(err)^alfa)*sign(err);
    end
```

Implementation of the ESO block

Different observers have been reported in the literature. The most commonly used for ADRC are the unknown input observer, the perturbation observer, and the ESO. For our ADRC, an ESO was used. An ESO not only estimates and compensates for system perturbations but also each state variable. An adequate ESO design ensures that all the closed-loop variables are bounded, as well as the tracking of the reference signal, even in the presence of uncertainty and perturbations. The ESO uses nonlinear gains for the state observer [45]. The gains fe and fe_1 are expressed in Eqs. (15) and (16). Note that, for fe , $\alpha = 0.25$, and, for fe_1 , $\alpha = 0.25$ [40].

$$fe = fal(e, 0.5, h) \quad (15)$$

$$fe_1 = fal(e, 0.25, h) \quad (16)$$

The precision of the estimation is determined using ω_0 . This is the only parameter in the ESO bandwidth. L is defined in Eq. (17) and corresponds to the bandwidth of the state observer [41].

$$L = [l1 \ l2 \ l3]^T = [ew_0, 3w_0^2, w_0^2]^T \quad (17)$$

The code developed for fe is as follows:

```
function y = fcn(err,h)
    alfa=0.5;
    y = fal(err,alfa,h);
end
```

The code developed for fe_1 is presented below.

```
function y = fcn(err,h)
    alfa=0.25;
    y = fal(err,alfa,h);
end
```

The ESO block implemented in Simulink can be seen in Fig. 3. This was done according to Eqs. (2) to (10), (13), (14), and (17).

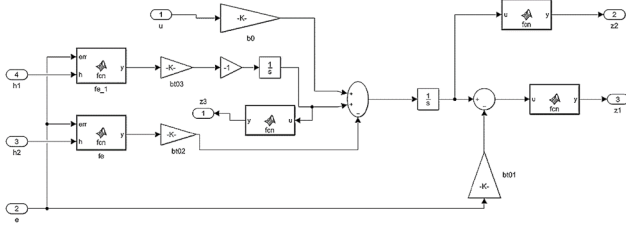


Figure 3. TD block in Simulink
Source: Authors

Implementation of the NLSEF block

The NLSEF block uses the differential signal of the TD and the real-time action of the ESO to control the plant, as shown in Eq. (18) [41].

$$U_0 = -fhan(e1, e2, r1, h1) \quad (18)$$

The NLSEF block was implemented using a MATLAB function in Simulink. The code used is shown below.

```
function u0 = fcn(r0,h0,c,e1,e2)
    u0=-fhan(e1,c*e2,r0,h0);
end
```

Implementation of the TD block

The code developed in MATLAB for the TD block is as follows:

```
function [fh,v2s] = fcn(h,r0,v0,v1,v2)
    v2s=v2;
    fh=fhan((v1-v0),v2,r0,h);
end
```

The TD block developed in Simulink can be seen in Fig. 4.

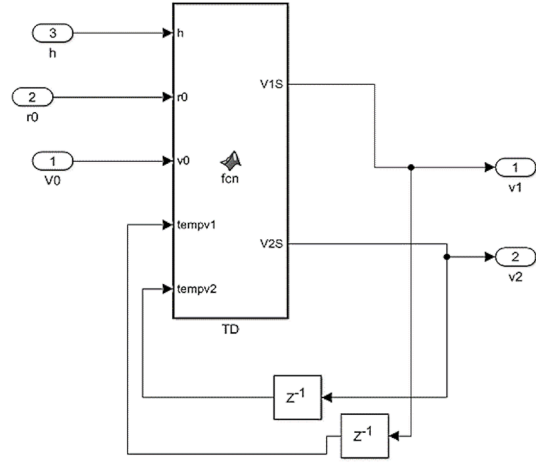


Figure 4. TD block in Simulink
Source: Authors

The scheme developed in Simulink for the ADCR is connected to the nonlinear model of the DC motor (Fig. 5). The implemented ADCR includes the ESO, TD, and NLSEF blocks introduced earlier. The parameters for the ADCR are obtained as shown in [46] i.e., $h = 0.01$, $\beta_{01} = 100$, $\beta_{02} = 300$, $\beta_{03} = 10\,000$, $r_0 = 30$, and $b_0 = 1$.

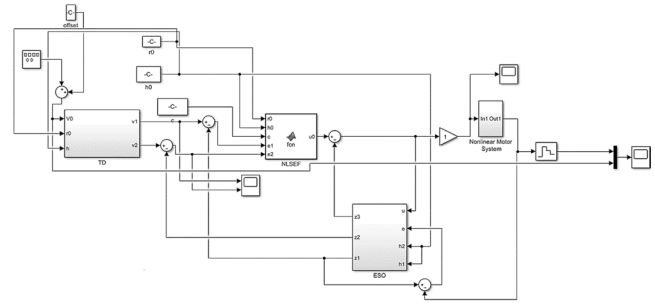


Figure 5. The ADCR and the nonlinear DC motor implemented in Simulink
Source: Authors

Implementation of the ADCR on the Raspberry Pi 3

To program the Raspberry Pi 3 from Simulink, the first step was to install MATLAB's support hardware package for the card. Then, in Simulink, the connection through the ethernet port was configured. This included setting the model of the card and its IP. Afterwards, the libraries developed for the Raspberry Pi 3 were installed. These libraries are used to connect an A/D or a D/A converter to the Pi 3, and they can read or write data from Simulink. The connection is carried out through an S-function. In this case, the converter used was a PCF8591F using the I2C protocol.

The implementation of the controller from Simulink on the Raspberry Pi 3 can be done in two ways, the first of which is maintaining the Pi 3 connected to the computer. This method features the use of scopes and sliders, wherein signals can be visualized in real time and sliders can be used

to change parameters while the Raspberry Pi 3 is controlling the system. To program the card in this way, the *Build model* option is selected in the peripheral section before running the model.

The second method is programming the Pi 3 and using it as an embedded system disconnected from the computer. To this effect, the *external* option must be selected in the simulation settings of the implemented scheme. In this work, the selected sample time was 0.05 s, as there is a noticeable delay when using sample times shorter than 0.01 s, which causes errors during real-time implementation. This delay seems to be due to the processing power of the card.

Fig. 6 presents the scheme of our implementation. Here, the Pi 3 is connected to a computer through an ethernet cable, and the card is connected to a PCF8591F A/D-D/A converter and a L298 H-bridge used to drive the DC motor through a PWM signal using a GPIO pin, wherein the turning sense is controlled by two additional GPIO pins. The L298 is connected to a Faulhaber DC motor coupled to a 10K Ω linear potentiometer, which is in turn connected to a voltage divider. This divider is connected to an instrumentation amplifier, which is in turn connected to the A/D-D/A converter. The converter is used to provide feedback to the Raspberry Pi 3 with information about the motor's position.

Our ADRC, developed in Simulink and implemented on the Raspberry Pi 3, can be seen in Fig. 7. This figure also shows the amplitude and offset correction from the input signal, which are employed to obtain the position of the motor in degrees. The ADRC uses this set of values and reference to calculate the error and uncertainty and compensate for them, using TD, ESO, and NSELF blocks to send a control signal to an A/D converter. Then, these digital signals are sent through GPIO pins to control the DC motor (PWM and turning sense). Simulink scopes can be used to examine the signals coming in and out of the Pi 3 in the computer while the simulation is running. Furthermore, the values of the reference or the ADRC can be modified in real time.

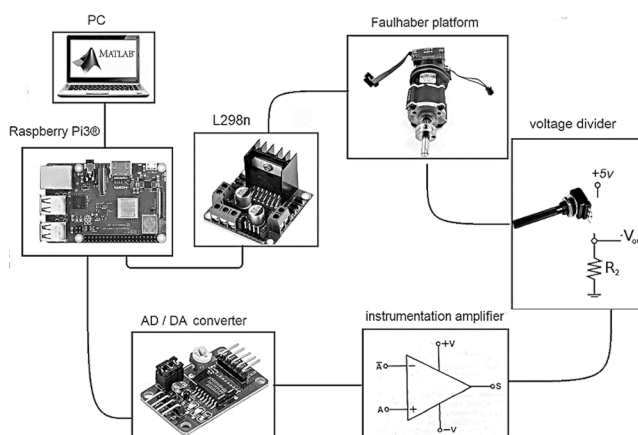


Figure 6. General scheme of the ADRC implemented on a DC motor
Source: Authors

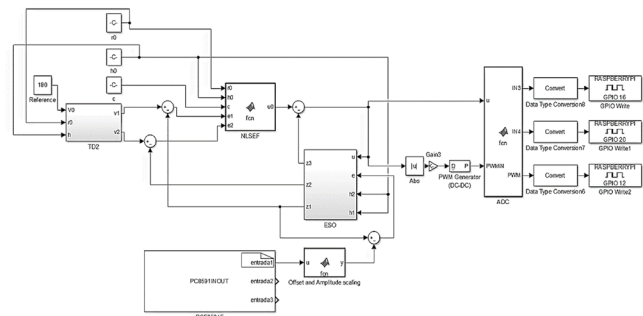


Figure 7. Interaction between Simulink and the Raspberry Pi 3
Source: Authors

Fig. 8 shows the implemented prototype, with the following components: 1) linear potentiometer, 2) Faulhaber DC motor, 3) H-bridge, 4) Raspberry Pi 3, 5) PCF8591 A/D-D/A converter, 6) Protoboard with instrumentation circuit, and 7) DC power supply.

Results and discussion

As previously mentioned, the ADRC was simulated, tuned, and implemented using Simulink and a Raspberry Pi 3, resulting in a robust control system with parameters that can be quickly and easily reconfigured.

Simulation results

Fig. 9 displays the output (motor position in radians) and reference signals obtained in the Simulink simulation using the nonlinear DC motor model (Fig. 2) when controlled by the ADRC. The ADRC effectively controlled the system for a given reference. The settling time was approximately 1.5 s. There was no noticeable overshoot, and the steady-state error was near zero. The control signal (Fig. 9) starts at 100% of the PWM duty cycle and decreases to zero as the output reaches the reference. The nonlinear motor model was controlled successfully in the simulation.



Figure 8. Physical implementation of the prototype
Source: Authors

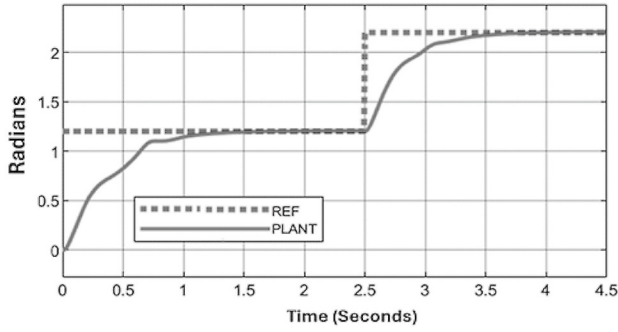


Figure 9. Motor position in the simulation with the ADRC
Source: Authors

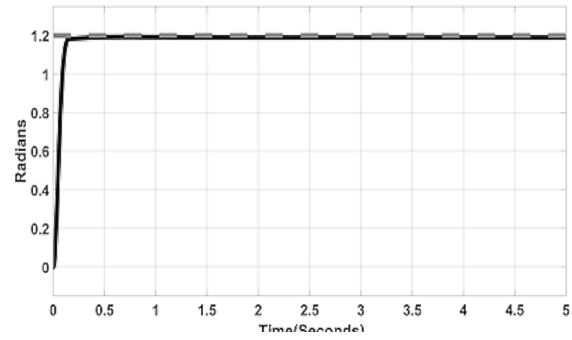


Figure 12. Output signal of the motor position during the simulation of the PI controller
Source: Authors

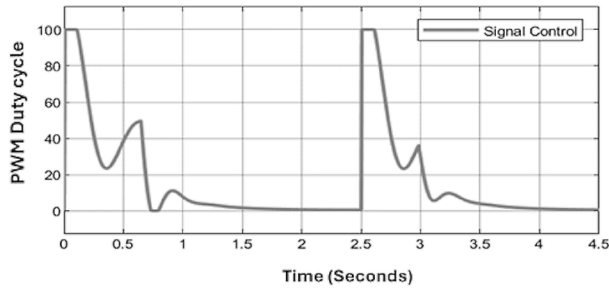


Figure 10. PWM duty cycle of the control signal obtained in the simulation
Source: Authors

To evaluate the controller's simulation performance, a PI controller was implemented, tuned, and simulated in Simulink using the nonlinear DC motor model. The parameters for the controller were $P = 10$ and $I = 0.1$. The integral of the absolute error (IAE), the integral of the time-weighted absolute error (ITAE), the integral of the squared error (ISE), and the integral of time-weighted squared error (ITSE) were calculated for both the PI controller and the ADRC. The simulation setup for the PI controller is shown in Fig. 11, and its response is illustrated in Fig. 12

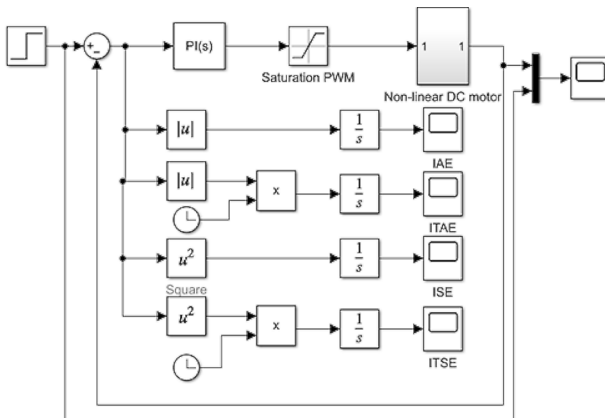


Figure 11. PI controller with its IAE, ITAE, ISE, and ITSE parameters in Simulink
Source: Authors

Fig. 13 compares the IAE values for the ADRC and the PID controller. The latter demonstrates a faster response, but it exhibits a persistent error accumulation due to its steady-state error, indicating higher sensitivity to system nonlinearities. In contrast, the ADRC shows a slower response but stabilizes around 1.5 s with no steady-state error, which highlights its ability to adapt to the model's nonlinear dynamics.

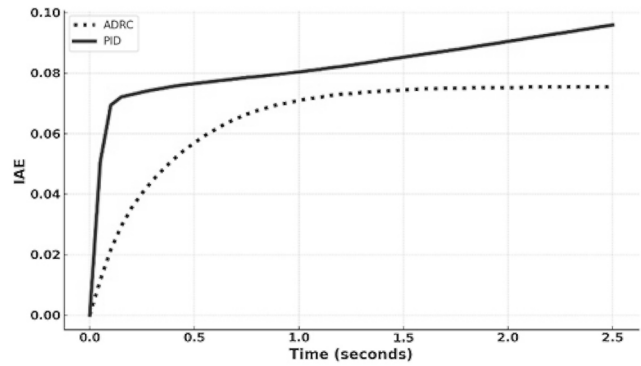


Figure 13. IAE of the control signal obtained in the simulation
Source: Authors

Fig. 14 plots the ISE of both controllers. The PID controller shows a rapid increase in this metric during the first 0.2 s, stabilizing at 0.3 s, with slight increases thereafter. In contrast, the ADRC exhibits a more gradual and continuous increase, stabilizing at 1 s with a lower final ISE than its PID counterpart. This indicates that, while the PID controller responds faster, the ADRC is more efficient at minimizing the accumulated error and provides greater long-term stability, making it more suitable for nonlinear models.

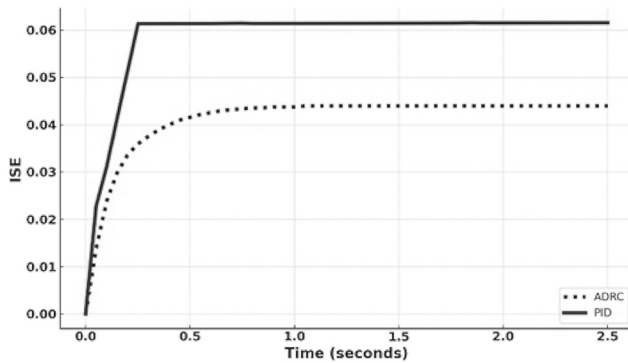


Figure 14. ISE of the control signal obtained in the simulation
Source: Authors

Fig. 15 shows the ITAE for both controllers. The PID controller starts slowly but continues accumulating errors throughout the analyzed interval, suggesting inefficiency in the long term and making it less suitable for systems requiring stability. In contrast, the ADRC shows a faster initial increase but stabilizes at 1 s, significantly reducing the accumulated errors. This behavior makes the ADRC more efficient and better suited for systems that demand precision and robustness.

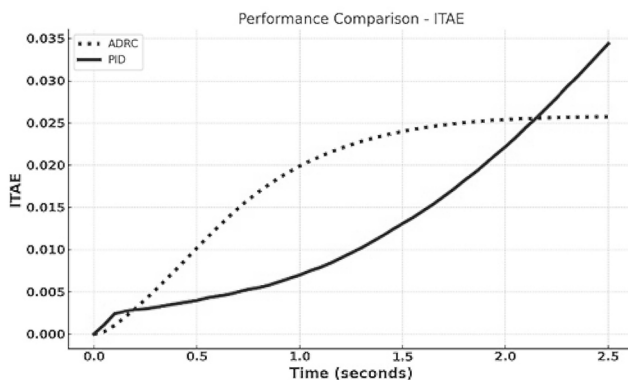


Figure 15. ITAE of the control signal obtained in the simulation
Source: Authors

Fig. 16 presents the ITSE for both controllers. The PID controller exhibits a gradual and uniform increase, reflecting a continuous error accumulation over time. In contrast, the ADRC shows a steeper initial growth but stabilizes quickly, demonstrating greater efficiency in minimizing the accumulated error.

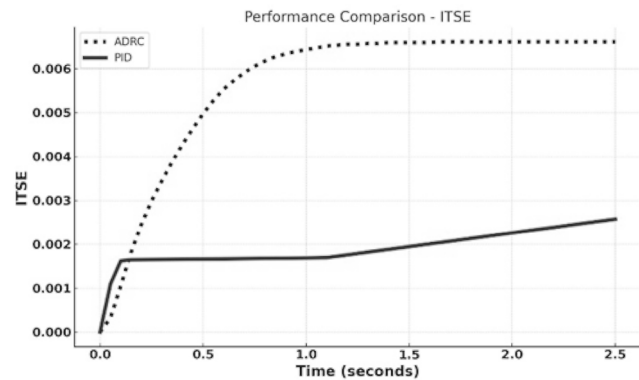


Figure 16. ITSE of the control signal obtained in the simulation
Source: Authors

An analysis of the aforementioned performance indices reveals that the ADRC excels at minimizing the accumulated error and maintaining stability in nonlinear systems, while the PID controller struggles with error accumulation over time, highlighting its limitations in such applications.

Table I presents a comparison in terms of settling time, overshoot, and steady-state error. The PID controller is faster than the ADRC but exhibits persistent steady-state errors. While this can be addressed by adjusting parameters, it leads to an overshoot greater than 10%. On the other hand, the PID controller is even slower and always exhibits overshoot. These findings are consistent with the results of [31], which compares ADRC, PI, and PID controllers on a DC motor.

Table I. Comparison of ADRC and PI controllers

Controller	Settling time (S)	Overshoot (%)	Steady state error (%)
ADRC	1.5	0	0
PI	0.3	0	1

Source: Authors

Implementation results

The results obtained from the Raspberry Pi 3-Simulink implementation are presented in Figs. 17 and 18.

Fig. 17 depicts the position of the motor and the reference signal in radians. The output signal is very noisy due to the brushes. The Faulhaber motor reaches the desired position without oscillation or overshoot, with a settling time of around 2 s. Moreover, the steady-state error is around zero, proving that ADRC can control the position of the motor without precise knowledge of the plant's dynamics – all this, while overcoming the noise. The parameters used in the ADRC were tuned using an approximated nonlinear model, but this is far from a close interpretation of the actual model, especially in terms of friction and noise. Moreover, there are delays related to the Raspberry Pi 3-Simulink connection, as well as others generated within the Raspberry Pi 3 due to

processing power limitations. This becomes evident when the sample time is set to values under 0.01 s.

Fig. 18 presents the control signal applied to the DC motor. It is noisy and quickly becomes zero after reaching the reference.

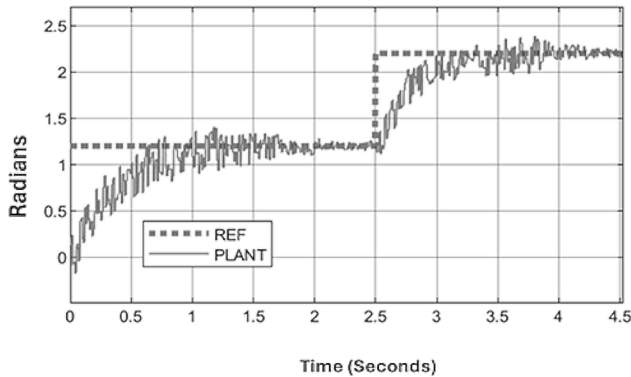


Figure 17. Output signals of the implemented system
Source: Authors

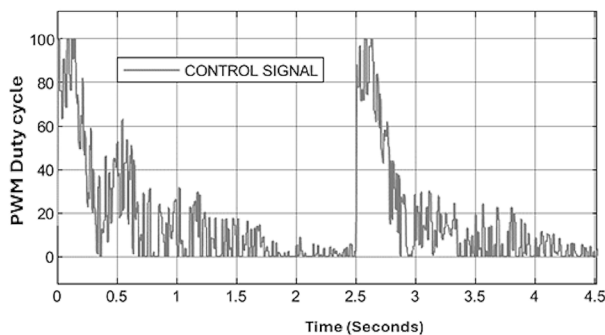


Figure 18. Control signal of the implemented system
Source: Authors

Conclusions

This paper presented the implementation of an active disturbance rejection control for a coreless brushed DC motor. Brushed motors are known for generating noise that can interfere with the control circuits. To effectively manage this motor, the general model of a nonlinear DC motor was used for tuning in simulation. The parameters obtained from the simulation were then applied in practice, demonstrating that ADRC requires minimal information about the plant to achieve effective control. The system actively estimates and compensates for dynamic and unknown disturbances, such as the noise generated by the motor's brushes. Tuning, simulation, and implementation were all carried out using Simulink and a Raspberry Pi 3. The latter was used as an embedded system, facilitating live data exchange with a computer via an Ethernet connection.

In this context, the careful selection of the sample time is crucial, as it must align with the board's processing power to avoid delays that could negatively impact controller

performance. The performance indices used (IAE, ITAE, ISE, and ITSE) highlight the advantages of ADRC over a PID controller when dealing with nonlinear systems. While the latter responds faster, it accumulates more error over time, making it less efficient and stable for long-term control. In contrast, ADRC stabilizes more quickly and minimizes the accumulated error, demonstrating greater robustness and precision. This analysis underscores the suitability of ADRC for applications that demand stability and adaptability to nonlinear dynamics. The development presented in this paper enables the implementation of a portable, easy-to-tune embedded control system, where simulation proved to be an effective tuning tool. Future work will focus on exploring additional nonlinear systems using ADRC, linear active disturbance rejection control (LADRC), and other control strategies. Additionally, deep learning-based tuning strategies will be tested, and other alternatives such as fuzzy controllers will be developed for implementation in Simulink's external mode using Raspberry Pi boards.

CRedit author statement

Mario Gonzalez, Oscar Sierra, and Oscar Higuera conceived the idea and conducted the background research. Oscar Higuera devised the methodology, and all authors collaborated with simulation, implementation, and data collection. The writing of the manuscript was entrusted to Mario González and Oscar Sierra, and its review and editing was performed by Oscar Higuera.

References

- [1] J. Han, "From PID to active disturbance rejection control," *IEEE Trans. Ind. Elect.*, vol. 56, no. 3, pp. 900–906, 2009. <https://doi.org/10.1109/TIE.2008.2011621>
- [2] Zhiqiang Gao, Yi Huang, and Jingqing Han, "An alternative paradigm for control system design," in *Proc. 40th IEEE Conf. Dec. Control* (Cat. No.01CH37228), vol. 5, no. February, pp. 4578–4585, 2001. <https://doi.org/10.1109/2001.980926>
- [3] T. A. Khaled, O. Akhrif, and I. A. Bonev, "Dynamic path correction of an industrial robot using a distance sensor and an ADRC controller," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 3, pp. 1646–1656, 2020. <https://doi.org/10.1109/TMECH.2020.3026994>
- [4] C. Liu, G. Luo, X. Duan, Z. Chen, Z. Zhang, and C. Qiu, "Adaptive LADRC-based disturbance rejection method for electromechanical servo system," *IEEE Trans. Ind. Appl.*, vol. 56, no. 1, pp. 876–889, 2019. <https://doi.org/10.1109/TIA.2019.2955664>
- [5] H. Jin, J. Song, W. Lan, and Z. Gao, "On the characteristics of ADRC: A PID interpretation," *Sci. China Info. Sci.*, vol. 63, no. 10, art. 209201, 2020.
- [6] L. Zhi, M. Huang, L. Qian, Z. Wang, Q. Wen, and W. Han, "Research on active disturbance rejection control with parameter autotuning for a moving mirror control system based on improved snake optimization," *Electronics*,

- vol. 13, no. 9, May 2024. <https://doi.org/10.3390/electronics13091650>
- [7] C. Du, Z. Yin, Y. Zhang, J. Liu, X. Sun, and Y. Zhong, "Research on active disturbance rejection control with parameter autotune mechanism for induction motors based on adaptive particle swarm optimization algorithm with dynamic inertia weight," *IEEE Trans. Power Electron.*, vol. 34, no. 3, pp. 2841–2855, Mar. 2019. <https://doi.org/10.1109/TPEL.2018.2841869>
 - [8] L. Xu, S. Zhuo, J. Liu, S. Jin, Y. Huangfu, and F. Gao, "Advancement of active disturbance rejection control and its applications in power electronics," in *IEEE Trans. Ind. Appl.*, vol. 60, no. 1, pp. 1680–1694, Jan.-Feb. 2024. <https://doi.org/10.1109/TIA.2023.3312653>
 - [9] H. Lu, X. Zhu, C. Ren, S. Ma, and W. Wang, "Active disturbance rejection sliding mode altitude and attitude control of a quadrotor with uncertainties," in *2016 12th World Cong. Intell. Control Autom. (WCICA)*, Guilin, China, 2016, pp. 1366–1371. <https://doi.org/10.1109/WCICA.2016.7578812>
 - [10] K. Xu, L. Lang, Q. Wei, and H. Ma, "Design of single leg foot force controller for hydraulic actuated quadruped robot based on ADRC," in *Chinese Control Conf., CCC*, vol. 2015-Sept, no. 61473304, pp. 1228–1233, 2015. <https://doi.org/10.1109/ChiCC.2015.7259809>
 - [11] X. Meng, H. Yu, J. Zhang, T. Xu, and H. Wu, "Liquid level control of four-tank system based on active disturbance rejection technology," *Measurement*, vol. 175, art. 109146, 2021. <https://doi.org/10.1016/j.measurement.2021.109146>
 - [12] H. Sira-Ramírez, J. Linares-Flores, A. Luviano-Juárez, and J. Cortés-Romero, "Ultramodelos globales y el control por rechazo activo de perturbaciones en sistemas no lineales diferencialmente planos," *Rev. Iberoamer. Autom. Infor. Ind.*, vol. 12, no. 2, pp. 133–144, 2015. <https://doi.org/10.1016/j.riai.2015.02.001>
 - [13] Y. Lei, J. Xu, and Q. Hao, "Application of ADRC in stability control of tank gun system," in *Proc. 2018 IEEE 7th Data Driven Control Learn. Syst. Conf., DDCLS 2018*, 2018, pp. 670–675. <https://doi.org/10.1109/DDCLS.2018.8515958>
 - [14] G. Pasolini, F. Zabini, A. Bazzi, and S. Olivieri, "A software defined radio platform with Raspberry Pi and Simulink," in *2016 24th Euro. Signal Proc. Conf. (EU-SIPCO)*, 2016, pp. 398–402. <https://doi.org/10.1109/EU-SIPCO.2016.7760278>
 - [15] A. Ferdjali, M. Stanković, S. Manojlović, R. Madonski, D. Bujaković, and A. Djenadbja, "Systematic design of nonlinear ADRC for laser seeker system with FPGA-based rapid prototyping validation," *Aircraft Eng. Aerospace Tech.*, vol. 94, no. 7, pp. 1087–1099, 2022. <https://doi.org/10.1108/AEAT-06-2021-0188>
 - [16] H. Zhang, Q. Zhang, Y. Zhang, A. Zhao, T. Ni, and K. Yang, "A DSP-based magnetic compensation system for optically pumped magnetometer," *IEEE Sens. J.*, 2023.
 - [17] H. K. Kondaveeti, D. Bandi, S. E. Mathe, S. Vappangi, and M. Subramanian, "A review of image processing applications based on Raspberry Pi," *2022 8th Int. Conf. Adv. Comp. Comm. Syst. (ICACCS)*, Coimbatore, India, 2022, pp. 22–28. <https://doi.org/10.1109/ICACCS54159.2022.9784958>
 - [18] Á. Ariza and C. N. Galvis, "RaspyControl Lab: A fully open-source and real-time remote laboratory for education in automatic control systems using Raspberry Pi and Python," *HardwareX*, vol. 13, art. e00396, 2023. <https://doi.org/10.1016/j.ohx.2023.e00396>
 - [19] S. E. Mathe, A. C. Pamorthy, H. K. Kondaveeti, and S. Vappangi, "A review on Raspberry Pi and its robotic applications," *2022 2nd Int. Conf. Art. Intell. Signal Proc. (AISP)*, Vijayawada, India, 2022, pp. 1–6. <https://doi.org/10.1109/AISP53593.2022.9760590>
 - [20] A. James, A. Seth, and S. C. Mukhopadhyay, "Programming Raspberry Pi for IoT system," in *IoT System Design*, S. C. Mukhopadhyay, Ed., Cham, Switzerland: Springer, 2022, vol. 41, pp. 51–79. https://doi.org/10.1007/978-3-030-85863-6_4
 - [21] D. Papakyriakou and I. Barbounakis, "Benchmarking and review of Raspberry Pi (RPI) 2B vs RPi 3B vs RPi 3B+ vs RPi 4B (8GB)," *Int. J. Comp. Appl.*, vol. 185, 2023. <https://doi.org/10.5120/ijca2023922693>
 - [22] E. Ilten and M. Demirtas, "Liquid level control interface design on Simulink external mode with Raspberry Pi," in *Proc. Int. Conf. Mod. Adv. Res.*, Konya, Turkey, Aug. 2023, pp. 85–88.
 - [23] K. Utari, N. Mulyaningsih, I. Astuti, Y. Bhakti, and Z. Zulherman, "Physics calculator application with MATLAB as a learning media to thermodynamics concept," *Momentum: Phys. Edu. J.*, vol. 5, no. 2, pp. 101–110, 2021. <https://doi.org/10.21067/mpej.v5i2.5133>
 - [24] R. P. Borase, D. K. Maghade, S. Y. Sondkar, and others, "A review of PID control, tuning methods and applications," *Int. J. Dynamics Control*, vol. 9, pp. 818–827, 2021. <https://doi.org/10.1007/s40435-020-00665-4>
 - [25] J. A. Niembro-Ceceña, R. A. Gómez-Loenzo, and J. Rodríguez-Reséndiz, "SoftCtrlDC-M: Embedded control software for brushed direct current motors," *SoftwareX*, vol. 25, art. 101643, Feb. 2024. <https://doi.org/10.1016/j.softx.2024.101643>
 - [26] S. Liu, M. Xue, Y. Qiu, X. Zhou, and Q. Zhao, "Design of the missile attitude controller based on the active disturbance rejection control," *J. Aerospace Tech. Manage.*, vol. 14, art. 1255, 2022. <https://doi.org/10.1590/jatm.v14.1255>
 - [27] M. Megrini, A. Gaga, and Y. Mehdaoui, "Processor in the loop implementation of artificial neural network controller for BLDC motor speed control," *Results Eng.*, vol. 23, art. 102422, Sep. 2024. <https://doi.org/10.1016/j.rinen.2024.102422>
 - [28] R. Zeng, J. Zhao, Y. Xiong, and X. Luo, "Active disturbance rejection control of five-phase motor based on parameter setting of genetic algorithm," *Processes*, vol. 11, no. 6, art. 1712, Jun. 2023. <https://doi.org/10.3390/pr11061712>
 - [29] M. V. Srikanth and N. Yadaiah, "A magnitude optimum approach for tuning Reduced-order ADRC with FOPDT models," in *2021 7th Indian Control Conf.*, 2021, pp. 46–51. <https://doi.org/10.1109/ICC54714.2021.9703133>
 - [30] W. Ai, M. Wu, X. Li, and X. Li, "Active disturbance rejection based iterative learning control for direct torque control of switched reluctance motor drive," in *Proc. 2021 IEEE 10th Data Driven Control Learn. Syst. Conf.*, May 2021, pp. 1379–1384. <https://doi.org/10.1109/DDCLS52934.2021.9455497>

- [31] M. M. Rahman, M. A. Al Mamon, and M. M. Rahaman, "Active disturbance rejection control based speed control of DC motor," in *2022 Int. Conf. Adv. Elect. Electron. Eng.*, Gazipur, Bangladesh, 2022. <https://doi.org/10.1109/ICAEEE54957.2022.9836391>
- [32] S. Shafi, P. S. Hamid, S. A. Nahvi, M. H. Koul, and M. A. Bazaz, "Active disturbance rejection control of angular position of a DC servo motor," *2023 Int. Conf. Power Inst. Energy Control (PIECON)*, Aligarh, India, 2023, pp. 1–5. <https://doi.org/10.1109/PIECON56912.2023.10085795>
- [33] S. Bose, Y. V. Hote and D. Sengupta, "Analysis and control of real-time DC servo system using linear ADRC," *2018 15th IEEE India Council Int. Conf. (INDICON)*, Coimbatore, India, 2018, pp. 1–6. <https://doi.org/10.1109/INDICON45594.2018.8987040>
- [34] E. Ilten, "Active disturbance rejection control of a DC motor with Raspberry Pi on Simulink external mode," *AS-Proceedings*, vol. 1, art. 124, 2023. <https://doi.org/10.59287/as-proceedings.124>
- [35] H. Sira-Ramírez, R. Castro-Linares, and G. Puriel-Gil, "An active disturbance rejection approach to leader-follower controlled formation," *Asian J. Control*, vol. 16, no. 2, pp. 382–395, 2014. <https://doi.org/10.1002/asjc.714>
- [36] C. Zhang and S. He, "Generalized output feedback active disturbance rejection control for uncertain lower-triangular nonlinear systems," *2016 35th Chinese Control Conf. (CCC)*, Chengdu, China, 2016, pp. 533–538. <https://doi.org/10.1109/ChiCC.2016.7553140>
- [37] D. Wu and K. Chen, "Frequency-domain analysis of nonlinear active disturbance rejection control via the describing function method," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3906–3914, 2013. <https://doi.org/10.1109/TIE.2012.2203777>
- [38] Y. Wang, J. Zhang, and W. Dong, "A new algorithm in blackbody temperature control system based on ADRC," *2017 IEEE Int. Conf. Mechatronics Autom.*, 2017, pp. 226–230. <https://doi.org/10.1109/ICMA.2017.8015818>
- [39] Z. Shen, "Design and simulation of naval gun servo controller based on ADRC," *Proc. 2019 11th Int. Conf. Meas. Tech. Mechatronics Autom.*, pp. 345–349, 2019. <https://doi.org/10.1109/ICMTMA.2019.00083>
- [40] P. Jiang, J. Y. Hao, X. P. Zong, and P. G. Wang, "Modeling and simulation of active-disturbance-rejection controller with Simulink," *2010 Int. Conf. Machine Learn. Cyber.*, 2010, vol. 2, no. July, pp. 927–931. <https://doi.org/10.1109/ICMLC.2010.5580604>
- [41] D. Liu, C. Min, J. Cui, and D. Feng, "Design of attitude controller for hyper velocity projectile based on active disturbance rejection control," in *Int. Conf. Guid. Nav. Control*, 2022, pp. 2709–2719. https://doi.org/10.1007/978-981-19-6613-2_264
- [42] R. Isermann, *Fault-diagnosis systems: An introduction from fault detection to fault tolerance*, 1st ed. Berlin, Germany: Springer-Verlag, 2006. <https://doi.org/10.1007/3-540-30368-5>
- [43] X. Ruan, X. Wang, X. Zhu, Z. Chen, and R. Sun, "Active disturbance rejection control of Single wheel robot," in *Proc. World Cong. Intell. Control Autom. (WCICA)*, 2015, pp. 4105–4110. <https://doi.org/10.1109/WCICA.2014.7053403>
- [44] K. Xu, L. Lang, Q. Wei, and H. Ma, "Design of single leg foot force controller for hydraulic actuated quadruped robot based on ADRC," *Chinese Control Conf., CCC*, 2015, vol. 2015-Sept, no. 61473304, pp. 1228–1233. <https://doi.org/10.1109/ChiCC.2015.7259809>
- [45] S. Liu, M. Xue, Y. Qiu, X. Zhou, and Q. Zhao, "Design of the missile attitude controller based on the active disturbance rejection control," *J. Aerospace Tech. Manage.*, vol. 14, art. e1322, 2022. <https://doi.org/10.1590/jatm.v14.1255>
- [46] X. Ruan, X. Wang, X. Zhu, Z. Chen, and R. Sun, "Active disturbance rejection control of Single wheel robot," in *Proc. World Cong. Intell. Control Autom. (WCICA)* (WCICA), 2015, pp. 4105–4110. <https://doi.org/10.1109/WCICA.2014.7053403>

Lightweight Deep Learning for Atrial Fibrillation Detection: Efficient Models for Wearable Devices

Aprendizaje profundo ligero para la detección de fibrilación auricular: modelos eficientes para dispositivos portátiles

Carlos A. Fajardo¹, Andrés S. Parra², and Tania Castellanos-Parada³

ABSTRACT

The timely and accurate detection of atrial fibrillation (AF) is crucial for an early intervention and for reducing the associated health risks. Wearable technology has emerged as a viable solution for continuous AF monitoring, but deploying accurate AF detection models on resource-constrained devices remains a challenge due to the high computational and memory demands. This study proposes a lightweight and efficient deep learning approach for real-time AF diagnosis in portable devices. We designed a series of convolutional neural network (CNN) models optimized for high accuracy while maintaining a minimal computational footprint. To further enhance efficiency, we explored deep learning compression techniques, including pruning, quantization, and knowledge distillation. Our results demonstrate that the proposed models achieve state-of-the-art accuracy while significantly reducing memory usage and computational complexity, making them suitable for real-time deployment. Additionally, we validated their feasibility by implementing them on a microcontroller, showcasing their practicality for wearable applications. This research paves the way for accessible, low-power, and high-accuracy AF detection in real-world settings, enabling early diagnosis and timely medical intervention without the need for continuous clinical supervision.

Keywords: cardiac arrhythmia, deep learning, detection, ECG, electrocardiogram, machine learning, portable device

RESUMEN

La detección temprana y precisa de la fibrilación auricular (FA) es fundamental para una intervención oportuna y la reducción de riesgos asociados. La tecnología *wearable* ha surgido como una solución viable para el monitoreo continuo de la FA, pero la implementación de modelos precisos de detección en dispositivos con recursos limitados sigue siendo un desafío debido a las altas demandas computacionales y de memoria. Este estudio propone un enfoque de aprendizaje profundo ligero y eficiente para el diagnóstico en tiempo real de la FA en dispositivos portátiles. Diseñamos una serie de modelos de redes neuronales convolucionales (CNN) optimizados para alcanzar alta precisión a la vez que mantenían un bajo costo computacional. Para mejorar aún más la eficiencia, exploramos técnicas de compresión de aprendizaje profundo, incluyendo poda, cuantización y destilación de conocimiento. Nuestros resultados demuestran que los modelos propuestos logran una precisión de última generación mientras reducen significativamente el uso de memoria y la complejidad computacional, lo que los hace adecuados para su implementación en dispositivos de borde. Además, validamos su viabilidad implementándolos en un microcontrolador, demostrando su aplicabilidad en *wearables*. Esta investigación abre el camino para una detección de FA accesible, de bajo consumo y de alta precisión en entornos reales, permitiendo un diagnóstico temprano y una intervención médica oportuna sin necesidad de supervisión clínica continua.

Palabras clave: arritmia cardíaca, aprendizaje profundo, detección, ECG, electrocardiograma, aprendizaje automático, dispositivo portátil

Received: May 20th 2025

Accepted: March 31st 2025

Introduction

Atrial fibrillation (AF) is the most common arrhythmia diagnosed in clinical practice, and its widespread prevalence is considered alarming, with some researchers predicting an epidemic within the next 10 to 20 years [1].

The primary method for detecting AF is analyzing the heart's electrical activity via an electrocardiogram (ECG), which captures patterns that reveal heart function and potential abnormalities. Early detection of AF is crucial in reducing morbidity and mortality. However, AF detection is particularly challenging due to its paroxysmal nature, where episodes occur sporadically and are often asymptomatic in early stages. Consequently, long-term monitoring is recommended for patients exhibiting occasional AF-related symptoms.

The systematic diagnosis of paroxysmal atrial fibrillation (PAF) is an important public health concern, as early identification allows for timely intervention with oral

anticoagulation and catheter ablation, treatments that can be curative when applied at the appropriate time [2]. In recent years, machine learning-based computer-aided diagnosis (CAD) systems have demonstrated high accuracy in AF detection [3], significantly improving clinical workflows [4].

Motivation and challenges

Despite these advancements, current deep learning models for AF detection face key limitations that hinder their deployment in wearable devices. Most high-performing

¹Electronics engineer, Universidad Industrial de Santander, Colombia. PhD, Universidad Industrial de Santander, Colombia. Affiliation: faculty professor, E-mail: cafajar@uis.edu.co

²Electronics engineer, Universidad Industrial de Santander, Colombia. Engineer in charge of control and instrumentation reliability. Ecopetrol. Barrancabermeja, Colombia. Email: andres.parraes@ecopetrol.com.co

³Electronics engineer, Universidad Industrial de Santander, Colombia. MSc, Universidad Industrial de Santander, Colombia. Affiliation: graduate student



Attribution 4.0 International (CC BY 4.0) Share - Adapt

models rely on architectures with millions of parameters, making real-time inference impractical due to high memory and computational demands. Transformer-based and ensemble models, while improving classification accuracy, introduce latency and energy inefficiency, making them unsuitable for edge deployment.

Our approach and contributions

To address these challenges, we propose a lightweight convolutional neural network (CNN) architecture optimized for real-time inference on resource-constrained devices. Our key contributions include:

- **An efficient deep learning model:** We designed a CNN with significantly fewer parameters than traditional models while maintaining state-of-the-art accuracy, making it viable for real-time applications.
- **Model compression strategies:** We explored pruning, quantization, and knowledge distillation to further reduce model size and computational cost while preserving performance.
- **Robustness to noisy signals:** We utilized Icentia11k, one of the largest public ECG datasets, ensuring better generalization. Our approach considers real-world ECG conditions, including the noise and signal variability that are common in wearable devices.
- **Hardware deployment and evaluation:** We implemented our model on a microcontroller, measuring performance in terms of accuracy, memory footprint, inference time, and computational complexity (FLOPs).

This research extends our previous work on low-complexity deep learning models for AF diagnosis [5], incorporating additional model optimizations, the evaluation of a larger dataset, and real-time hardware implementation.

This study follows a structured approach to address the challenges of AF detection in resource-constrained environments. Section 2 presents the dataset used in this research, detailing its characteristics and relevance to the problem at hand. Building upon this foundation, Section 3 introduces the proposed deep learning models, highlighting their design choices and suitability for efficient ECG classification. To further enhance the practicality of our approach, Section 4 explores various compression strategies, including knowledge distillation, pruning, and quantization, with the aim of reducing computational complexity while maintaining classification performance. The implementation of these models on resource-limited hardware is discussed in Section 5, where we evaluate real-time inference feasibility. In Section 6, we present a comprehensive analysis of our results, comparing them against those of related studies and providing key insights into the effectiveness of our method for AF detection. Finally, Section 7 summarizes the study's contributions, discusses its implications, and outlines potential directions for future research.

Data description

Analyzing the heart's electrical activity is the main technique for detecting arrhythmias. This electrical activity is recorded by electrodes on the skin and plotted as a voltage vs. time graph (Fig. 1). This graph is called an *electrocardiogram* (ECG).

There are different types of waves in an ECG signal, such as P-waves (atrial depolarization), T- and U-waves (ventricular repolarization), and QRS complexes (ventricles depolarization). These waves are used to identify anomalies in the behavior of the heart, like cardiac arrhythmia [6, 7].

AF is caused by the asynchronous contraction of the atria due to the fast firings of electrical impulses. AF is characterized by the lack of sinus P-waves, irregular and fast ventricular contraction, an irregular and unusual RR interval, and atrial heart rate oscillating between 140 to 160 beats per minute (bpm) [8].

Atrial flutter is another type of cardiac arrhythmia. Unlike AF, during an atrial flutter episode, the electrical activity of the atria is synchronous, and the atrial heart rate oscillates between 250 to 350 bpm.

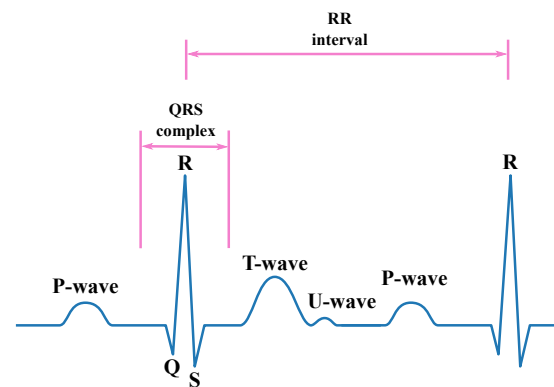


Figure 1. ECG waveform and its main patterns (P-, T-, and U-waves, QRS complexes, and RR intervals)

Source: Authors

In this work, we used the Icentia11k dataset [9], which contains normal sinus rhythms, noises, and AF and atrial flutter signals. As suggested by [4], we merged the AF and atrial flutter signals into one single class.

The Icentia11k dataset was selected for the following reasons. First, it is the largest public ECG dataset, with 11 000 patients and 2 billion labeled beats. This helps to reduce generalization error. Second, this dataset was collected using a single-lead heart monitor device, which allows for a more straightforward future deployment in wearable hardware. Finally, the Icentia11k dataset has a large number of noise signals (around 40%). As previously mentioned, the ability to deal with noise signals is an important goal in the development of wearable devices.

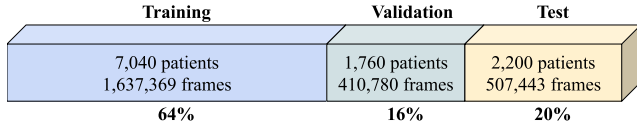
The signals were acquired using the CardioSTAT device [10]. Each patient used the device for a period of two weeks. This device automatically detects the heartbeat. Subsequently, the corresponding labels were verified by a cardiologist who analyzed the complete recording. The dataset has 50 segments per patient, each with 1 048 576 samples. From each of these segments, four frames with 2 049 samples were extracted. Table 1 shows the dataset statistics.

Table 1. Dataset statistics

Statistic	# Units
Number of Patients	11 000
Number of labeled beats	2 774 054 987
Sample Rate	250 Hz
Frame size	2 049 Samples
Total number of frames	2 555 592

Source: Authors

We split the dataset into training, validation, and test sets, using 64%, 16%, and 20% of the data, respectively (Fig. 2). It is important to note that each set has different patients; there are no shared patients between sets.


Figure 2. Strategy used to split the data

Source: Authors

Models

This research aimed to empirically find the smallest and most accurate deep learning model. It proposes a CNN inspired by [4], [11], and [12]. To this effect, we tested several configurations by varying the kernel size, depth, and width.

The network architecture consists of N residual blocks (Fig. 3), where N , I_{ch} , and S_j are hyperparameters that allow varying the model size.

The first and last layers of the model are special cases due to this pre-activation block structure [11]. The model uses residual blocks as a means to deal with the vanishing or exploding gradient problem [13]. Moreover, the model includes skip-connections, which favor the propagation of information in deep neural networks, and it uses batch normalization to keep values within bounds and avoid saturation. It applies dropout to prevent overfitting during training, followed by a rectified linear activation unit (ReLU). Each residual block culminates in average pooling. Furthermore, maxpooling layers were also used in the skip connections to maintain dimensional consistency when the two separate paths joined back together at each block. The model finishes with a fully connected layer, followed by the softmax function, which has three outputs corresponding to the probability of each class (normal, arrhythmia, or noise).

Compact models

Neural networks are typically oversized [14]. Thus, this research aimed to find the smallest and most accurate model empirically. The goal was to determine the minimum values for the model's kernel size, width, and depth. This process focused on the convolutional layers' number of trainable parameters, as they are the majority in the model.

In the proposed model (Fig. 3), the first layers have the same number of initial output channels (I_{ch}). In the residual blocks, the number of output channels (C_{out}^j) is given by

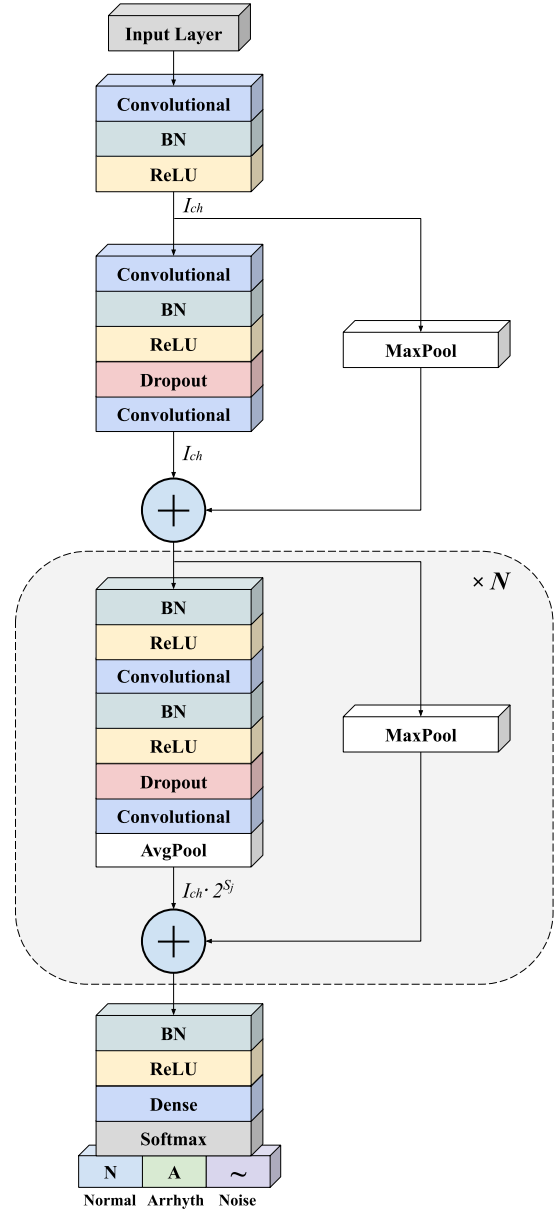


Figure 3. The CNN consists of N residual blocks and accepts raw ECG data as input. It outputs a prediction of normal (N), cardiac arrhythmia (A), or noise signals (~). The number of output channels for the first layers is I_{ch} . The number of output channels for the layers in the residual blocks is $I_{ch} \cdot 2^{S_j}$. The depth and width of the network are changed using the hyperparameters N , I_{ch} , and s_j

Source: Authors

$$C_{out}^j = I_{ch} \cdot 2^{S_j}, \quad (1)$$

where j is the number of the residual block, and s_j and I_{ch} are hyperparameters that allow setting the increment of the output channels throughout the model. This work also applied different strides in the residual blocks.

Let C_{in}^j and C_{out}^j be the input and output channel sizes for the 1D-convolutional layer in the j th-residual block (for notation simplicity, assume that the input and output activations have the same spatial size). The convolutional layer will have $C_{out}^j \cdot C_{in}^j$ filters. Thus, for a kernel size of K , the total number of trainable parameters in the 1-D convolutional layer will be

$$P_{conv} = K \cdot C_{in}^j \cdot C_{out}^j + C_{out}^j, \quad (2)$$

where C_{out}^j and C_{in}^j are calculated using Eq. (1), adjusting the corresponding value of s_j . The last C_{out}^j corresponds to the bias parameters.

We tested several configurations by varying the model's kernel size (K), depth (N), and width (I_{ch} and s_j). Kernel size, a highly influential hyperparameter, underwent extensive training. Our findings indicate that the best value for kernel size is $K = 16$. We also tested different ways to apply downsampling to the CNN by setting a stride greater than 1. The late downsampling strategy [15] was used, which has been shown to improve accuracy on a limited budget of parameters. This strategy aims to generate large activation maps at the beginning of the network, which can lead to higher classification accuracy.

We trained several models with the aim of determining their F1-score. Fig. 4a illustrates the top-performing models of our study. Notably, with just 73 343 parameters, we achieved an F1-score of 0.909, and larger models did not report a significant improvement. Therefore, we decided to focus on the five smallest models (CNN1-CNN5). Fig. 4b shows the number of floating-point operations (FLOPs) for our five CNN models. This number is a measurement used to assess computational complexity. As expected, the larger models yielded higher FLOPs values.

Fig. 5 provides a visual representation of these five models, showcasing their architectural components. Input and output layers are depicted in gray boxes, while the red boxes indicate special-cased blocks. The blue boxes correspond to the N residual blocks, and their output dimensions are highlighted. As an example, the CNN5 model encompasses two red special-cased blocks, each with four output channels. Furthermore, the model incorporates 13 blue residual blocks, wherein the number of output channels progressively increases every fourth residual block. Note that, as the number of filters grows, the output dimension undergoes subsampling.

Training

The deep learning models were implemented in Python, using the Keras library with a TensorFlow backend. The networks were trained using the Adam optimizer with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$) and a mini-batch size of 128. The initial learning rate was set to 0.001 and was reduced by a factor of 10 when the loss plateaued. To prevent overfitting, we employed early stopping, terminating training when the validation loss failed to improve for a predefined number of epochs. The model with the lowest validation loss was selected for the final evaluation.

Compressed models

Several strategies to compress deep learning models have been proposed, which aim to reduce memory and computational requirements without significantly compromising accuracy [14]. By applying these compression techniques, deep learning models can be made more compact, making them easier to deploy on edge devices. In this study, we used knowledge distillation,

pruning, and quantization as compression methods. The subsequent subsections detail the implementation of these strategies.

Knowledge distillation

Knowledge distillation (KD) is a strategy for model compression, whose foundations were established by [16] and were then improved and generalized [17]. KD aims to transfer knowledge from a well-trained larger model (teacher) to a small one (student). [17] suggested that, when a teacher assigns a relatively high probability to the wrong categories, the teacher can offer valuable information regarding similarities between classes, which can be used to train the student. However, in most cases, the teacher assigns very high probabilities to the correct classes and close-to-zero probabilities to the wrong classes. In these cases, the transferred information from the teacher to the student is very similar to the one-hot ground truth labels. To tackle this problem, [17] introduced the concept of *temperature* (T) to calculate the *softmax* equation as follows:

$$p_i(z_i; T) = \frac{\exp(\frac{z_i}{T})}{\sum_{j=1}^C \exp(\frac{z_j}{T})}, \quad (3)$$

where z_i are the logits and C is the number of classes. When T is set to a larger value, the probability distribution becomes softer. The soft probabilities calculated using Eq. (3) are known as *soft labels*, whereas the ground-truth labels are *hard labels*.

In KD, the student is trained using two different loss functions. The first one calculates the Kullback-Leibler divergence between the teacher and the student's soft labels. The second function is the standard cross-entropy (L_{ce}), calculated with model predictions and ground-truth hard labels. Thus, in KD, the complete loss function L is a linear combination between L_{kl} and L_{ce} :

$$L = \alpha L_{ce} + (1 - \alpha) L_{KD}. \quad (4)$$

KD was applied using teachers with as much as 10 times more parameters in order to improve the performance metrics of the student models (CNN1-CNN5). Many tests using different values of α and T , i.e., Eqs. (3) and (4), were carried out, obtaining better results when using high temperatures (> 15) and small α values (< 0.5).

Pruning strategy

Pruning is a compression technique used to reduce the model size by removing non-important connections. [18] proposed a magnitude-based weight pruning strategy to reduce model size and facilitate the deployment of DNNs on wearable and mobile devices. The method prunes network connections in such a way that the original accuracy is preserved. The pruning steps can be consulted in more detail in [18].

We pruned the models using two pruning schedules: zeroing weights constantly throughout the training process or gradually zeroing weights in a polynomial decay fashion. Better results were obtained with the constant pruning schedule. The models CNN1-CNN5 were pruned with 50%

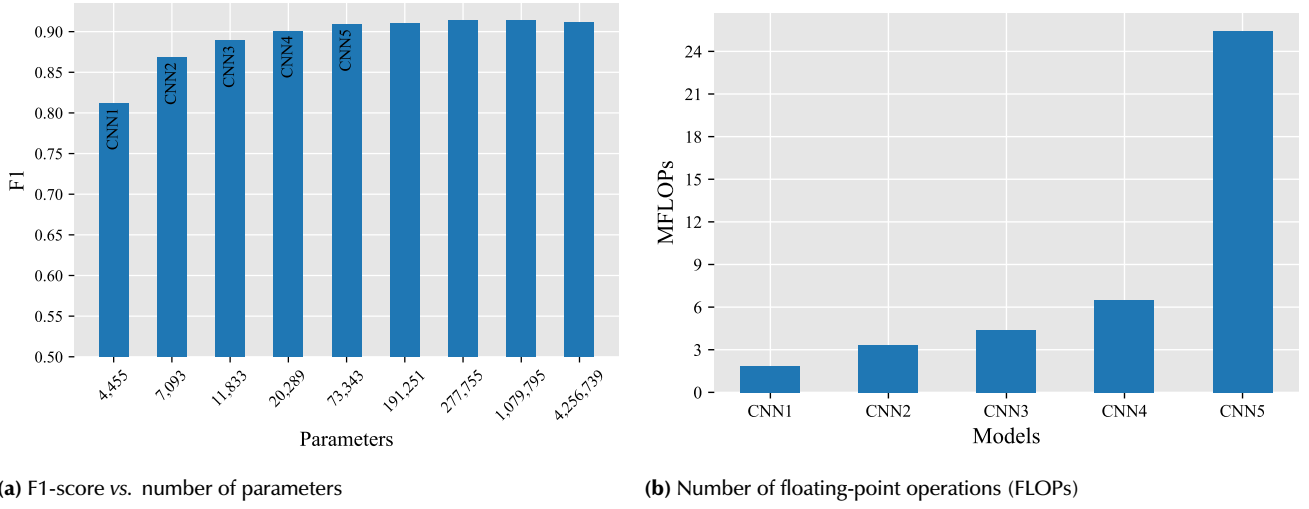


Figure 4. F1-scores and FLOPs for the models CNN1-CNN5

Table 2. Evaluation of different compression methods

Model	Baseline		KD		Pruning		Quantization	
	F1	Size (KB)	F1	Size (KB)	F1	Size (KB) / (% reduction)	F1	Size (KB) / (% reduction)
CNN1	0.81	31	0.82	31	0.80	25 / (21%)	0.80	8 / (75%)
CNN2	0.87	48	0.87	48	0.85	38 / (21%)	0.86	12 / (75%)
CNN3	0.89	63	0.90	63	0.89	46 / (26%)	0.88	16 / (75%)
CNN4	0.90	98	0.90	98	0.89	69 / (29%)	0.89	24 / (75%)
CNN5	0.91	296	0.91	296	0.91	193 / (35%)	0.90	74 / (75%)

Source: Authors

sparsity because using a higher value yielded a poor trade-off between F1-score and size. Nevertheless, models larger than CNN5 allowed for pruning with a sparsity greater than 50% without compromising the metrics.

Quantization

This research used an 8-bit quantization implementation based on [19]. Considering an unsigned 8-bit representation, we mapped the floating-point variable to 256 values (0, 255). To this effect, two parameters were considered: scale (S) and zero-point (Z), which allow mapping the floating-point value (r) to the corresponding 8-bit value (q) using the following equation:

$$r = S(q - Z). \quad (5)$$

Results obtained from the compression methods

Table 2 presents the results of applying these three compression strategies to our proposed models. The table shows the F1-score and model size (in KB). The percent reduction in model size, achieved through pruning and quantization, is also presented.

KD did not enable a significant improvement in the F1-score and failed to reduce the model size. Conversely, pruning reduced the model size and, in certain cases, it even improved performance due to the regularization effect of the pruning strategy. Remarkably, quantization emerged as the most effective method, reducing the model size by 75% with only a slight decrease of 0.01 in the F1-score.

Hardware deployment

We evaluated the models by implementing the inference stage in a Teensy 4.1 microcontroller based on ARM Cortex-M7, working at a frequency of up to 600 MHz. It has 8 MB of FLASH memory and 1 MB of RAM [20].

Table 3 presents the performance of our implemented models. It includes the following metrics [21]:

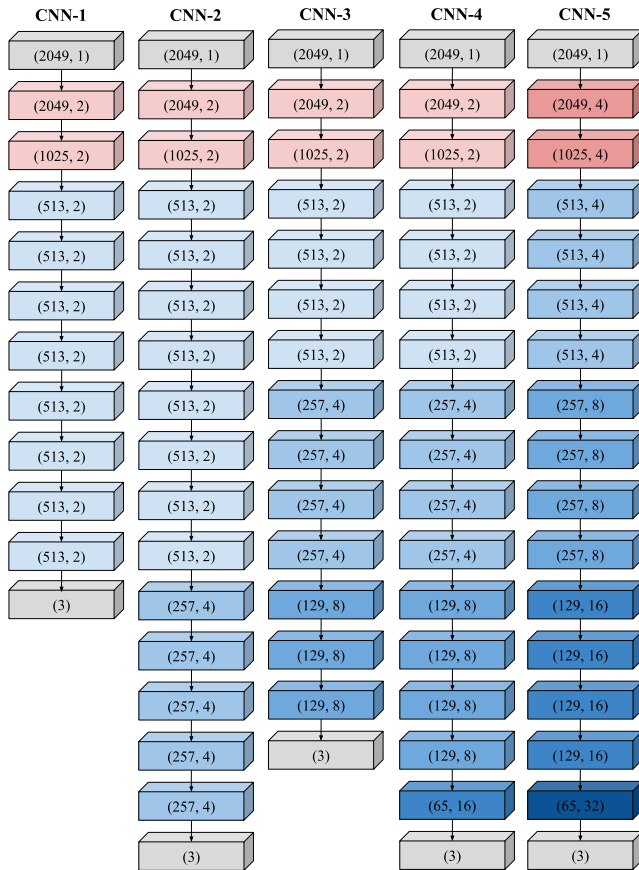
- **Model size** is reported in terms of the number of parameters used.
- **Accuracy** is measured using the F1-score, which assesses the model's performance in terms of precision and recall.
- **Computational cost** is measured in the number of millions of FLOPs required for inference.
- **Peak memory footprint** indicates the maximum amount of memory (in KB) consumed by the models during inference.
- **Inference time** (in ms) is measured at both the minimum and maximum operating frequencies.

The results presented in Table 3 demonstrate the feasibility of implementing the models on this resource-constrained device. The peak memory footprint in the flash memory ranged from 1% (CNN1) to 4% (CNN5) of the available memory. RAM usage ranged from 11% (CNN1) to 53% of the available memory. The inference time was sufficiently small to support real-time inference. However, it is important to

Table 3. Evaluation metrics of the models on the Teensy 4.1

Model	Size # parameters	F1	MFLOPs	Peak Memory		Inference time (ms)	
				Flash (KB)	RAM (KB)	150 MHz	600 MHz
CNN1	4455	0.81	1.84	63.7	115.3	227	58
CNN2	7093	0.87	3.35	77.9	148.1	479	122
CNN3	11 833	0.89	4.39	96.3	213.7	559	145
CNN4	20 289	0.90	6.53	131.1	295.7	816	210
CNN5	73 343	0.91	25.45	344.0	529.0	2039	599

Source: Authors

**Figure 5.** Architecture used in our five models. Each box represents a residual block and its output dimension. Boxes in gray are input/outputs, boxes in red are the special-cased blocks, and boxes in blue represent the N residual blocks.

Source: Authors

note that the RAM requirement may be a critical factor to consider.

Discussion

Among the classical algorithms used for ECG signal classification are support vector machines [26], naive Bayes [27], linear and quadratic discriminant analysis, and random forests [28], among others. However, these classical techniques typically require two steps: feature extraction (pre-processing) and classification. This feature engineering process often relies on domain expertise and can introduce biases. In contrast, deep learning approaches can learn hierarchical features directly from raw signals, eliminating the need for manual feature extraction. Moreover, reducing pre-processing steps

can lower computational costs and energy consumption, making deep neural networks more suitable for real-time applications.

Deep learning techniques have significantly advanced ECG classification and arrhythmia detection [29, 4, 30, 31]. CNNs can capture temporal and spatial patterns in ECG signals [23, 32], while recurrent neural networks (RNNs) and transformer-based models enhance the modeling of sequential dependencies [33].

Recent approaches integrate hybrid deep learning models that combine convolutional and recurrent architectures to enhance feature extraction and classification performance [34]. Additionally, temporal convolutional networks (TCNs) have emerged as an alternative to recurrent models, offering greater efficiency in capturing long-term dependencies while reducing computational complexity [35]. Ensemble methods that combine multiple deep learning models have also been explored to improve robustness and generalization [36].

These studies primarily aim to improve classification accuracy, often at the expense of increased model size. However, larger models pose challenges regarding memory consumption and computational complexity [14, 24], which are critical for deployment on resource-constrained edge devices. The inclusion of recurrent layers { such as long short-term memory (LSTM) cells } or attention mechanisms can further elevate computational costs, making real-time implementation impractical [37].

Unlike previous deep learning models that prioritize accuracy at the expense of high computational complexity, our approach balances efficiency and performance. As shown in Table 4, models with a large number of parameters often achieve state-of-the-art accuracy but are impractical for wearable devices due to their excessive memory and power requirements.

Table 4 provides an analysis of our proposed models vs. recent ECG classification studies that use large datasets (more than 8000 patients) and explicitly report model parameters. This comparison focuses on key factors such as the size of the dataset, the number of classes, the deep learning techniques used, and the classification performance. Given the challenges of a direct comparison due to variations in the datasets, the number of ECG leads, and model objectives, our analysis prioritizes studies that align with our focus on both accuracy and computational efficiency. Specifically, we included research that reports on model parameters, allowing for estimations of computational cost, a crucial factor for deployment on wearable devices. Furthermore, we prioritized studies

Table 4. Recent studies, on larger data sets, classifying ECG signals using Deep Learning techniques.

Study	Database	Total data & classes	Number of patients	Technique	Number of parameters	Results
[4]	Zio Monitor	91 232 records, 12 classes	53 549	CNN	10 473 635	F1 = 0.837
[22]	PhysioNet Challenge 2017	8528 records, 4 classes	8528	CNN	262 344	F1 = 0.820
[23]	1st China Physiological Signal Challenge	9831 records, 8 classes	9831	ATI-CNN	4 984 640	F1 = 0.812
[24]	Icentia11k	550 000 records, 4 classes	11 000	Contrastive CNN (few-shot)	250 000	F1 = 0.801*
[25]	ECG5000	5000 records	Not specified	Temporal CNN (TCN)	10 200	F1 = 0.8413
CNN1 (Ours)	Icentia11k	550 000 records, 4 classes	11 000	CNN	4455	F1 = 0.813
CNN2 (Ours)	Icentia11k	550 000 records, 4 classes	11 000	CNN	7093	F1 = 0.872
CNN3 (Ours)	Icentia11k	550 000 records, 4 classes	11 000	CNN	11 833	F1 = 0.891
CNN4 (Ours)	Icentia11k	550 000 records, 4 classes	11 000	CNN	20 289	F1 = 0.902
CNN5 (Ours)	Icentia11k	550 000 records, 4 classes	11 000	CNN	73 343	F1 = 0.909

Source: Authors

reporting the F1-score, as accuracy alone can be misleading in unbalanced datasets.

A notable trend in Table 4 is that models trained on larger datasets require more parameters for effective generalization. Our study used one of the largest publicly available ECG datasets in terms of both signal count and patient coverage. Remarkably, our CNN-4 model, with only 20 289 parameters, achieved an F1-score of 0.902, demonstrating high efficiency relative to its size. Furthermore, its 8-bit quantized version reduced the memory footprint by 75% while incurring a minimal F1-score reduction of 0.01. This result underscores the balance between model performance and deployment feasibility, and therefore the importance of designing efficient architectures for real-world applications.

Future work should focus on optimizing deep learning architectures to further reduce computational complexity while maintaining high accuracy. Lightweight models, KD [38], and hardware-aware optimizations are promising directions to bridge the gap between accuracy and efficiency in ECG classification.

Conclusions and future work

This study addressed the challenges of atrial fibrillation detection in edge devices by proposing deep learning models optimized for low computational complexity. Unlike previous studies that focus primarily on accuracy, often at the expense of increased model size and computational demands, our approach achieved a balance between efficiency and performance. The proposed CNN-4 model, with only 20 289 parameters, exhibited a competitive classification performance while significantly reducing memory and power requirements, making it well-suited for

real-time applications on portable devices. Additionally, the quantized version of our model achieved a 75% reduction in memory footprint while maintaining minimal losses regarding the F1-score, further reinforcing its suitability for resource-constrained environments.

Despite these promising results, certain limitations must be considered. The performance of compressed models may be affected by variations in ECG signal quality, and their generalization to different datasets requires further validation. Moreover, while our models were evaluated using a public ECG dataset, real-world deployment necessitates pre-clinical testing under diverse conditions, including different sensor types and patient demographics.

Future work will explore hybrid compression techniques, such as the combination of pruning and quantization, to further reduce model size while preserving classification accuracy. We will also investigate the feasibility of ultra-low-bit quantization (e.g., fewer than eight bits) to enhance energy efficiency. Another critical direction involves deploying and optimizing our models on specialized hardware platforms, including FPGAs and System-on-Chip (SoC) architectures, in order to ensure efficient inference on edge devices.

Additionally, the development of a proof of concept for a wearable AF detection device will be pursued, integrating our models into real-time data acquisition systems. This step is crucial for validating the practical applicability of our approach in clinical settings.

Ultimately, this research contributes to advancing AF detection on edge devices, addressing the growing need for early diagnosis and intervention. By optimizing deep learning models for real-world deployment, we aim to

improve patient outcomes and support the adoption of AI-driven diagnostic tools in wearable healthcare technologies.

Acknowledgment

The authors would like to thank the CPS research group of Universidad Industrial de Santander and C-BRIC from Purdue University.

CRedit author statement

All authors: conceptualization, methodology, software, validation, formal analysis, investigation, writing (original draft, writing, review, and editing), data curation.

Conflicts of interest

The authors declare that they have no conflict of interest with regard to the publication of the results of this research.


References

- [1] C. A. Morillo *et al.*, "Atrial fibrillation: The current epidemic," *J. Ger. Card.*, vol. 14, no. 3, p. 195, 2017. <https://doi.org/10.11909/j.issn.1671-5411.2017.03.011>.
- [2] J. Primo *et al.*, "Prevalence of paroxysmal atrial fibrillation in a population assessed by continuous 24-hour monitoring," *Rev. Port. Card.*, vol. 36, no. 7-8, pp. 535–546, 2017. <https://www.revportcardiol.org/en-prevalence-paroxysmal-atrial-fibrillation-in-articulo-S217420491730199X>.
- [3] C. Y. Chenggong Xie, Zhao Wang, J. Liu, and H. Liang, "Machine learning for detecting atrial fibrillation from ecgs: Systematic review and meta-analysis," *RCM*, vol. 25, no. 1, p. 8, 2024. <https://doi.org/10.31083/j.rcm2501008>.
- [4] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, 2019. <http://dx.doi.org/10.1038/s41591-018-0268-3>.
- [5] C. A. Fajardo, A. S. Parra, T. Castellanos-Paradaa, and K. Roy, "Low-complexity deep learning models for accurate atrial fibrillation diagnosis," in *26th Eur. Conf. Art. Intel.*, 2023. https://ecai2023.eu/conf-data/ecai2023/files/STAIRS/stairs2023_05.pdf.
- [6] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *IEEE Trans. Inst. Measure.*, vol. 50, no. 3, pp. 808–812, 2001. <https://doi.org/10.1109/19.930458>.
- [7] R. Avanzato and F. Beritelli, "Automatic ecg diagnosis using convolutional neural network," *Electronics*, vol. 9, no. 6, p. 951, 2020. <https://doi.org/10.3390/electronics9060951>.
- [8] J. A. Castillo, Y. C. Granados, and C. A. Fajardo, "Patient-specific detection of atrial fibrillation in segments of ecg signals using deep neural networks," *Cien. Ing. Neogranadina*, vol. 30, no. 1, pp. 45–58, 2020. <https://revistas.unimilitar.edu.co/index.php/rcin/article/view/4156/4249>.
- [9] S. Tan, G. Androz, A. Chamseddine, P. Fecteau, A. Courville, Y. Bengio, and J. P. Cohen, "Icentia11k: An unsupervised representation learning dataset for arrhythmia subtype discovery," *arXiv preprint arXiv:1910.09570*, 2019. <https://arxiv.org/abs/1910.09570>.
- [10] CardioSTAT, "Ambulatory Cardiac Monitoring," 2021. <https://www.cardiostat.com/>.
- [11] P. Rajpurkar, A. Y. Hannun, M. Haghighpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint: 1707.01836*, 2017. <https://arxiv.org/abs/1707.01836>.
- [12] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," *Comp. Card.*, vol. 44, pp. 1–4, 2017. <https://www.cinc.org/archives/2017/pdf/066-138.pdf>.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2016. <https://doi.org/10.1109/CVPR.2016.90>, pp. 770–778.
- [14] G. C. Marinó, A. Petrini, D. Malchiodi, and M. Frasca, "Deep neural networks compression: A comparative survey and choice recommendations," *Neurocomputing*, vol. 520, pp. 152–170, 2023. [Online]. Available: <https://doi.org/10.1016/j.neucom.2022.11.072>
- [15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 mb model size," *arXiv preprint: 1602.07360*, 2016. <https://arxiv.org/abs/1602.07360>.
- [16] C. Bucilua, R. Caruana, and A. Niculescu-Mizil, "Model compression," in *Proc. 12th ACM SIGKDD Int. Conf. Know. Disc. Data Mining*, 2006. <https://dl.acm.org/doi/10.1145/1150402.1150464>, pp. 535–541.
- [17] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint: 1503.02531*, 2015. <http://arxiv.org/abs/1503.02531>.
- [18] S. Han, J. Pool, J. Tran, and W. J. Dally, "Learning both weights and connections for efficient neural networks," *arXiv preprint: 1506.02626*, pp. 1–9, 2015. <http://arxiv.org/abs/1506.02626>.
- [19] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2018. <https://ieeexplore.ieee.org/document/8578384>, pp. 2704–2713.
- [20] SparkFun-Electronics, "Teensy 4.1." <https://www.sparkfun.com/products/16771>.
- [21] O. A. Ademola, M. Leier, and E. Petlenkov, "Evaluation of deep neural network compression methods for edge devices using weighted score-based ranking scheme," *Sensors*, vol. 21, no. 22, p. 7529, 2021. <https://www.mdpi.com/1424-8220/21/22/7529>.
- [22] J. Rubin, S. Parvaneh, A. Rahman, B. Conroy, and S. Babaeizadeh, "Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ecg recordings," *J. Electrocard.*, vol. 51, no. 6, pp. S18–S21, 2018. <https://www.sciencedirect.com/science/article/abs/pii/S0022073618303315>.

- [23] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network," *Info. Fus.*, vol. 53, no. June 2019, pp. 174–182, 2020. <https://dl.acm.org/doi/abs/10.1016/j.inffus.2019.06.024>.
- [24] K. Fonseca, S. Osorio, J. Castillo, and C. Fajardo, "Contrastive learning for atrial fibrillation detection in challenging scenarios," in *2022 30th Eur. Signal Proc. Conf. (EUSIPCO)*. IEEE, 2022. <https://doi.org/10.23919/EUSIPCO55093.2022.9909842>, pp. 1218–1222.
- [25] A. R. Iskandar, S. Jundi, H. Rifai, and N. Rizoug, "Ecg classification using an optimal temporal convolutional network for remote health monitoring," *Sensors*, vol. 23, no. 3, p. 1697, 2023. <https://doi.org/10.3390/s23031697>.
- [26] S. Asgari et al., "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Comp. Biol. Med.*, vol. 60, pp. 132–142, 2015. <https://doi.org/10.1016/j.combiomed.2015.03.005>.
- [27] R. J. Martis et al., "Automated detection of atrial fibrillation using bayesian paradigm," *Knowledge-based Syst.*, vol. 54, pp. 269–275, 2013. <https://doi.org/10.1016/j.knosys.2013.09.016>.
- [28] M. Mohebbi and H. Ghassemian, "Detection of atrial fibrillation episodes using svm," in *2008 30th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.* IEEE, 2008. <https://doi.org/10.1109/iembs.2008.4649119>, pp. 177–180.
- [29] S. Parvaneh, J. Rubin, S. Babaeizadeh, and M. Xu-Wilson, "Cardiac arrhythmia detection using deep learning: A review," *J. Electrocard.*, vol. 57, pp. S70–S74, 2019. <https://doi.org/10.1016/j.jelectrocard.2019.08.004>.
- [30] A. H. Ribeiro, D. M. Tavares de Oliveira, P. R. Gomes, D. Canazart, M. Ferreira, C. R. Andersson, A. L. Ribeiro, T. B. Schon, W. Meira, and L. P. Rocha, "Automatic diagnosis of the 12-lead ecg using a deep neural network," *Nature Comm.*, vol. 11, no. 1, p. 1760, 2020.
- [31] O. Yildirim, U. B. Baloglu, R. Tan, and U. R. Acharya, "Arrhythmia classification using transformer-based deep learning model," *Biomed. Signal Proc. Control*, vol. 81, p. 104477, 2023.
- [32] J. Rubin, S. Parvaneh, A. Rahman, B. Conroy, and S. Babaeizadeh, "Densely connected convolutional networks for ecg classification," *J. Electrocard.*, vol. 57, pp. S20–S24, 2018. [10.1016/j.jelectrocard.2018.10.004](https://doi.org/10.1016/j.jelectrocard.2018.10.004).
- [33] Y. Oh, H. Moon, D. Kim, J. Park, and D. Kim, "Transformer-based arrhythmia detection on ecg signals," *Sensors*, vol. 22, no. 8, p. 2966, 2022.
- [34] I. J. Selvam, M. Madhavan, and S. K. Kumarasamy, "Detection and classification of electrocardiography using hybrid deep learning models," *Hellenic J. Card.*, vol. 81, pp. 75–84, 2025.
- [35] A. R. Ismail, S. Jovanovic, N. Ramzan, and H. Rabah, "Ecg classification using an optimal temporal convolutional network for remote health monitoring," *Sensors*, vol. 23, no. 3, p. 1697, 2023. <https://doi.org/10.3390/s23031697>.
- [36] T. Mahmud, A. Barua, D. Islam, M. S. Hossain, R. Chakma, K. Barua, M. Monju, and K. Andersson, "Ensemble deep learning approach for ECG-based cardiac disease detection: Signal and image analysis," in *2023 Int. Conf. Info. Comm. Tech. Sust. Dev. (ICT4SD)*, 2023, pp. 70–74.
- [37] Z. Zhao, X. Dong, J. Liu, Y. Guo, and Y. Yao, "ECG classification using a lightweight transformer-based model," *Biomed Signal Proc. Control*, vol. 78, p. 103931, 2022.
- [38] E. Tanghatari, M. Kamal, A. Afzali-Kusha, and M. Pedram, "Federated learning by employing knowledge distillation on edge devices with limited hardware resources," *Neurocomputing*, vol. 531, pp. 87–99, 2023.

Influence of the Synthesis Route on the Structural and Magnetic Properties of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ Nanoparticles

Influencia de la ruta de síntesis en las propiedades estructurales y magnéticas de nanopartículas de $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$

Jenny A. Mera-Córdoba ¹, Jhon F. Betancur-Pérez ², Nicolás A. Salazar-Henao ³
Luis Alejandro Galeano ⁴ Andrés Rosales-Rivera ⁵, and Diego F. Coral-Coral ⁶

ABSTRACT

Cancer is one of the leading causes of death worldwide, significantly impacting public health, with current treatment options that often have side effects on patients. In this context, magnetic hyperthermia emerges as a non-invasive alternative that utilizes magnetic nanoparticles to generate heat and destroy cancer cells. In this vein, this research sought to synthesize magnetic $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (LSMO) nanoparticles in order to study the effect of the synthesis route on the production of nanoparticles with optimal properties for biomedical applications. LSMO samples were synthesized via the sol-gel, ceramic, and Pechini methods. These samples were characterized through X-ray diffraction (XRD), small-angle X-ray scattering (SAXS), thermogravimetric analysis (TGA), and vibrating sample magnetometry (VSM) in order to study their structure, morphology, and magnetic behavior. The nanoparticles obtained via the Pechini method exhibited the best crystalline structure, the smallest size, and reduced magnetic properties. This work allowed identifying the ceramic method as the synthesis route that produces nanoparticles suitable for biomedical applications, as demonstrated through the numerical calculation of the specific absorption rate (SAR). The results indicated that the SAR of nanoparticles synthesized using this method is 30 times greater than that of samples synthesized using the other two methods under equal field amplitude and frequency conditions.

Keywords: ceramic synthesis, hyperthermia applications, manganites, Pechini synthesis, sol-gel synthesis

RESUMEN

El cáncer es una de las principales causas de muerte a nivel mundial y tiene un impacto significativo en la salud pública, con opciones de tratamiento actuales que suelen generar efectos secundarios en los pacientes. En este contexto, la hipertermia magnética surge como una alternativa no invasiva que utiliza nanopartículas magnéticas para generar calor y destruir células cancerígenas. Dado lo anterior, esta investigación buscó sintetizar nanopartículas magnéticas de $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (LSMO) para estudiar el efecto de la ruta de síntesis en la producción de nanopartículas con propiedades óptimas para aplicaciones biomédicas. Las muestras de LSMO fueron sintetizadas mediante los métodos sol-gel, cerámico y Pechini. Estas muestras fueron caracterizadas mediante difracción de rayos X (XRD), dispersión de rayos X a bajo ángulo (SAXS), análisis termogravimétrico (TGA) y magnetometría de muestra vibrante (VSM), con el fin de estudiar su estructura, morfología y comportamiento magnético. Las nanopartículas obtenidas mediante el método Pechini presentaron la mejor estructura cristalina, el menor tamaño y propiedades magnéticas reducidas. Este trabajo permitió identificar el método cerámico como la ruta de síntesis que produce nanopartículas adecuadas para aplicaciones biomédicas, como se demostró mediante el cálculo numérico de la tasa específica de absorción (SAR). Los resultados indicaron que el SAR de las nanopartículas sintetizadas con este método es 30 veces mayor que el de las muestras sintetizadas mediante los otros dos métodos bajo iguales condiciones de amplitud de campo y frecuencia.

Palabras clave: aplicaciones de hipertermia, manganitas, síntesis cerámica, síntesis sol-gel, síntesis Pechini

Received: August 16th, 2024

Accepted: May 19th, 2025

¹ PhD(c) in Sustainable Development, Universidad de Manizales, Colombia. MSc in Sustainable Development and Environment, Universidad de Manizales, Colombia. PhD Fellow, Ministry of Science, Technology, and Innovation (Min-ciencias). Affiliation: Research Group in Nanostructured Functional Materials, Universidad CESMAG, Colombia. Email: jennyalmc@gmail.com

² PhD in Agricultural Science, Universidad Nacional de Colombia. Affiliation: Research professor, Universidad de Manizales, Colombia. E-mail: jbetancur@umanizales.edu.co

³ MSc in Physics, Universidad Nacional de Colombia, Colombia. Affiliation: Magnetism and Advanced Materials Laboratory, Universidad Nacional de Colombia, Manizales campus, Manizales, Colombia. E-mail: nashenao@gmail.com

⁴ PhD in Inorganic Chemistry and Chemical Reactivity, University of Salamanca, Spain. Affiliation: Functional Materials and Catalysis Research Group (GIMFC), Universidad de Nariño, Colombia. E-mail: luis.alejandrogaleano@gmail.com

⁵ PhD in Physics, Universidade Federal de Pernambuco, Brazil. Affiliation: Magnetism and Advanced Materials Laboratory, Universidad Nacional de Colombia, Manizales campus, Manizales, Colombia. E-mail: arosalesr@unal.edu.co

⁶ PhD in Physics, Universidad Nacional de La Plata, Argentina. Affiliation: Research Group in Science and Technology of Ceramic Materials, Department of Physics, Universidad del Cauca, Colombia. Email: dfcoral@unicauca.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Introduction

It is well known that particle size, morphology, and crystalline homogeneity influence the properties of materials at the nanometric scale. In the case of manganites such as $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ (LSMO), these properties can be modified through different synthesis techniques. Particularly, particle size can be controlled using different synthesis methods and processing conditions [1]–[4].

The properties of LSMO that depend on particle size are critical for biomedical applications, e.g., saturation magnetization (M_s), and it has been demonstrated that particle size range influences the magnetic response [1], [5]. Thus, the specific absorption rate (SAR), a key parameter for magnetic hyperthermia applications of nanoparticles, is directly proportional to M_s^2 [6], [7]. Likewise, particle heating efficiency can be used to activate thermo-sensitive polymers for drug delivery applications [8], [9]. In magnetic imaging diagnosis, particle size also plays an important role in relation to the longitudinal and transverse relaxation times [10]. In addition, improving the magnetic properties of LSMO nanoparticles, as well as coating them with carboxyl-containing polymers such as PVA, opens new possibilities for use in RNA extraction by magnetic separation, contributing to the diagnosis of viral agents such as SARS-CoV-2, as recently proposed by [11].

Most of these applications have employed iron oxide magnetic nanoparticles (IOMNPs) such as Fe_3O_4 and $\gamma\text{-Fe}_2\text{O}_3$. Several works have demonstrated that nanoparticles with a lower Curie temperature favor heating in the therapeutic range (41–47 °C) for magnetic hyperthermia applications [12]–[15]. Typical IOMNPs exhibit a Curie temperature of around 858 K [14], [16], while manganite nanoparticles ($\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$) fall within the range of 350–370 K [15]. This can be tuned by means of the Sr content [17]–[19], which makes LSMO a promising family of compounds for biomedical applications.

Lanthanum manganites (LaMnO_3) are complex manganese oxides with a perovskite-like crystalline structure, which can be understood as a set of regular MnO_6 octahedra in contact with each other through their vertices. In this structure, oxygen ions are located at the vertices, while La occupies the center of the cube. When La^{3+} ions are partially replaced by divalent ions such as Sr^{2+} in a specific proportion, LSMO perovskite is formed, and a double exchange interaction between Mn^{3+} and Mn^{4+} ions takes place, favoring the change in crystalline phase from an orthorhombic structure (Pbnm) to a rhombohedral one (R3c) [4], [20]–[22], with the latter being responsible for the structural, electrical, and magnetic properties of LSMO [23], [24].

The use of LSMO nanoparticles in magnetic hyperthermia is possible due to their biocompatibility. Several studies on these nanoparticles have revealed a low cytotoxicity, making them suitable for biomedical applications, particularly for magnetic hyperthermia. *In vitro* cytotoxicity tests conducted

on HeLa cell lines using MTT assays demonstrated a concentration-dependent effect, with approximately 92% cell viability at 1 mg/mL and 63% at 5 mg/mL, suggesting minimal toxicity at lower concentrations [25]. This low cytotoxicity can be attributed to the hydroxyl group coating on the nanoparticles, which enhances their biocompatibility. Similar results were reported using concentrations of 0.66, 0.9, and 1.02 mg/mL after 96 hours, demonstrating that LSMO nanoparticles significantly reduce cell viability in a time- and dose-dependent manner; the viability decreased by 69, 82, and 86%, respectively [26]. Likewise, LSMO nanoparticles, particularly those with silica coatings, have shown promising biocompatibility and controlled cytotoxicity, which makes them suitable for magnetic hyperthermia cancer therapy. *In vitro* biocompatibility assessments using murine fibroblast (L929) cells indicated a high viability, exceeding 70% at 1 mg/mL and reaching over 93% at 250 $\mu\text{g/mL}$, which again confirms these nanoparticles' minimal toxicity at lower concentrations [27].

This work analyzes the influence of synthesis pathways on the magnetic and morphological properties of LSMO nanoparticles. Additionally, it explores how these properties affect heat generation in radiofrequency fields, with the aim of proposing LSMO nanoparticles as heat mediators in hyperthermia experiments for cancer treatment. Studies on magnetic nanoparticles (MNPs) with a LSMO perovskite-type structure in the field of hyperthermia are still scarce, which justifies further exploration.

To this effect, three different synthesis methods, including the traditional sol-gel process [28], the ceramic method [29], and the Pechini method [30, 31] were employed to prepare LSMO nanoparticles. In the ceramic method, the metal oxides or carbonates are used as precursors, which are mixed and calcined at high temperatures to form nanoparticles. Generally, the resulting compounds have an inhomogeneous size distribution and impurity phases [29]. In wet chemical approaches like the Pechini and sol-gel methods, the formation of nanoparticles occurs during the decomposition of previously formed organic-inorganic complexes. These processes provide smaller nanoparticles, along with a better control of stoichiometry, size, and morphology as well as fewer defects [32]. Once the nanoparticles had been characterized, we analyzed the influence of these synthesis methods on their structural, morphological, and magnetic properties.

To bridge the knowledge gap between LSMO nanoparticles and their potential biomedical applications, we performed a morphological, structural, and magnetic characterization of synthesized LSMO nanoparticles, with the aim of identifying the method that offers the most homogeneous LSMO nanoparticles as well as improved magnetic properties.

This document is organized as follows. In the *Experimentation* section, the methods used are presented. The *Results and discussion* section outlines the results and correlates the properties obtained. This section is divided into structural,

thermal, and magnetic characterization, concluding with specific absorption rate (SAR) simulations to examine the application of LSMO nanoparticles in hyperthermia. Finally, the main conclusions of this work are presented.

Experimentation

Synthesis methods

Pechini's method

Citric acid and ethylene glycol (1:4 molar ratio) were initially mixed for 30 min under constant stirring at 80 °C in a reflux system. Subsequently, high purity nitrates were added in a 1:1 molar ratio (citrate:cation), i.e., $\text{La}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ (99.99%), $\text{Sr}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (99.9 %) and $\text{Mn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (99.9 %), which had been previously dissolved in water, and heated at 80 °C for an additional 30 min. Once a pH of 5.0 was reached, it was maintained by adding NH_4OH . Esterification was performed in an electric oven at 140 °C, which, in turn, enabled the removal of traces of H_2O . The result was heated to 180 °C to remove the excess of ethylene glycol released during polymerization and oligomerization. The obtained solid was then calcined at 250 °C for 8 h. Finally, the powders were calcined at 650 °C for 12 h [33]. The sample synthesized through this method was labeled as M1.

Ceramic method

In this method, La_2O_3 (99.99%), $\text{Sr}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (99.9%), and MnO_2 (99%) precursors were mixed in stoichiometric quantities and grinded in agate mortar for 45 min. The obtained powders were calcined in a muffle at 1100 °C for 12 h, with subsequent sintering at the same temperature for an additional 12 h [34]. The sample synthesized by this route was labeled as M2.

Sol-gel method

In the sol-gel synthesis of LSMO nanoparticles, the required molar amounts of the metal salts $\text{La}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ (99.99%), $\text{Sr}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (99.9%), and $\text{Mn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (99.9 %) were dissolved in bi-distilled water, and then a mixture of citric acid:ethylene glycol (1:4) was added to the solution. The molar ratio between the salts and the mixture was 1:10. The solution was heated to 80 °C under vigorous stirring until a polymer gel was formed as a result of the poly-esterification reaction, which was then pyrolyzed at 200 °C. The obtained powder was further heated at 400 °C for 2 h and at 650 °C for 12 h [24], [35]. This solid was labeled as M3.

Structural, morphologic, and magnetic characterization

The structural properties of the LSMO powders were studied using X-ray diffraction (XRD). The samples were irradiated with Cu K α radiation ($\lambda=1.5406$ Å) from an

X-ray source powered at constant voltage (40 kV) and constant current (40 mA). Diffraction data were acquired with a LynxEye linear detector in a 2θ range of 2.000-69.999°, with a step of 0.024°. A qualitative analysis by comparison with the PDF-2(2016) diffraction dataset was conducted, and a quantitative analysis was carried out by fitting the observed diffraction patterns with the inorganic crystal structure database (ICSD) by means of the Rietveld method. The reported percentages correspond to the relationship between the polycrystalline phases, quantified without considering any content of amorphous material. Additionally, the crystal size (d_{XRD}) was estimated from XRD data via the Scherrer equation.

Particle size and morphology were studied using small-angle X-ray scattering (SAXS). To this effect, SAXS patterns were obtained in the National Synchrotron Light Laboratory at Campinas, Brazil. The powder samples were sealed with Kapton tape into a 1 mm wide sample holder. For the SAXS measurements, the samples were illuminated with monochromatic X-ray radiation ($\lambda=1.489$ Å), and the resulting 2-D scattering patterns were recorded using a MAR165-CCD camera. To achieve a q -range of 0.14-5.09 nm^{-1} a sample-detector length of 878.8 mm was used. An empty sample holder was also measured in order to subtract the Kapton contribution to the sample's scattering.

The obtained SAXS patterns were fitted using the mathematical model for the scattered intensity (I) of spherical particles as a function of the wave vector (q), as expressed in Eq. (1).

$$I(q) = \int_0^\infty f(r) F^2(r, q, \Delta\eta) dr + I_B \quad (1)$$

where $f(r)$ is the lognormal size distribution of primary particles of radius r , characterized by the median radius (r_o) and the standard deviation (σ) for the Gaussian distribution of the variable $u = \ln(r/r_o)$.

$F(r, q, \Delta\eta)$ is the spherical form factor [36] that represents the scattering amplitude produced by spherical particles of radius r , as given in Eq. (2)

$$F(r, q, \Delta\eta) = 4\pi r^3 \Delta\eta \left(\frac{\sin(qr) - qr \cos(qr)}{(qr)^3} \right) \quad (2)$$

where $\Delta\eta$ is the electronic contrast difference between the particle and the matrix, and I_B is the background intensity. From the fitted parameters (r_o and σ), the mean particle radius ($R_{\text{SAXS}} = r_o e^{\sigma^2/2}$) and the standard deviation ($sd = R_{\text{SAXS}} \sqrt{\sigma^2 - 1}$) were calculated. Note that the fitting processes were carried out using SASfit software [37].

Thermal stability was assessed by means of thermogravimetric analysis (TGA) curves using a TA Instruments TGA-Q500. The equipment was previously calibrated using the Curie temperature of a nickel sample (99.99% purity). In order to

avoid oxidation during the thermal process, the oven was previously purged with a constant flux of nitrogen for 30 min. For thermal decomposition analysis, the sample mass was measured throughout the temperature ramp (from 25 to 800 °C) at a constant rate of 5 °C/min under a nitrogen atmosphere (60 mL/min flux).

The magnetic characterization of each sample was carried out using a vibrating sample magnetometer (VSM-VersaLab by Quantum Design). Isothermal specific magnetization (M) loops under an applied magnetic field (H) with a maximum H of 3.0 T for different temperatures were taken from a known mass of powder samples. The selected temperatures ranged between 50 and 325 K. For magnetization cycles analysis, the particle mean magnetic moment ($\langle\mu\rangle$), the saturation magnetization (M_s), and coercive field (H_c) were calculated.

A flowchart of our methodology is presented in Fig. 1.

Results and discussion

XRD and SAXS analyses

The XRD patterns are shown in Fig. 2a. For all samples, the most intense diffraction peaks were indexed for a rhombohedral structure with the space group R3, corresponding to the crystalline structure of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ [1]–[3].

The quantitative analysis performed using the Rietveld method revealed that the main perovskite crystal phase was $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ for samples M1 and M3 (74.2 and 30.6%). The diffraction peaks can be indexed using a cubic crystalline structure. The calculated lattice parameters were $a = 5.50(9)$ Å, $5.52(1)$ Å, and $5.56(6)$ Å for samples M1, M2, and M3, respectively. Other crystalline phases for M1 included $x = 0.33$ (21.4%) and $x = 0.4$ (4.4%). As reported in the literature [37], [38], a lower incidence of undesired crystalline phases might even be achieved through longer sintering at the same synthesis temperature. For sample M2, the phase corresponding to $x =$

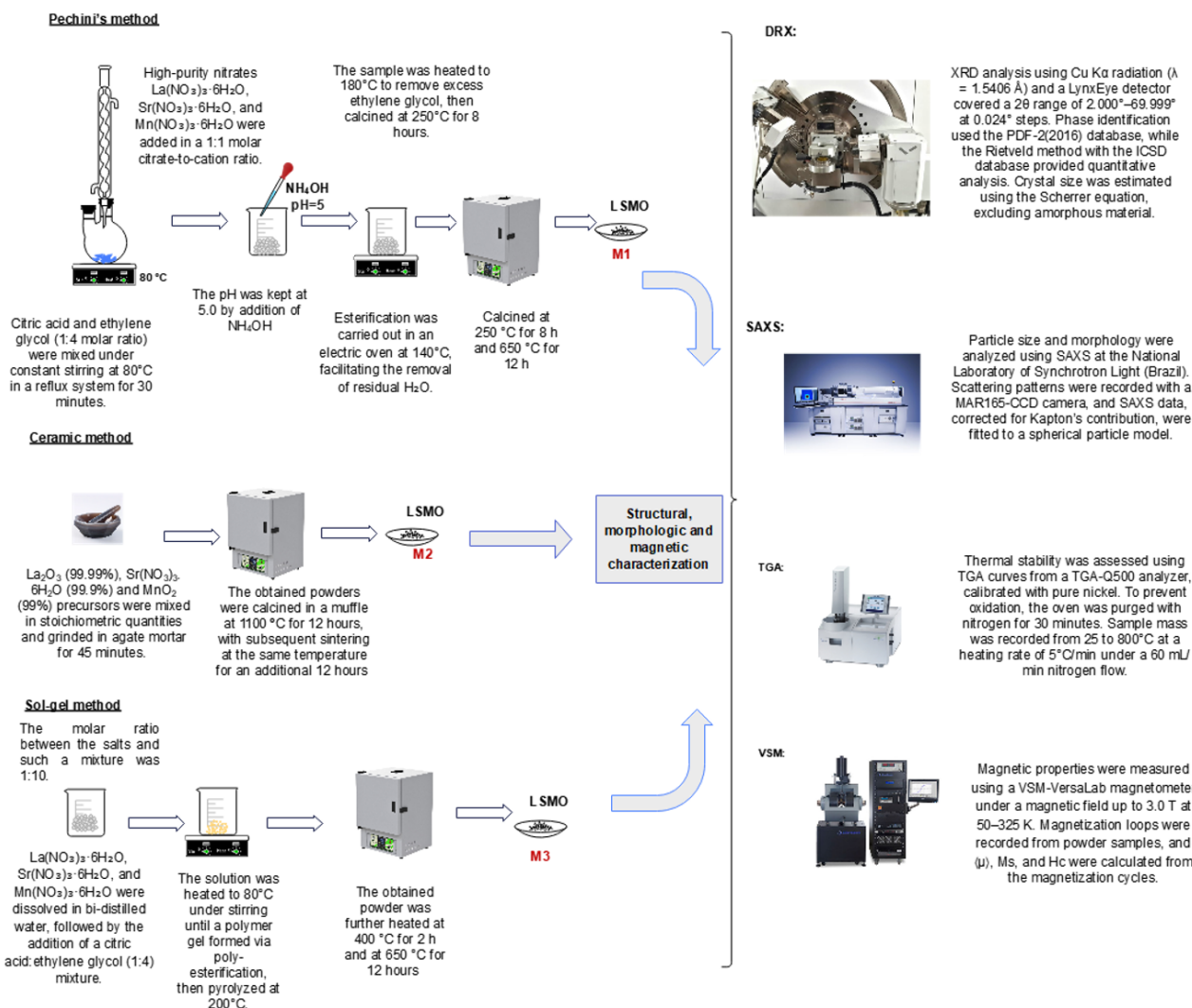


Figure 1. Methodology flowchart

Source: Authors

0.3 was not observed, and the main crystalline phase was $x = 0.33$ (38.5%). For this sample, precursors and intermediate compounds such as $\text{La}(\text{MnO}_3)$ (27.1%) and $\text{La}(\text{OH})_3$ (15.2%) were observed alongside Sr and La silicates (20.2 %). In Fig. 2b, the diffraction peaks for $\text{La}(\text{MnO}_3)$, $\text{La}(\text{OH})_3$, and the silicates are labeled with +, o, and *, respectively.

In the case of sample M3, $\text{Sr}(\text{CO}_3)$ (32.4%) and $\text{LaSr}_4\text{Mn}_5\text{O}_{13}$ (37.0%) were observed. In Fig. 2c, the diffraction peaks for $\text{Sr}(\text{CO}_3)$ are labeled with arrows.

The above indicates that the best synthesis procedure to obtain $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ was the Pechini method (M1). Sample M3, synthesized via the sol-gel method, reported a low reaction yield and low crystallinity, which can be attributed to the fact that the pH of the reaction was not controlled.

Some studies have stated that the homogeneity and stability of metal citrate solutions may largely depend on the pH, which is particularly important in systems with several different metals. This is useful in optimizing the formation of stable species of metal citrate and avoiding the precipitation of individual hydroxides [39], [40]. This parameter is relevant in controlling the sample particle size [41].

Once the samples had been structurally analyzed, SAXS was used to prove whether the samples were composed of particles in the nanometric range. The measured SAXS patterns, the curves fitted using Eq. (1), and the best-fitted particle mean size (R_{SAXS}) and size standard distribution (s.d) values are shown in Fig. 2 and Table I.

The SAXS curve fitting revealed that the spherical model was sufficient to fit all the data. The wide size distribution obtained from the fitting is presented in the form of insets in Fig. 3 (b, c, d), indicating a polydispersity of over 55% for samples M1 and M3, as well as over 93% for sample M2. The mean radius (R_{SAXS}) is in agreement with the crystal size (d_{XRD}) measured using the XRD peak at 32.7° and the Scherrer equation – if each crystal is assumed to be a sphere of diameter d_{XRD} .

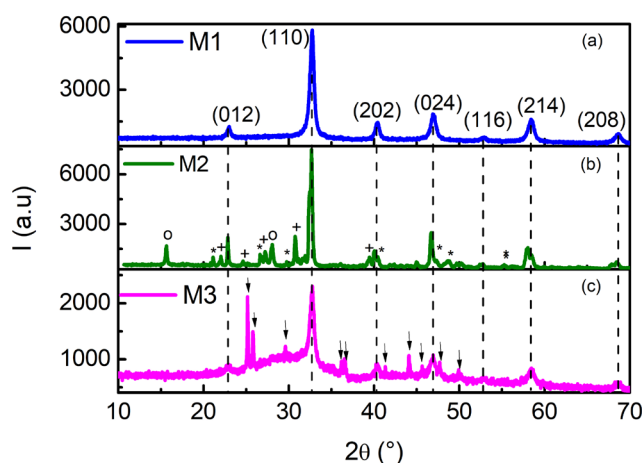


Figure 2. X-ray diffraction patterns for all the samples. The indexed peaks are related with the $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ crystalline structure.

Source: Authors

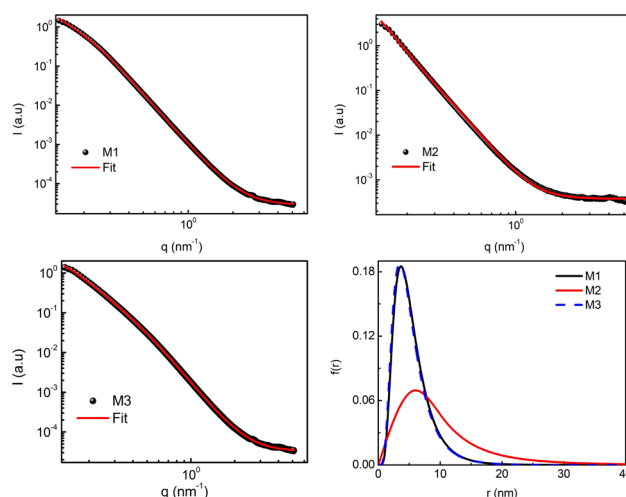


Figure 3. SAXS pattern for samples M1, M2, and M3. Solid lines correspond to the best-fit curves obtained using Eq. (1). The particle number size distribution obtained via the SAXS fitting procedure is also presented.

Source: Authors

Thermal degradation

TGA curves are shown in Fig. 4. These analyses were performed on the final products after calcination. Fig. 3a presents the percent mass loss vs. the temperature for all the samples.

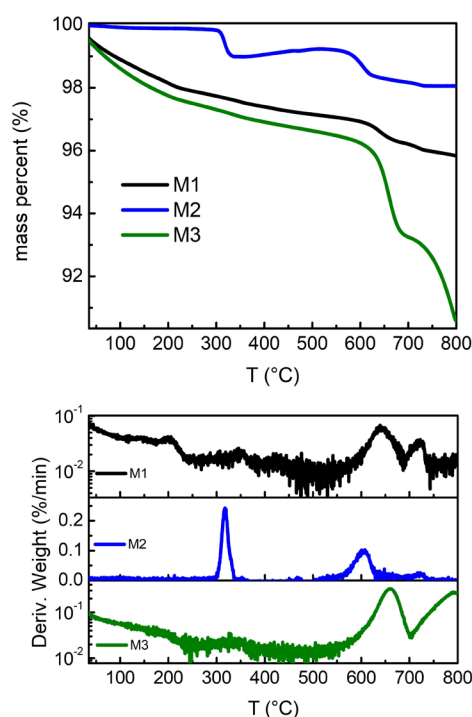


Figure 4. Thermal decomposition of all samples. a) Mass loss curves. Arrows indicate the two main thermal decomposition processes. b) Differential mass loss curves.

Source: Authors

In this figure, sample mass losses take place as the temperature increases. For samples M1 and M3, two different thermal events are observed (indicated with arrows in Fig. 3a). First, a slight mass reduction between room temperature and 550 °C, followed by an abrupt mass loss between 550 and 700 °C. This second reduction is clearly more pronounced for sample M3. Meanwhile, sample M2 exhibits an abrupt mass reduction between 250 and 300 °C, as well as a second reduction in the range of 550 to 650 °C. In Fig. 4b, these stages are more clearly seen as peaks, whose centers correspond to the mean decomposition temperatures.

The observed thermal behavior is related to the chemical composition of the samples, considering that each chemical phase decomposes at a different temperature. Hence, the initial slight mass loss of samples M1 and M3 may be associated with sample dehydration.

In Fig. 5, the second derivative of the TGA curve (d^2W/dT^2) serves as a powerful analytical tool that enhances the resolution of thermal events by revealing inflection points not readily apparent in the original or first-derivative data [42]. For all the samples, the second derivative exhibits multiple transitions corresponding to distinct decomposition steps, indicating the breakdown of the material into several subphases. These subtle changes provide a precise delineation of the onset, peak, and termination temperatures of each thermal event, allowing for a more comprehensive understanding of the material's thermal stability and compositional complexity. The rapid drop in the second derivative at the onset of the first decomposition event (25-100 °C) indicates a sudden, pronounced change in the reaction kinetics, consistent with the beginning of thermal decomposition. This behavior suggests that a specific subphase, which is less thermally stable, undergoes an abrupt breakdown, leading to a swift acceleration of the mass loss rate, typically associated with physisorbed water. Essentially, it marks a well-defined thermal event where the material's structure rapidly transitions from a stable state to a decomposing phase, highlighting the onset of a critical degradation step [43].

In Fig. 4b, a peak related to the thermal decomposition of unreacted precursors [44], [45] was observed for all the samples between 250 and 300 °C. For sample M2, this peak was more pronounced, and the mass loss corresponded to 1.0%. In samples M1 and M3, although this peak was also observed, the mass loss was rather negligible. This correlates well with the unreacted precursors observed via XRD for sample M2. The peak observed between 550 and 700 °C corresponds to the thermal decomposition of organic remnants from the synthesis process [2]. In this stage, mass losses of 2.3, 0.9, and 7.1 were recorded for samples M1, M2, and M3, respectively. The mass percentages after the thermal process were 95.85, 98.05, and 90.61% for samples M1, M2, and M3, which could correspond to dehydrated LSMO nanoparticles [2].

Magnetic properties

Fig. 6 shows the magnetization curves at different temperatures for each sample. The magnetization (M) was normalized using the $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ phases (LSMO) present in each sample according to the XRD analyses. This value increased with the magnetic field (H) until saturation was reached (M_s). Likewise, the saturation decreased when the temperature increased, as expected for magnetic systems [1], [5]. The insets in Fig. 4 show the coercive field (H_c) for each temperature. As expected, the coercive field did tend to zero when the temperature increased [46].

Fig. 6d presents the magnetization curves at 300 K. These curves were fitted using the Langevin equation convoluted with a lognormal distribution of magnetic moments ($f(\mu)$):

$$M = \int_0^\infty N\mu \left[\text{Coth} \left(\frac{\mu\mu_0 H}{k_B T} \right) - \frac{k_B T}{\mu\mu_0 H} \right] f(\mu) d\mu + \chi_{par} H \quad (3)$$

In this model, μ is the particle magnetic moment, N is the mass density of magnetic moments, χ_{par} represents the susceptibility associated with paramagnetic contributions to the magnetization process, and $\mu_0 = 4\pi \times 10^{-7} \text{ Tm/A}$ denotes the vacuum magnetic permeability.

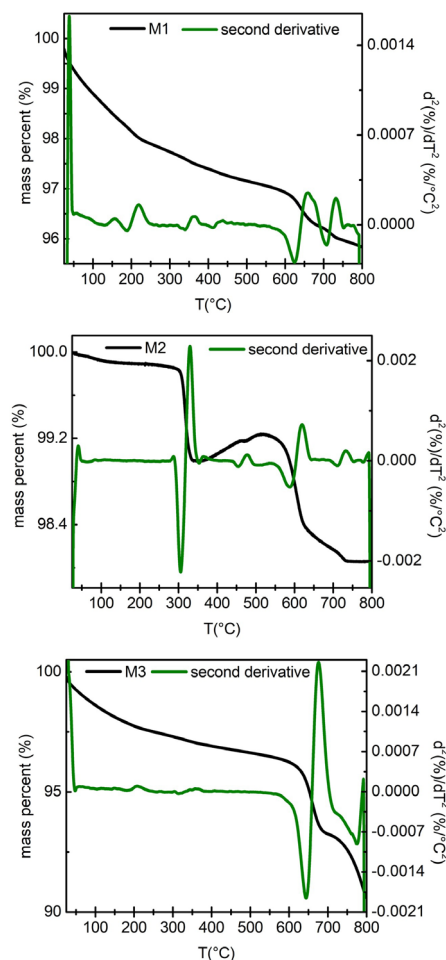


Figure 5. Second-derivative analysis of TGA curves

Source: Authors

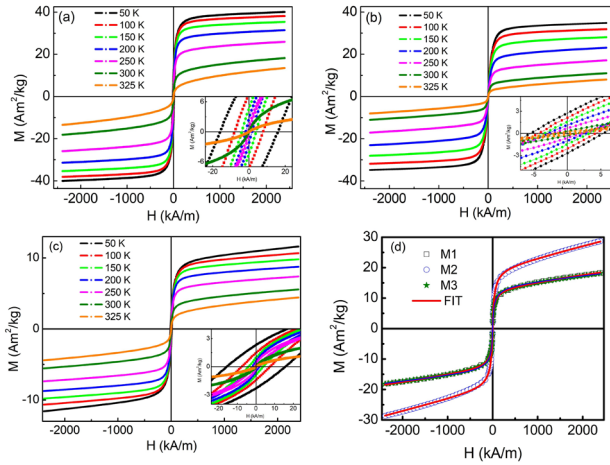


Figure 6. M vs. H curves at different temperatures, measured by means of VSM, for samples a) M1, b) M2, and c) M3. Insets: zoom around the zero-field. d) Magnetization curve at 300 K for samples M1, M2, and M3. Continuous lines represent best-fit curve obtained using Eq. (3).
Source: Authors

From the fittings, the magnetic saturation was calculated as $M_s = N\langle\mu\rangle$, where $\langle\mu\rangle$ is the mean magnetic moment calculated from the fitted lognormal distribution $f(\mu)$. The results showed that the 300 K magnetization curves can be fully fitted using Eq. (3), which indicates that, at 300 K, the LSMO nanoparticles behaved as super-paramagnetic [47].

The fitted $\langle\mu\rangle$ and M_s values are presented in Table 1. Both parameters were similar for samples M1 and M3. For sample M2, $\langle\mu\rangle$ was around 1.6 times the value obtained for M1. This value was equal to the ratio between the crystallite volumes of samples M1 and M2, which indicates that, the higher the particle size, the higher the magnetic moment. This behavior has already been observed for similar $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles [48]. Eq. (3) was also used to determine M_s and H_c for each temperature.

The fitting procedure was enough to fit the curve, even for the lowest temperature. In order to determine the coercive field from the fittings, the Langevin equation was displaced for $\pm H_c$. This value corresponds to a fitting parameter.

Table 1. Structural, morphological, and magnetic parameters of the samples. Here, d_{XRD} is the crystallite size obtained from X-ray diffraction, R_{SAXS} is the mean particle radius (with a standard distribution s.d.) obtained from SAXS, $\langle\mu\rangle$ is the mean particle magnetic moment (with a standard distribution s.d.), M_s is the magnetic saturation, H_c is the coercive field, T_B is the magnetic blocking temperature, K_{eff} is the magnetic anisotropy constant, M_0 is the magnetic saturation extrapolated to 0 K, and T_c is the Curie temperature.

Sample	d_{XRD} (nm)	$R_{\text{SAXS}} \pm \text{sd}$ (nm)	$\langle\mu\rangle \pm \text{sd}$ (μ_B)	M_s ($\text{Am}^2/\text{kg}_{\text{LSMO}}$)	H_c (kA/m)	T_B (K)	K_{eff} (J/m^3)	M_0 ($\text{Am}^2/\text{kg}_{\text{LSMO}}$)	α	T_c (K)
M1 Pechini	14.5 ± 0.7	5.4 ± 3.0	6570 ± 42	13.6 ± 0.9	0	189 ± 18	11175 ± 1281	40.3 ± 0.6	2.49 ± 0.02	358 ± 28
M2 Ceramic	17.0 ± 0.7	8.2 ± 7.7	10534 ± 550	19.8 ± 0.1	1.45 ± 0.04	458 ± 4	2740 ± 27	89.7 ± 2.5	1.22 ± 0.06	353 ± 123
M3 Sol-gel	14.6 ± 0.7	5.3 ± 3.0	6690 ± 77	14.4 ± 0.6	0.27 ± 0.01	186 ± 20	11168 ± 1574	33.6 ± 0.5	2.71 ± 0.02	367 ± 28

Source: Authors

Fig. 7 presents the thermal behavior of the coercive field and the magnetic saturation. In Fig. 7a, the behavior of H_c vs. $T^{1/2}$ is shown. The coercive field decreased with temperature until a value of zero was reached. For superparamagnetic nanoparticles, a linear behavior is expected – it is observed at lower temperatures. The theoretical temperature at which $H_c = 0$ is known as the *blocking temperature* (T_B). When $T > T_B$, nanoparticles are expected to behave as superparamagnetic particles, and, when $T < T_B$, the magnetic anisotropy energy is higher than the thermal energy. Then, the thermal fluctuation ceases, and the particles' magnetic moments become blocked.

The H_c vs. $T^{1/2}$ curves were fitted using the thermal-activated Stoner-Wohlfarth model [49]–[51]:

$$H_c = H_k \left(1 - \left(\frac{T}{T_B} \right)^{1/2} \right) \quad (4)$$

In Eq. (4), $H_k = 0.96 \frac{K_{\text{eff}}}{\mu M_{s0K}}$ is the anisotropy field, K_{eff} is the effective magnetic anisotropy constant, μ denotes the particle magnetic moment, M_{s0K} is saturation magnetization at $T = 0$ K, and T_B is the blocking temperature. It has been demonstrated that this method enables a good approximation to the blocking temperature [7],[52].

The fitted values of T_B and K_{eff} are shown in Table 1. They are in good agreement with those measured for similar nanoparticles [48], [53], [54]. In Fig. 7b, the fitted M_s is plotted against temperature. For all the samples, M_s decreased with temperature. This behavior is typical of ferromagnetic materials. The temperature at which $M_s = 0$ is called the *Curie temperature* (T_c). In order to obtain information from the M_s vs. T curves, these values were fitted using the spin-wave model for the dependence of M_s on the temperature T [55]:

$$M_s(T) = M_0 (1 - B_0 T^\alpha) \quad (4)$$

where M_0 is the saturation magnetization at 0 K and B_0 is a constant whose value depends on the exchange integral J . The coefficient α generally takes the value of 3/2 for bulk

ferromagnetic materials and, for some bulk spinel ferrites, α ranges between 0.2 and 2.0. The fitted values for M_0 and α are presented in Table 1 and the best-fit curves are shown in Fig. 7b.

The fitted value for M_0 was similar for samples M1 and M3, and the value of M_0 found for sample M2 was similar to those reported for larger particles [42], [56]. The lower magnetization of samples M1 and M3 could be related to the formation of a magnetically frustrated layer on the particle's surface – these effects were more marked when the particle size was reduced [57].

The fitted α values were 2.49 for M1, 1.22 for M2, and 2.71 for M3. It has been demonstrated that the α value depends on particle size: the smaller the size, the higher the α value [48]. This was confirmed via XRD and SAXS size analysis: M2 exhibited the highest mean particle size in comparison with M1 and M3. The α value fitted for sample M2 was the nearest to $\alpha = 3/2$, indicating that the particles almost reached the bulk state.

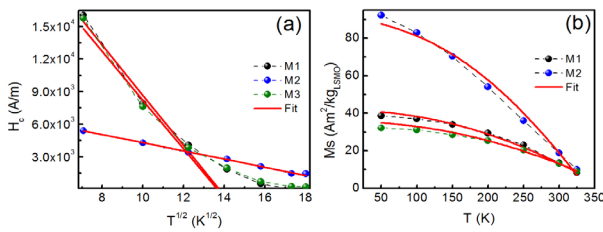


Figure 7. Thermal behavior of a) the coercive field (H_c) and b) the magnetic saturation

Source: Authors

From the fittings, the value of T_c was calculated as $T_c = \left(\frac{1}{B_0} \right)$. These values are presented in Table 1, using $B_0 = 4.28 \times 10^{-7} \text{ K}^{-2.49}$, $7.54 \times 10^{-4} \text{ K}^{-1.22}$, and $1.11 \times 10^{-7} \text{ K}^{-2.71}$ for samples M1, M2, and M3, respectively. The results are in good agreement with those expected for $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ nanoparticles [42], [46], [48].

Specific absorption rate calculations for magnetic hyperthermia

Once the structural and magnetic characterization had been performed, these results were used to study, by means of simulation, the possible applications of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles in magnetic hyperthermia. As explained elsewhere, if the magnetic properties and the size distribution are known, it is possible to numerically simulate the SAR [7], [49]. One of the most important biomedical applications of magnetic nanoparticles is magnetic hyperthermia (MH).

In MH, a set of magnetic nanoparticles is submerged in a radiofrequency (RF) magnetic field. These nanoparticles absorb energy from the field and dissipate it to the medium as heat. This heat can be used to kill cancer cells. The key

parameter to measure nanoparticles' ability to transduce magnetic energy into heat is the SAR. Said heat is produced by magnetic relaxation processes governed by the imaginary part of the magnetic susceptibility (χ''). Thus, the SAR can be modeled as follows:

$$\text{SAR} = \mu_0 \pi f H_0^2 \int_0^\infty \chi''(f, \tau) g(\tau) d\tau \quad (6)$$

where $\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$ is the vacuum magnetic permeability; f and H_0 are the RF field frequency and amplitude, respectively; $g(\tau)$ is the Gaussian distribution of the relaxation times (τ); and $\chi''(f, \tau) = \frac{2\pi f \tau}{1 + (2\pi f \tau)^2} \chi_0$, with $\chi_0 = \frac{\mu_0 M_s^2 V}{3k_B T}$ being the low field d.c. specific susceptibility.

τ is defined as the time it takes for the magnetic moment to switch from one equilibrium position to another. If the particle is immersed in a solid matrix, a whole rotation is not allowed, and magnetic moment switching takes place internally. This is known as the Néel relaxation process. The associated relaxation time is

$$\tau = \tau_0 e^{\frac{K_{\text{eff}} V}{k_B T}} \quad (7)$$

where $\tau_0 = 10^{-10} \text{ s}$ is the characteristic time, K_{eff} is the effective magnetic anisotropy constant, V is the particle volume, k_B is the Boltzmann constant, and T is the temperature. As observed, a particle size distribution ($g(r)$) generates a relaxation time distribution ($g(\tau)$).

Thus, the behavior of SAR vs. H_0 at 100 kHz can be calculated. The results are presented in Fig. 8.

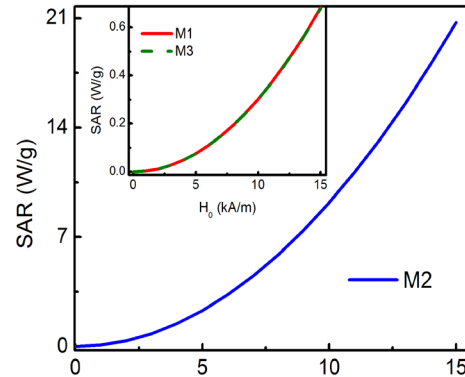


Figure 8. Specific absorption rate calculations vs. RF magnetic field amplitude at 100 kHz

Source: Authors

Moreover, the mean Néel relaxation time can be calculated: for samples M1 and M3, $\tau \approx 0.2 \text{ ns}$, and, for M2, $\tau = 0.1 \text{ ns}$. This indicates that, the lower the relaxation time, the higher the SAR.

As observed in Fig. 8, sample M2 exhibits the best SAR behavior, which means that this sample is a good candidate for biomedical applications. In addition, M2 has the highest saturation magnetization and the lowest Curie temperature.

From a chemical point of view, M2 is composed of $\text{La}_{0.66}\text{Sr}_{0.33}\text{MnO}_3$ instead of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$, the main chemical phase for samples M1 and M3. This slight increase in the Sr content enhances the magnetic properties and the SAR of M2.

Conclusions

In this work, three different ways to synthesize $\text{La}_x\text{Sr}_{1-x}\text{MnO}_3$ magnetic nanoparticles were tested. Although the precursors were fixed to obtain the phase $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$, this was only possible through the Pechini and sol-gel methods. In contrast, the ceramic method yielded the phase $\text{La}_{0.66}\text{Sr}_{0.33}\text{MnO}_3$. The structural characterization revealed that unreacted precursors and other manganite phases (different from $x=0.3$ and $x=0.33$) were present in all samples, implying low reaction yields, especially for the sol-gel and the ceramic methods. Meanwhile, the magnetic characterization revealed that a slight increase in Sr content leads to enhanced magnetic properties, making the sample prepared through the ceramic method (M2) a good candidate for biomedical applications, as confirmed by SAR calculations. Thus, this study demonstrated that the synthesis pathway influences the properties of LSMO nanoparticles and, consequently, their applications in magnetic hyperthermia. The results presented herein may inform improvements in the synthesis of LSMO nanoparticles for biomedical applications, especially regarding their magnetic properties. This could include magnetic hyperthermia and drug delivery. Future work is expected to address the cytotoxicity effects of LSMO nanoparticles in cell cultures and determine the heat generated in RF fields under different experimental conditions (*in vitro*, *ex-vivo*, and *in vivo*).

Acknowledgements

We acknowledge the financial support from Universidad CESMAG (Colombia). We would like to thank Universidad de Nariño, the X-ray Laboratory of Universidad Industrial de Santander, and the Brazilian Synchrotron Light Laboratory (LNLS) for refining the physicochemical characterization of the materials.

CRedit author statement

All authors: conceptualization, methodology, software, formal analysis, investigation, data curation, writing (original draft, review, and editing).

Conflicts of interest

The authors of this paper declare that they have no conflicts of interest.

References

- [1] N. D. Thorat, K. P. Shinde, S. H. Pawar, K. C. Barick, C. A. Betty, and R. S. Ningthoujam, "Polyvinyl alcohol: an efficient fuel for synthesis of superparamagnetic LSMO nanoparticles for biomedical application," *Dalt. Trans.*, vol. 41, no. 10, art. 3060, 2012. <https://doi.org/10.1039/c2dt11835a>
- [2] N. D. Thorat *et al.*, "Highly water-dispersible surface-functionalized LSMO nanoparticles for magnetic fluid hyperthermia application," *New J. Chem.*, vol. 37, no. 9, art. 2733, 2013. <https://doi.org/10.1039/c3nj00007a>
- [3] N. D. Thorat, V. M. Khot, A. B. Salunkhe, A. I. Prasad, R. S. Ningthoujam, and S. H. Pawar, "Surface functionalized LSMO nanoparticles with improved colloidal stability for hyperthermia applications," *J. Phys. D Appl. Phys.*, vol. 46, no. 10, art. 105003, Mar. 2013. <https://doi.org/10.1088/0022-3727/46/10/105003>
- [4] K. Navin and R. Kurchania, "A comparative study of the structural, magnetic transport and electrochemical properties of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ synthesized by different chemical routes," *Appl. Phys. A*, vol. 126, no. 2, art. 100, Feb. 2020. <https://doi.org/10.1007/s00339-019-3269-2>
- [5] P. V. Hendriksen, S. Linderoth, and P.-A. Lindgård, "Finite-size effects in the magnetic properties of ferromagnetic clusters," *J. Magn. Magn. Mater.*, vol. 104–107, pp. 1577–1579, Feb. 1992. [https://doi.org/10.1016/0304-8853\(92\)91461-2](https://doi.org/10.1016/0304-8853(92)91461-2)
- [6] R. E. Rosensweig, "Heating magnetic fluid with alternating magnetic field," *J. Magn. Magn. Mater.*, vol. 252, pp. 370–374, 2002. [https://doi.org/http://dx.doi.org/10.1016/S0304-8853\(02\)00706-0](https://doi.org/http://dx.doi.org/10.1016/S0304-8853(02)00706-0)
- [7] M. B. Fernández van Raap, D. F. Coral, S. Yu, G. A. Muñoz, F. H. Sánchez, and A. Roig, "Anticipating hyperthermic efficiency of magnetic colloids using a semi-empirical model: A tool to help medical decisions," *Phys. Chem. Chem. Phys.*, vol. 19, no. 10, pp. 7176–7187, 2017. <https://doi.org/10.1039/C6CP08059F>
- [8] M. Soleymani, M. Edrissi, and A. Mohammad Alizadeh, "Thermosensitive polymer-coated $\text{La}_{0.73}\text{Sr}_{0.27}\text{MnO}_3$ nanoparticles: Potential applications in cancer hyperthermia therapy and magnetically activated drug delivery systems," *Polym. J.*, vol. 47, pp. 797–801, 2015. <https://doi.org/10.1038/pj.2015.66>
- [9] B. T. Mai, S. Fernandes, P. B. Balakrishnan, and T. Pellegrino, "Nanosystems based on magnetic nanoparticles and thermo- or pH-responsive polymers: An update and future perspectives," *Acc. Chem. Res.*, vol. 51, no. 5, pp. 999–1013, May 2018. <https://doi.org/10.1021/acs.accounts.7b00549>
- [10] M. Di Marco, C. Sadun, M. Port, I. Guilbert, P. Couvreur, and C. Dubernet, "Physicochemical characterization of ultrasmall superparamagnetic iron oxide particles (USPIO) for biomedical application as MRI contrast agents," *Int. J. Nanomedicine*, vol. 2, no. 4, pp. 609–22, 2007. <http://www.ncbi.nlm.nih.gov/pubmed/18203428>
- [11] S. B. Somvanshi, P. B. Kharat, T. S. Saraf, S. B. Somvanshi, S. B. Shejul, and K. M. Jadhav, "Multifunctional nano-magnetic particles assisted viral RNA-extraction protocol for potential detection of COVID-19," *Mater. Res. Innov.*, vol. 25,

- no. 3, pp. 169–174, May 2020. <https://doi.org/10.1080/14328917.2020.1769350>
- [12] A. Jordan et al., "Presentation of a new magnetic field therapy system for the treatment of human solid tumors with magnetic fluid hyperthermia," *J. Magn. Magn. Mater.*, vol. 225, no. 1–2, pp. 118–126, Jan. 2001. [https://doi.org/10.1016/S0304-8853\(00\)01239-7](https://doi.org/10.1016/S0304-8853(00)01239-7)
- [13] C. S. S. R. Kumar and F. Mohammad, "Magnetic nanomaterials for hyperthermia-based therapy and controlled drug delivery," *Adv. Drug Deliv. Rev.*, vol. 63, no. 9, pp. 789–808, Aug. 2011. <https://doi.org/10.1016/j.addr.2011.03.008>
- [14] N. D. Thorat et al., "Superparamagnetic gadolinium ferrite nanoparticles with controllable Curie temperature-cancer theranostics for MR-imaging-guided magneto-chemotherapy," *Eur. J. Inorg. Chem.*, vol. 2016, no. 28, pp. 4586–4597, Oct. 2016. <https://doi.org/10.1002/ejic.201600706>
- [15] S. Lotfi, S. Bahari, A. Bahari, and M. Roudbari, "Magnetic performance and evaluation of radiofrequency hyperthermia of perovskite $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$," *J. Supercond. Nov. Magn.*, vol. 31, no. 7, pp. 2187–2193, Jul. 2018. <https://doi.org/10.1007/s10948-017-4475-9>
- [16] T. N. Brusentsova, N. A. Brusentsov, V. D. Kuznetsov, and V. N. Nikiforov, "Synthesis and investigation of magnetic properties of Gd-substituted Mn–Zn ferrite nanoparticles as a potential low-TC agent for magnetic fluid hyperthermia," *J. Magn. Magn. Mater.*, vol. 293, no. 1, pp. 298–302, May 2005. <https://doi.org/10.1016/j.jmmm.2005.02.023>
- [17] M. H. Ehsani, P. Kameli, M. E. Ghazi, F. S. Razavi, and M. Taheri, "Tunable magnetic and magnetocaloric properties of $\text{La}_{0.6}\text{Sr}_{0.4}\text{MnO}_3$ nanoparticles," *J. Appl. Phys.*, vol. 114, no. 22, art. 223907, Dec. 2013. <https://doi.org/10.1063/1.4846758>
- [18] R. T. Salakhova et al., "The frequency dependence of magnetic heating for $\text{La}_{0.75}\text{Sr}_{0.25}\text{MnO}_3$ nanoparticles," *J. Magn. Magn. Mater.*, vol. 470, pp. 38–40, Jan. 2019. <https://doi.org/10.1016/j.jmmm.2017.11.126>
- [19] R. Epherre et al., "Manganite perovskite nanoparticles for self-controlled magnetic fluid hyperthermia: about the suitability of an aqueous combustion synthesis route," *J. Mater. Chem.*, vol. 21, no. 12, art. 4393, 2011. <https://doi.org/10.1039/c0jm03963b>
- [20] A. J. Campbell, G. Balakrishnan, M. R. Lees, D. McK. Paul, and G. J. McIntyre, "Single-crystal neutron-diffraction study of a structural phase transition induced by a magnetic field in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$," *Phys. Rev. B*, vol. 55, no. 14, pp. R8622–R8625, Apr. 1997. <https://doi.org/10.1103/PhysRevB.55.R8622>
- [21] P. Y. Vanina, A. A. Naberezhnov, V. I. Nizhankovskii, and R. F. Mamin, "Temperature evolution of the magnetic properties of lanthanum-strontium manganites," *St. Petersburg. Polytech. Univ. J. Phys. Math.*, vol. 2, no. 3, pp. 175–180, Oct. 2016. <https://doi.org/10.1016/j.sjpm.2016.08.001>
- [22] M. Eshraghi and P. Kameli, "Structural and magnetic properties of microwave assisted sol-gel synthesized $\text{La}_{0.9}\text{Sr}_{0.1}\text{MnO}_3$ manganite nanoparticles," *ISRN Nanotechnol.*, vol. 2014, pp. 1–6, 2014. <https://doi.org/10.1155/2014/867139>
- [23] L. Dimesso, "Pechini processes: An alternate approach of the sol-gel method, preparation, properties, and applications," in *Handbook of Sol-Gel Science and Technology*, L. Klein, M. Aparicio, and A. Jitianu, Eds., Cham: Springer International Publishing, 2016, pp. 1–22. https://doi.org/10.1007/978-3-319-19454-7_123-1
- [24] Y. Shlapa, S. Solopan, A. Belous, and A. Tovstolytkin, "Effect of synthesis method of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ manganite nanoparticles on their properties," *Nanoscale Res. Lett.*, vol. 13, no. 1, art. 13, Dec. 2018. <https://doi.org/10.1186/s11671-017-2431-z>
- [25] M. Aneja, A. Tovstolytkin, and G.S. Lotey, "Superparamagnetic LaSrMnO_3 nanoparticles for magnetic nanohyperthermia and their biocompatibility," *J. Magn. Magnet. Mater.*, vol. 442, pp. 423–428, 2017. <https://doi.org/10.1016/j.jmmm.2017.06.106>
- [26] N. K. Prasad, K. Rathinasamy, and D. Bahadur, " T_c -tuned biocompatible suspension of $\text{La}_{0.73}\text{Sr}_{0.27}\text{MnO}_3$ for magnetic hyperthermia," *J. Biomed. Mater. Res. Part B App. Biomater.*, vol. 85B, pp. 409–416, 2007. <https://doi.org/10.1002/jbm.b.30959>
- [27] A.B. Tewari, R. Sharma, and D. Sharma, "Magnetic hyperthermia cancer therapy using rare earth metal-based nanoparticles: An investigation of Lanthanum strontium Manganite's hyperthermic properties," *Res. Eng.*, vol. 20, art. 101537, 2023. <https://doi.org/10.1016/j.rinen.2023.101537>
- [28] T. Wang et al., "Mechanochemical effects on microstructure and transport properties of nanocrystalline $\text{La}_{0.8}\text{Na}_{0.2}\text{MnO}_3$ ceramics," *J. Alloys Comp.*, vol. 458, no. 1–2, pp. 248–252, Jun. 2008. <https://doi.org/10.1016/j.jallcom.2007.04.023>
- [29] S. Daengsakul et al., "Magnetic and cytotoxicity properties of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($0 \leq x \leq 0.5$) nanoparticles prepared by a simple thermal hydro-decomposition," *Nanoscale Res. Lett.*, vol. 4, no. 8, pp. 839–845, Aug. 2009. <https://doi.org/10.1007/s11671-009-9322-x>
- [30] P. Kameli, H. Salamati, and A. Aezami, "Effect of particle size on the structural and magnetic properties of $\text{La}_{0.8}\text{Sr}_{0.2}\text{MnO}_3$," *J. Appl. Phys.*, vol. 100, no. 5, art. 053914, Sep. 2006. <https://doi.org/10.1063/1.2345036>
- [31] P. Dey and T. K. Nath, "Enhanced grain surface effect on the temperature-dependent behavior of spin-polarized tunneling magnetoresistance of nanometric manganites," *Appl. Phys. Lett.*, vol. 87, no. 16, art. 162501, Oct. 2005. <https://doi.org/10.1063/1.2089179>
- [32] A. O. Turkey, M. M. Rashad, A. M. Hassan, E. M. Elnaggar, and M. Bechelany, "Optical, electrical and magnetic properties of lanthanum strontium manganite $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ synthesized through the citrate combustion method," *Phys. Chem. Chem. Phys.*, vol. 19, no. 9, pp. 6878–6886, 2017. <https://doi.org/10.1039/C6CP07333F>
- [33] J. Mera, M. Mera, C. Cordoba, O. Paredes, and O. Morán, " $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles synthesized via the (Pechini) polymeric precursor method," *J. Supercond. Nov. Magn.*, vol. 26, no. 7, pp. 2553–2556, Jul. 2013. <https://doi.org/10.1007/s10948-012-1570-9>
- [34] H. A. Reshi and V. Shelke, "Grain size induced metal-insulator transition in $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ compounds," *J. Nano-Electron. Phys.*, vol. 5, no. 4, art. 04053, 2013. https://jnep.sumdu.edu.ua/en/full_article/1131
- [35] Z. Jiráček, J. Kuličková, V. Herynek, M. Maryško, J. Koktan, and O. Kaman, "Titania-coated manganite nanoparticles:

- Synthesis of the shell, characterization and MRI properties," *J. Magn. Magn. Mater.*, vol. 427, pp. 245–250, Apr. 2017. <https://doi.org/10.1016/j.jmmm.2016.10.097>
- [36] O. Glatter and O. Kratky, *Small angle x-ray scattering*. Cambridge, MA, USA: Academic Press, 1982.
- [37] J. Kohlbrecher, "SASfit: A program for fitting simple structural models to small angle scattering data," Paul Scherrer Institute, 2014. [Online]. Available: <http://kur.web.psi.ch/sans1/SANSSoft/sasfit.html>
- [38] M. H. Ehsani, M. E. Ghazi, and P. Kameli, "Effects of pH and sintering temperature on the synthesis and electrical properties of the bilayered $\text{LaSr}_2\text{Mn}_2\text{O}_7$ manganite prepared by the sol-gel process," *J. Mater. Sci.*, vol. 47, no. 15, pp. 5815–5822, Aug. 2012. <https://doi.org/10.1007/s10853-012-6481-4>
- [39] P. Vaqueiro and M. A. López-Quintela, "Influence of complexing agents and pH on yttrium-iron garnet synthesized by the sol-gel method," *Chem. Mater.*, vol. 9, no. 12, pp. 2836–2841, Dec. 1997. <https://doi.org/10.1021/cm970165f>
- [40] A. Abreu, S. M. Zanetti, M. A. S. Oliveira, and G. P. Thim, "Effect of urea on lead zirconate titanate- $\text{Pb}(\text{Zr}_{0.52}\text{Ti}_{0.48})\text{O}_3$ -nanopowders synthesized by the Pechini method," *J. Eur. Ceram. Soc.*, vol. 25, No. 5, pp. 743–748, Feb. 2005. <https://doi.org/10.1016/j.jeurceramsoc.2004.02.021>
- [41] S. Daengsakul et al., "A simple thermal decomposition synthesis, magnetic properties, and cytotoxicity of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles," *Appl. Phys. A*, vol. 96, no. 3, pp. 691–699, Aug. 2009. <https://doi.org/10.1007/s00339-009-5151-0>
- [42] T. Székely, G. Várhegyi, and F. Till, "The determination and use of the second derivative thermogravimetric function (DDTG) and the calculation of the kinetic constants of some decomposition reaction types," *J. Therm. Analysis*, vol. 5, pp. 227–237, 1973. <https://doi.org/10.1007/bf01950371>
- [43] J. Gaitán-Álvarez et al., "Thermogravimetric, devolatilization rate, and differential scanning calorimetry analyses of biomass of tropical plantation species of Costa Rica torrefied at different temperatures and times," *Energies*, vol. 11, art. 696 2018. <https://doi.org/10.3390/en11040696>
- [44] P. K. Yap et al., "Growth and magnetic behaviours of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles synthesized via thermal treatment method," *Sains Malaysiana*, vol. 48, no. 2, pp. 369–375, Feb. 2019. <https://doi.org/10.17576/jsm-2019-4802-14>
- [45] H. Pfeiffer, "relaxation behaviour of magnetic particle assemblies due to thermal fluctuations," *Phys. Status Solidi*, vol. 120, no. 1, pp. 233–245, Jul. 1990. <https://doi.org/10.1002/pssa.2211200121>
- [46] C. P. Bean and J. D. Livingston, "Superparamagnetism," *J. Appl. Phys.*, vol. 30, no. 4, pp. S120–S129, 1959. <https://doi.org/10.1063/1.2185850>
- [47] D. H. Manh, P. T. Phong, P. H. Nam, D. K. Tung, N. X. Phuc, and I.-J. Lee, "Structural and magnetic study of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanoparticles and AC magnetic heating characteristics for hyperthermia applications," *Phys. B Condens. Matter*, vol. 444, pp. 94–102, Jul. 2014. <https://doi.org/10.1016/j.physb.2014.03.025>
- [48] E. C. Stoner and E. P. Wohlfarth, "A mechanism of magnetic hysteresis in heterogeneous alloys," *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.*, vol. 240, no. 826, pp. 599–642, 1948. <http://www.jstor.org/stable/91421>
- [49] J. García-Otero, A. García-Bastida, and J. Rivas, "Influence of temperature on the coercive field of non-interacting fine magnetic particles," *J. Magn. Magnet. Mater.*, vol. 189, no. 3, pp. 377–383, Nov. 1998. [https://doi.org/10.1016/S0304-8853\(98\)00243-1](https://doi.org/10.1016/S0304-8853(98)00243-1)
- [50] O. Moscoso-Londoño et al., "Different approaches to analyze the dipolar interaction effects on diluted and concentrated granular superparamagnetic systems," *J. Magn. Magnet. Mater.*, vol. 428, 2017. <https://doi.org/10.1016/j.jmmm.2016.12.019>
- [51] J.-L. Ortiz-Quiñonez, L. García-González, F. E. Cancino-Gordillo, and U. Pal, "Particle dispersion and lattice distortion induced magnetic behavior of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ perovskite nanoparticles grown by salt-assisted solid-state synthesis," *Mater. Chem. Phys.*, vol. 246, art. 122834, May 2020. <https://doi.org/10.1016/j.matchemphys.2020.122834>
- [52] I. Radelytskiy et al., "Magnetic anisotropy of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ nanopowders," *J. Magn. Magn. Mater.*, vol. 335, pp. 11–16, Jun. 2013. <https://doi.org/10.1016/j.jmmm.2013.01.031>
- [53] A. D Souza, P. D. Babu, S. Rayaprol, M. S. Murari, L. D. Mendonca, and M. Daivajna, "Size control on the magnetism of $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$," *J. Alloys Comp.*, vol. 797, pp. 874–882, Aug. 2019. <https://doi.org/10.1016/j.jallcom.2019.05.004>
- [54] A. E. Berkowitz, W. J. Schuele, and P. J. Flanders, "Influence of crystallite size on the magnetic properties of acicular $\gamma\text{-Fe}_2\text{O}_3$ particles," *J. Appl. Phys.*, vol. 39, no. 2, pp. 1261–1263, Feb. 1968. <https://doi.org/10.1063/1.1656256>
- [55] W. Xia, Z. Pei, K. Leng, and X. Zhu, "Research progress in rare earth-doped perovskite manganite oxide nanostructures," *Nanoscale Res. Lett.*, vol. 15, no. 1, art. 9, Dec. 2020. <https://doi.org/10.1186/s11671-019-3243-0>
- [56] P. V. Hendriksen, S. Linderöth, and P.-A. Lindgård, "Finite-size modifications of the magnetic properties of clusters," *Phys. Rev. B*, vol. 48, no. 10, pp. 7259–7273, Sep. 1993. <https://doi.org/10.1103/PhysRevB.48.7259>
- [57] D. F. Coral and J. A. Mera-Córdoba, *Nanomedicine and Nanotechnology: Magnetic Nanoparticles Applications in Cancer Treatment (Nanotecnología y Medicina: Aplicaciones de Nanopartículas Magnéticas en el Tratamiento del Cáncer)*, 1st ed. Pasto, Colombia: Editorial CESMAG, 2018. <https://doi.org/10.15658/CESMAG18.090101>

Effect of Ship Size on the Hull Dimension Ratios of Purse Seiners

Efecto del tamaño de la embarcación sobre los coeficientes de dimensiones del casco de los barcos cerqueros

Dennys Dunker de la Torre Cortez ¹

ABSTRACT

In the preliminary ship design stage, the first approximations of speed, stability, structure, weight, and cost, among others, are reached by selecting the appropriate hull dimension ratios. This research was motivated by the noticeable lack of systematic data on these ratios in the literature on American-type semi-industrial purse seiners. Regression analysis was performed to investigate the relationship between hull dimension ratios and vessel size from a database composed of 130 purse seiners of the Peruvian fleet. This work aimed to determine the effect of vessel size, as characterized by length, breadth, depth, gross tonnage, net tonnage, and fish hold capacity, on the main hull dimension ratios of purse seiners, *i.e.*, length/breadth, length/depth, and breadth/depth. A compilation of hull statistics for the ship attributes and ratios was elaborated, with the findings indicating that length significantly affects the slimness of American-type semi-industrial purse seiners, whose structural strength and ship weight increase with length. These vessels exhibit an invariable minimum stability according to standard criteria since a breadth-to-depth ratio of 2 is maintained. In this vein, depth is half the value of breadth. The other parameters do not affect the hull dimension ratios. The estimated regression equations and the patterns presented in this research can be used in the preliminary design stage of American-type semi-industrial purse seiners.

Keywords: hull dimension ratios, purse seiner, ship size, regression analysis, hull statistics

RESUMEN

En la etapa preliminar de diseño de barcos, se alcanzan las primeras aproximaciones de velocidad, estabilidad, estructura, peso y costo, entre otros, mediante la selección de proporciones adecuadas para las dimensiones del casco. Esta investigación fue motivada por la notable falta de datos sistemáticos sobre estas proporciones en la literatura sobre los cerqueros semi-industriales de tipo americano. Se realizó un análisis de regresión para investigar la relación entre las proporciones de las dimensiones del casco y el tamaño del barco a partir de una base de datos compuesta por 130 cerqueros de la flota peruana. El objetivo de este trabajo fue determinar el efecto del tamaño del barco, caracterizado por su eslora, manga, profundidad, tonelaje bruto, tonelaje neto y capacidad de bodega de pescado, sobre las principales proporciones de las dimensiones del casco de los cerqueros, *i.e.*, eslora/manga, eslora/ profundidad y manga/ profundidad. Se elaboró una recopilación de estadísticas del casco para los atributos y proporciones del barco, y los resultados indican que la eslora afecta significativamente la delgadez de los cerqueros semi-industriales de tipo americano, cuya resistencia estructural y peso aumentan con la eslora. Estos barcos presentan una estabilidad mínima invariable de acuerdo con los criterios estándar, pues se mantiene una proporción manga/ profundidad de 2. En este sentido, la profundidad es la mitad del valor de la manga. Los otros parámetros no afectan las proporciones de las dimensiones del casco. Las ecuaciones de regresión estimadas y los patrones presentados en esta investigación pueden utilizarse en la etapa preliminar de diseño de los cerqueros semi-industriales de tipo americano.

Palabras clave: relaciones de dimensiones del casco, cerquero, tamaño del barco, análisis de regresión, estadísticas del casco

Received: August 24th, 2024

Accepted: March 3rd, 2025

Introduction

In the preliminary ship design stage, the first approximations of speed, stability, structure, weight, and cost are reached by selecting the main hull proportions, also referred to as the *hull dimension ratios*, a definition that will be adopted henceforth. Through these ratios, designers can prioritize ship qualities such as seakeeping, velocity, stability, and so on. To accomplish this task, they need an available source of updated hull dimension ratios. Recently, hull dimension ratios for the preliminary design stage have been a hot research topic, *e.g.*, in chemical tankers [1] and Grand Canal ships [2].

[3] developed and compared simple regressions with equivalent formulations available in the literature, proving the better approximations of the trends for 260 non-sister container ships. Similarly, [4] arrived at a practical formula

¹ Naval engineer, Master of Science with an emphasis on Investment Projects, PhD of Science with an emphasis on Energy, Universidad Nacional de Ingeniería, Peru. Affiliation: Research professor, Multidisciplinary Transport Research Group, Department of Mechanical Engineering, Universidad Nacional de Ingeniería. Email: ddelatorrec@uni.edu.pe



to assist naval architects in the preliminary design of cruisers with regard to gross tonnage. Likewise, [5] proposed a procedure for determining the optimal characteristics of multi-purpose cargo vessels (MPCVs) in the preliminary design stage. Their procedure relies on a statistical analysis of the MPCV database.

[6] reported a length over breadth ship ratio (L/B) equal to 4 m for small crafts, which included fishing boats less than 30 m in length. Instead, for vessels 30-130 m in length, a range that covers coasters and general cargo ships, L/B varies according to the formula $L/B = 0.025 \cdot L + 3.25$. Furthermore, for ships over 130 m in length, L/B is around 6.5. For volume carrier ships, a category that includes fishing and cargo ships whose depth is limited due to stability considerations, the breadth over depth ratio (B/D) is around 1.65 [7].

[8] obtained formulas for determining the general patterns and dependence of fishing vessel dimensions on their main design attributes. The studied fishing vessels included purse seiners, trawlers, trawlers, and seiners. These formulas were the result of statistically analyzing 1080 fishing vessels in operation with over 100 GRT and classified under the RMRS class. They relate B to L and D to L . Said formulas are $B = 0.8 \cdot L^{2/3}$ and $D = 0.2 \cdot L^{0.87}$ for vessels between 15 and 120 m in length.

[9] studied the main attributes of 2000 Chinese fishing vessels (fishing tackle vessels, trawlers, and purse seiners). They performed a regression analysis and obtained the relationships between L and other attributes. They developed formulas relating B to L and D to L , i.e., $B = 0.142 \cdot L + 2.433$; $D = 0.07207 \cdot L + 1.129$ for vessels between 15 and 115 m.

[10] presented an estimated regression equation for fishing vessels, wherein L/B is a function of the length between perpendiculars L_{pp} : $L/B = 1.8006 \cdot \ln(L_{pp}) - 2.367$. He also presented the range L_{pp}/D : $8.2 \leq L_{pp}/D \leq 9.0$. He used the IHS Fairplay World Shipping Encyclopedia, which includes a database comprising 637 fishing vessels 20-130 m in length.

[11] considered fishing trawler hulls in their numerical method to determine the stability of fishing vessels. They presented their optimal L/B range, which varies from 4.4 to 5.8.

When designing a fishing vessel, the naval architect considers a list of requirements from the ship owner. i.e., length, fish hold capacity, gross tonnage, net tonnage, speed, power, and range. The fulfilment of these requirements constitutes the preliminary design stage, where the professional usually follows a method based on a parent ship or on statistics. The statistical method is particularly useful when information on parent ship lines is unavailable. The parent ship lines of a semi-industrial purse seiner are usually based on the American or European types, as detailed in a work by the Food and Agriculture Organization (FAO) [12].

American-type purse seiners, with the bridge and accommodation forward and the working deck aft, are usually built in the eastern Pacific Ocean, e.g., in the Peruvian sea. The lines of American-type purse seiners are not always accessible for conducting a systematic series analysis. Consequently, the design process must rely on statistical methods, which aim to determine how the hull dimension ratios are affected by the ship's size to predict its quality. *Ship size* may refer to any of the following characteristics: linear ship dimensions, volume, or deadweight. On the other hand, ship qualities may include speed, stability, structure, weight, seakeeping, or cost. This research was motivated by the noticeable lack of systematic data and regression models on hull dimension ratios in the literature on American-type semi-industrial purse seiners.

This work seeks to determine the influence of purse seiner size on hull dimension ratios. It analyzes the main ship size characteristics, i.e., length (L), breadth (B), gross tonnage (GT), net tonnage (NT), and fish hold capacity (FHC). The hull dimension ratios analyzed herein are L/B , L/D , and B/D . This research relies on the database of fishing vessels published by the Peruvian Government's Ministry of Production (PRODUCE) [13], and it provides exhaustive information on hull statistics, regression and correlation analysis, significance tests, and the patterns revealed, which will aid in the preliminary design stage of semi-industrial purse seiners.

Methodology

Ship database

This research collected data on purse seiners from the records of the vessel registry of PRODUCE [13] for the 1996-2000 period, employing a collection of 130 typical American-type semi-industrial purse seiners with a section plan similar to that of the UCB systematic series [14]. These vessels have two knuckles on the hull and a refrigerated sea water (RSW) conservation system. Table I summarizes the characteristics of the database. Regression and correlation analyses, as well as significance tests, were conducted to investigate the effect of purse seiner size on hull dimension ratios.

Table I. Size characteristics in the purse seiner database

Main characteristic	Symbol	SI Units	Minimum	Maximum
Length	L	m	21.7	77.0
Breadth	B	m	6.0	11.6
Depth	D	m	2.60	5.73
Gross tonnage	GT	-	21.82	312.63
Net tonnage	NT	-	72.5	1006.7
Fish hold capacity	FHC	m ³	89.2	868.3

Source: Author

Regression analysis methodology

The main hull dimension ratios used in this study are L/B, L/D, and B/D, as listed in [15]. L/B represents a measure of slimness, a ratio related to ship speed. For example, “fast ships require larger length to beam [breadth] ratios than slow ships”, according to [16, p. 201]. Additionally, the L/D ratio affects the structural integrity and weight of a ship [17], and B/D provides information on stability, as indicated by the inclination of the height of the center of gravity above the keel. B/D relates to the ship stability because \overline{KM} is a function of B and \overline{KG} is a function of D [18].

During the ship concept design stage, the operational requirements are defined in terms of engineering parameters. In the subsequent preliminary design phase, the final ship proportions, arrangements, power plant type, and structural layout are determined to meet said requirements [19]. Shipowners usually give the shipbuilder four pieces of information: the type of vessel, the deadweight, the required service speed, and the route on which the new vessel will operate [20]. Regarding the latter, a length or breadth constraint may be imposed by the dimensions of docks and canal locks. The ship's linear dimension is mainly represented by L, and the two other relevant linear dimensions are B and D.

A second measure of ship size is volume, which is characterized by GT and NT. The former represents the total enclosed volume of the ship, and the latter considers the volume of the cargo and passenger spaces. *Tonnage* was defined by the International Convention on Tonnage Measurement of Ships held in 1969 [16]. Another measure of ship size that is related to fishing vessel volume is the FHC.

The GT equation is as follows:

$$GT = K_1 V \quad (1)$$

where:

- $K_1 = 0.2 + 0.02 \cdot \log_{10} V$
- V = total volume of all enclosed spaces of the vessel in m^3

NT is a measure of the carrying capacity of a vessel. It can be regarded as an indicator of revenue-generating capacity. It is determined as a function of the molded volume of the vessel. This equation also includes the volume of the cargo space plus the volume of the passenger spaces. The NT equation is as follows:

$$NT = K_2 V_c \left(\frac{4T}{3D} \right)^2 + K_3 \left(N_1 + \frac{N_2}{10} \right) \quad (2)$$

where:

- $K_2 = 0.2 + 0.02 \cdot \log_{10} V_c$
- $K_3 = 1.25 \left(\frac{GT + 10000}{10000} \right)$
- V_c = total volume of the cargo spaces in m^3
- D = molded depth amidships in m
- T = molded draught amidships in m
- N_1 = number of passengers in cabins with no more than eight berths
- N_2 = number of other passengers
- The NT value is not less than 30% of the GT.

Finally, deadweight (DWT) is the third measure of ship size. It is the weight of cargo that can be carried at full load draft. Specifically, DWT is the difference between the full load displacement weight and the lightship weight. This measure is mainly used in bulkers, container ships, tankers, and general cargo ships. For fishing vessels, FHC is used instead of DWT.

Within the simple regression methodology, the coefficient of determination was quantified and compared against the values found in the literature to assess regression accuracy. The equation for the coefficient of determination (R^2) is as follows [21]:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^*)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

where y_i are n observations, y_i^* represents the predicted values, and \bar{y} is the mean value of the observations.

The confidence intervals for the true slope and the intercept are presented below.

$$\left. \begin{aligned} b_1 - t_{\alpha/2} s_{b_1} &\leq \beta_1 \leq b_1 + t_{\alpha/2} s_{b_1} \\ b_0 - t_{\alpha/2} s_{b_0} &\leq \beta_0 \leq b_0 + t_{\alpha/2} s_{b_0} \end{aligned} \right\} \quad (4)$$

where s_{b_1} is the standard error of the slope, s_{b_0} is the standard error of the intercept, β_1 is the true slope, β_0 is the true intercept, and t denotes Student's t with d.f. = $n-2$ degrees of freedom at a 95% confidence level.

Are the true slope and intercept different from zero? The hypotheses and criteria in this regard are:

Coefficient	Hypotheses	Criteria
Slope	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	$p_{value} < 0.05$

$$\begin{aligned} \text{Intercept} \quad H_0 : \beta_0 &= 0 \\ H_0 : \beta_0 &\neq 0 \quad p_{\text{value}} < 0.05 \end{aligned}$$

Generally, three characteristics representing the ship size are used in the database: linear dimensions, volume characteristics, and DWT. Notably, as fishing vessels are categorized as volume carriers, DWT was not used in this research. There are a few relevant hull dimension ratios (L/B, L/D, B/D) and other measures related to immersion depth. However, as the database does not include information on the latter, only L/B, L/D, and B/D were considered.

Results and discussion

Hull statistics

This subsection discusses central tendency, dispersion, and distribution shape. In statistics, the central tendency of the data is represented by the mean or the median (when the data are continuous). Table II shows that the mean and median are markedly different for GT, which was quantified using Eq. (1); NT, which was quantified using Eq. (2); and FHC, as obtained from measurements. Therefore, using the mean or median to determine the central tendency of variables relies on the distribution shape or on skewness and kurtosis.

In our case, the skewness for all variables exceeds 0.5, except for L, B, D, and L/B. The kurtosis is higher than 0.5, except for B, so the median represents the central tendency of all variables except B. Their distribution shape denotes right skewness, and the frequency is higher than expected. B is the only variable with a normal distribution, so its mean represents its central tendency.

Although the median and mean of B/D are the same and equal to two, the distribution of B/D is non-normal. In general terms, we can say that the value of D is taken as half

of B. Standard deviation and range values were used to set the number of bins in the corresponding histograms.

Fig. 1 shows the distribution of L as representative of the linear dimension variables. This distribution exhibits a long tail to the right and a concentration between 40 and 48 m in length. Fig. 2 suggests that 70% of American-type semi-industrial purse seiners in the Peruvian fleet are about over 300 GT, and about 42% of vessels are over 400. The GT distribution is skewed right and concentrated between 300 and 500 GT.

The FHC distribution does not follow a normal distribution (Fig. 3). Roughly five vessels have extreme FHCs, i.e., over 700 m³. The FHC ranges from 400 to 700 m³. High positive values regarding skewness, kurtosis, and the variation in the histograms indicate that the variables (except B) do not have a normal distribution. In light of the above, the median was selected as the best measure of central tendency for the Peruvian fleet of American-type semi-industrial purse seiners. This fleet is composed of vessels with a length of about 33 to 55 m, about 300 to 500 GT, and about 300 to 500 m³ of FHC.

The histograms reveal an upper L limit of 70 m instead of 77 m, as initially shown in Table I. Vessels over 70 m in length belong to a new category of ships, with a GT over 600, a NT over 175, and a FHC over 600 m³. The existence of large fishing vessels in the Peruvian fleet may be due to the need for catching larger species than anchovy, like tuna, in order to preserve the catch for more time using advanced processing facilities, improve energy use, and sail out to sea and stay there longer.

The hull statistics of the Peruvian fishing fleet resemble those of other fleets from the west coast of the American continent because anchovy and other small species are mainly caught using American-type purse seiners. Thus, the results provide meaningful information for the preliminary stage of fishing vessel design.

Table II. Descriptive statistics of ship size for the Peruvian fleet (American-type semi-industrial purse seiners)

Statistic	Length (m)	Breadth (m)	Depth (m)	L/B	L/D	B/D	Net tonnage	Gross tonnage	Fish hold capacity (m ³)
Mean	43.5	9.0	4.4	4.81	9.73	2.02	113.7	376.5	446.3
Standard error	0.8	0.1	0.1	0.06	0.12	0.01	3.6	14.5	12.3
Median	43.3	8.9	4.4	4.70	9.42	2.00	110.5	366.5	440.0
Standard Deviation	8.7	1.1	0.6	0.63	1.34	0.14	41.4	164.7	140.5
Kurtosis	1.6	-0.3	0.9	1.37	2.51	10.91	4.2	1.6	1.5
Skewness	0.2	-0.01	-0.4	0.42	1.01	2.37	0.9	0.8	0.1
Range	52.3	6.6	3.1	3.88	9.25	1.14	290.8	934.2	779.0

Source: Author

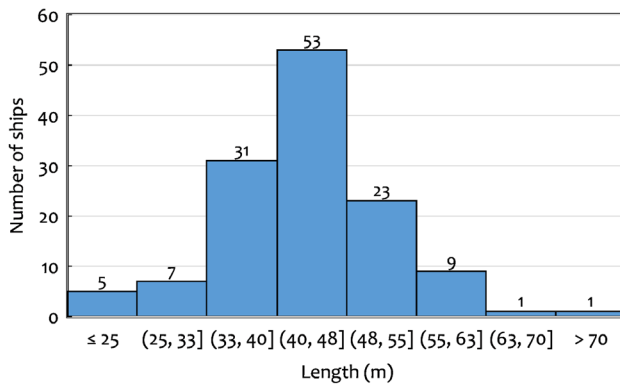


Figure 1. Length histogram for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

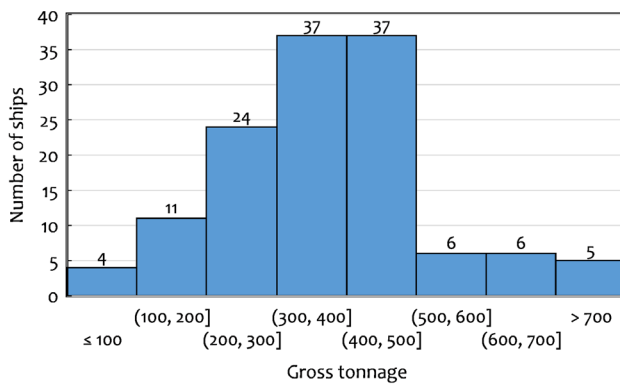


Figure 2. Gross tonnage histogram for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

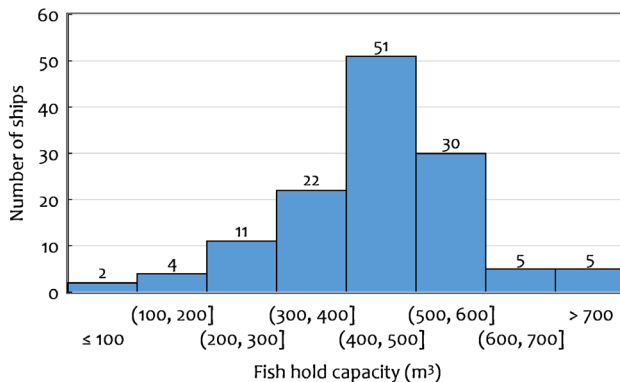


Figure 3. Fish hold capacity histogram for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

Regression analysis

In the model, the response variables are L/B, L/D, and B/D. In addition, the predictor variables are L, B, D, GT, NT, and FHC. Table III presents the estimated slope b_1 and the intercept b_0 , as well as the R^2 for all combinations of the response and predictor variables. The table shows that a straight line does not fit the data of the response and predictor variables, except for the combinations between L/B and L and L/D and L. Since the L/B regression yields an

R^2 value of 0.66, we could say that L explains 66% of the variation in L/B. Similarly, L explains 67% of the L/D variation because the regression yields an R^2 of 0.67.

The unexplained variation reflects factors not included in our model (e.g., later elongations, intact stability criteria, seakeeping qualities, and speed and power considerations) or just plain random variation. In this case, there is, in theory, a reason to believe that causation exists; specifically, increased length increases the slimness of the vessel, as well as the structural integrity and weight of a ship. The estimated regression equation of L/B as a function of L is:

$$\frac{L}{B} = 2,25 + 0,06 \cdot L \quad (5)$$

Each extra meter of L will add an average of 0.06 units to L/B. The intercept alone is not significant because no vessel can have $L = 0$ m.

The estimated regression equation of L/D as a function of L is:

$$\frac{L}{D} = 4,28 + 0,13 \cdot L \quad (6)$$

The L/D value is at least 4.28. An extra meter of L will increase L/D by 0.13 units.

The fit between the other combinations of response and predictor variables is poor, as indicated by R^2 values close to zero. Consequently, there are no useful regression equations for the remaining combinations. However, naval architects can use the mean or median while considering the hull statistics discussed in the previous section.

Fig. 4 indicates a strong positive relationship between L/B and L. A straight line can be constructed to fit the data. Many points are scattered when L ranges between 33 and 55 m, so the R^2 decreases to 0.66. Fig. 5 shows a strong positive relationship between L/D and L. The data are scattered and uniformly spread around the line.

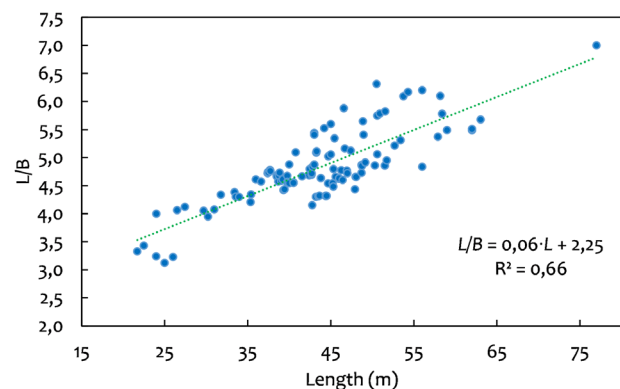


Figure 4. Regression analysis of L/B as a function of L for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

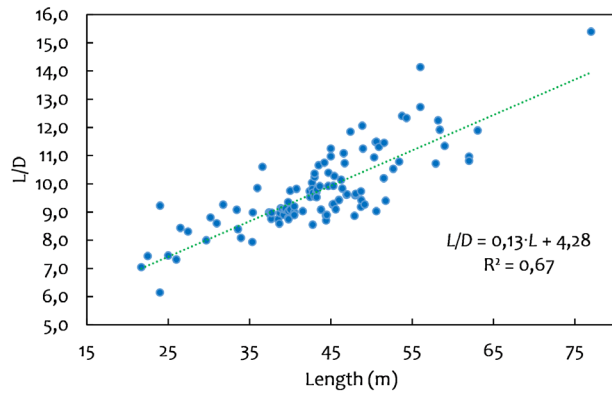


Figure 5. Regression analysis of L/D as a function of L for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

The scatter plot of B/D and L (Fig. 6) shows no linear relationship. The values of B/D range from 1.75 to 2.25 regardless of L. The spread of the data indicates no correlation between the variables, which is also confirmed by the low R^2 of the obtained linear trend line.

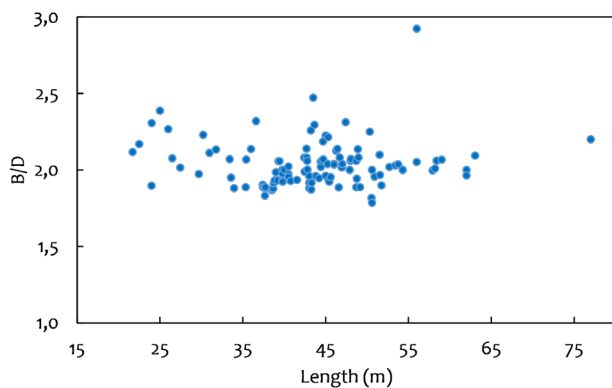


Figure 6. Scatter plot of B/D and L for the American-type semi-industrial purse seiners in the Peruvian fleet

Source: Author

The confidence intervals for the true slope and intercept can be considered trustworthy given the normality in the residuals histogram. The residuals normal probability plots

of L/B and L/D exhibit $R^2 = 0.844$ for L/B vs. L, and 0.834 for L/D vs. L, thus confirming residuals normality.

The confidence intervals for the coefficients, as described in Eq. (4), are not widened because there is no heteroscedasticity in the residuals vs. L plot. The residuals plot does not show a left-to-right pattern.

The model's fit is not overstated because there is no pattern of non-independent errors. The Durbin-Watson statistics were nearly 2, i.e., 1.62 for L/B vs. L, and 1.59 for L/D vs. L. For the two well-fitted equations (L/B vs. L and L/D vs. L), nonlinearity studies were conducted. The linear model was identified as the best fit for L/B vs. L since the coefficient of the quadratic term did not significantly differ from zero.

Similarly, for L/D vs. L, the linear model was deemed optimal, even though the coefficient of the quadratic term was not significantly different from zero. This decision was supported by a marginal increase in R^2 : from 0.666 (linear) to 0.679 (quadratic), representing a mere 2% improvement. Introducing a squared predictor increases model complexity and reduces the number of degrees of freedom for significance testing, which justifies the preference for the simpler linear model.

According to the results, as L increases in the database, ship slimness and weight increase, and structural integrity demands more stiffness. Conversely, ship stability, represented by B/D, remains invariable. The regression analyses of the remaining variable combinations evidenced no correlations.

Table IV shows that the confidence intervals for β_1 and β_2 do not include zero for the combinations between L/B and L and L/D and L. The true slope and the intercept are nonzero for the corresponding equations, as indicated by their low p-values (≈ 0), according to Eq. (4). These nonzero values are not due to chance. Namely, the relationships between L/B and L and L/D and L are statistically significant; L/B and L/D can be predicted by L.

Although the confidence intervals show that the true slope and the intercept of the other variable combinations are

Table III. Regression analysis of the coefficients for the hull dimension ratios of the American-type semi-industrial purse seiners in the Peruvian fleet

	L/B			L/D			B/D		
	Slope (b_1)	Intercept (b_0)	R^2	Slope (b_1)	Intercept (b_0)	R^2	Slope (b_1)	Intercept (b_0)	R^2
Length (m)	0.06	2.25	0.66	0.13	4.28	0.66	0.00	2.00	0.00
Breadth (m)	0.17	3.29	0.09	0.51	5.13	0.18	0.03	1.75	0.05
Depth (m)	0.45	2.80	0.15	0.59	7.08	0.06	-0.08	2.38	0.09
Net tonnage	0.00	4.05	0.19	0.01	8.19	0.16	0.00	2.04	0.00
Gross tonnage	0.00	3.92	0.36	0.00	7.76	0.41	0.00	2.00	0.00
Fish hold capacity (m^3)	0.00	3.84	0.23	0.00	7.80	0.21	0.00	2.05	0.00

Source: Author

Table IV. Confidence intervals for the true slope and intercept, and p-values for the significance tests

		L/B			L/D			B/D		
		95% lower	95% upper	p-value	95% lower	95% upper	p-value	95% lower	95% upper	p-value
Length (m)	β_0	2.27	2.93	0.00	4.78	6.66	0.00	2.02	2.28	0.00
	β_1	0.04	0.06	0.00	0.07	0.11	0.00	-0.01	0.00	0.02
Breadth (m)	β_0	3.21	4.61	0.00	7.04	10.29	0.00	2.04	2.41	0.00
	β_1	0.03	0.18	0.01	-0.06	0.28	0.21	-0.04	0.00	0.02
Depth (m)	β_0	3.72	4.75	0.00	9.63	12.00	0.00	2.40	2.62	0.00
	β_1	0.02	0.24	0.02	-0.49	0.00	0.05	-0.13	-0.08	0.00
Net tonnage	β_0	4.47	4.84	0.00	9.41	10.27	0.00	2.07	2.16	0.00
	β_1	0.00	0.00	0.02	0.00	0.00	0.37	0.00	0.00	0.00
Gross tonnage	β_0	4.37	4.72	0.00	9.12	9.97	0.00	2.06	2.15	0.00
	β_1	0.00	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00
Fish hold c. (m ³)	β_0	4.43	4.82	0.00	9.37	10.28	0.00	2.08	2.17	0.00
	β_1	0.00	0.00	0.01	0.00	0.00	0.46	0.00	0.00	0.00

Source: Author

Table V. Comparison of the obtained regression formulas against those reported in the literature

This database	[7]	[10]	[8]	[9]
$L/B=0.06.L+2.25$	$L/B=0.025.L+3.25$	$L/B=1.8006 \cdot \ln(L_{pp})-2.367$	$L/B=1.25.L^{1/3}$	$L/B=L/(0.142.L+2.433)$
$B/D=2$	$B/D=1.65$	---	---	---
$L/D=0.13.L+4.28$	---	8.6 in average	$L/D=5.L^{0.13}$	$L/D=L/(0.07207.L+1.129)$

Source: Author

significantly different from zero, the response variables could not be predicted by their corresponding predictor variables because their R^2 is nearly zero. Thereupon, the best prediction for B/D is the median, i.e., the mean should not be used to predict B/D since its distribution is non-normal.

The median and mean of B/D are 2.0, which is higher than the value obtained by [10] (1.65), although this previous study does not specify the type of fishing vessel analyzed. The B/D value of 2 can be considered as a unique characteristic of American-type purse seiners.

Comparison against previous studies

Tables V and VI and Figs. 7 and 8 compare the estimated regression formulas for the analyzed database against those reported in the literature. To evaluate the goodness of fit of said formulas, their R^2 value was calculated using Eq. (3). The R^2 formula considered the database as the observed values and the predictions as the estimated values. Table VI presents the calculated R^2 for the hull dimension ratios as a function of L. Additionally, Table VI shows that only the formulas developed by [9] exhibit a certain degree of goodness of fit for the analyzed database. It is important to recall that these formulas only apply to purse seiners.

Table VI. Comparison of the R^2 obtained against the values reported in the literature

R^2	$L/B=f(L)$	$B/D=f(L)$	$L/D=f(L)$
This database	0.66	0	0.67
[7]	0	0	---
[10]	0	---	---
[8]	0.06	---	0
[9]	0.45	---	0.40

Source: Author

Fig. 7 compares this work's regression formula against four other regressions for L/B estimation. These estimations follow two different tendencies for L values lower than 35 m and greater than 55 m, which is only the case for the formula of [9]. The other regression formulas do not predict the L/B values of the database analyzed in this work. Fig. 8 shows that only the formula of [9] predicts the L/D values when L is between 35 and 55 m. These figures illustrate that the estimated slope and intercept of L/D as a function of L are twice the values for L/B as a function of L. This is expected because $B/D = 2$ for the studied database.

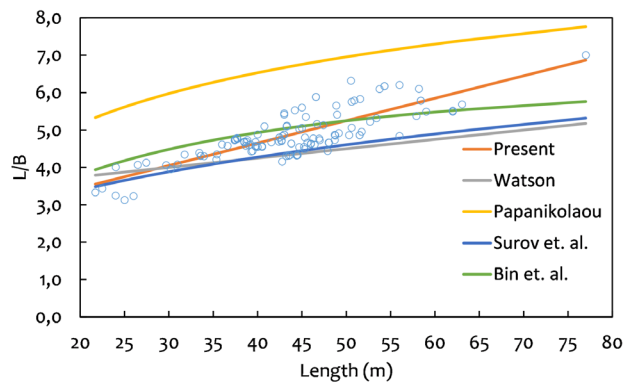


Figure 7. Length-to-breadth ratio regression compared to the reports by [7], [10], [8], and [9]

Source: Author

The regression results contrast those in [9] when L is around 55 m or higher, which could be due to the fact that fishing vessels can be elongated, keeping the master frame intact to enlarge the FHC and increase their speed. This type of hull modification is a common practice in the fishing fleets of the American Pacific coasts.

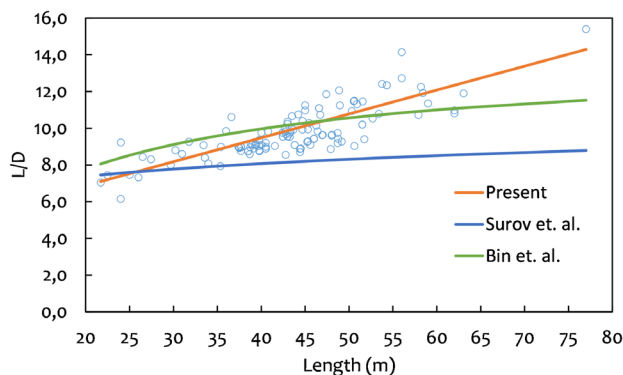


Figure 8. Length-to-depth-ratio regression compared to the reports by [7], [10], [8], and [9]

Source: Author

Application example

B can be obtained from Eq. (5) if L is known. Then, D is obtained by dividing B by the B/D median ($=2$). Ship attributes like GT , NT , and FHC can be estimated through their corresponding median.

For example, if $L = 50$ m, L/B is 5.25 ($L/B = 2.25 + 0.06 \cdot 50 = 5.25$). Afterwards, B is obtained by dividing L by 5.25, or $B = L / 5.25 = 50 / 5.25 \approx 9.5$ m. Next, D is quantified by dividing B by 2, i.e., $D = B / 2 = 9.5 / 2 = 4.75$, or D is obtained via Eq. (6) ($L/D = 4.28 + 0.13 \cdot L$). Then, D is 4.63 m. In this vein, the GT can be roughly estimated as 366.5, the NT as 110.5, and the FHC as 440 m³.

Conclusions

Length significantly affects the slimness of a ship. Increasing values of L increase L/B , as per the estimated regression equation $L/B = 2.25 + 0.06 \cdot L$. A straight line can be drawn in the scatter plot of L/B and L with a good fit. The confidence intervals of the true slope and intercept do not include zero. The true slope and intercept are nonzero, as indicated by their low p -values. Since L/B (ship slimness) is directly related to speed, the positive correlation between L/B and L suggests that designers may have prioritized speed over other vessel qualities when determining the hull dimensions analyzed in this work. It should be considered that, as L increases, the designer must enhance structural strength to counteract the greater bending stresses, which yields an increase in the ship's overall weight. In this vein, it is consistent that length significantly affects the L/D ratio, as per the formula $L/D = 4.28 + 0.13 \cdot L$.

In summary, the estimated regression equations for L/B and L/D as a function of length, along with the constant B/D , suggest that longer ships tend to be slimmer, prioritizing speed and stability. When B is predicted using the L/B equation, D can be estimated as half of B – statistically, the former is twice the latter. This proportion reflects the designer's focus on ensuring good stability, as B influences the metacentric height while D denotes the height of the center of gravity. For American-type semi-industrial purse seiners, this proportion remains statistically constant, prioritizing a high metacentric height over the height of the center of gravity. These vessels are consistently designed to maximize speed and ensure stability, often at the expense of other qualities.

On the other hand, B , D , GT , NT , and FHC do not affect the hull dimension ratios L/B , L/D , and B/D . This affirmation does not exclude the impact of said parameters on the ship concept design stage. This means that the hull dimension ratios are determined independently by the aforementioned variables.

A new category of ships has emerged given the need to catch species larger than anchovies, e.g., tuna. The histograms and scatter plots presented in this work reveal an upper limit to the validity of the regression equations. For fishing vessels between 20 and 70 m in length, the equations hold true. For vessels exceeding 70 m, a different regression equation, tailored to a distinct ship category, would be required. These larger ships have a GT over 700 and a FHC over 700 m³. They allow for more time spent on fishing, improve energy use, and allow sailing out to sea and staying there longer. This new category scattered the data points, resulting in low R^2 values.

Future research must be conducted to investigate the relationship between hull form coefficients, ship linear dimensions, and hull dimension ratios. Furthermore, a new concept design for an industrial purse seiner could be elaborated while considering the information presented herein.

Acknowledgements

The author would like to thank the reviewers for their comments, which helped to improve the manuscript; as well as the Vice-Principal for Research of Universidad Nacional de Ingeniería for the financial support provided under Rectoral Resolution N.º 0892-2024-UNI.

References

- [1] S. Gorgen, I. Altin, and M. Ozkok, "Prediction of main particulars of a chemical tanker at preliminary ship design using artificial neural network," *Ships Offshore Struct.*, vol. 13, no. 5, pp. 459–465, Jul. 2018. <https://doi.org/10.1080/17445302.2018.1425337>
- [2] X. Peng, H. Lei, and W. Zhong, "Research on estimation method of gross registered tonnage of ships in Grand Canal," in *Modeling Risk Management for Resources and Environment in China*, 1st ed., D. D. Wu and Y. Zhou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 101–108. <https://doi.org/10.1007/978-3-642-18387-4>
- [3] B. Rinauro, E. Begovic, F. Mauro, and G. Rosano, "Regression analysis for container ships in the early design stage," *Ocean Eng.*, vol. 292, no. December 2023, art. 116499, Jan. 2024. <https://doi.org/10.1016/j.oceaneng.2023.116499>
- [4] A. Lepore, L. Mocerino, B. Palumbo, E. Rizzuto, E. Rossi, and L. Vitiello, "A supervised statistical learning approach to the preliminary design of cruise ship gross tonnage," *Appl. Ocean Res.*, vol. 143, no. November 2023, art. 103837, Feb. 2024. <https://doi.org/10.1016/j.apor.2023.103837>
- [5] M. Kalajdžić and N. Momčilović, "A step toward the preliminary design of seagoing multi-purpose cargo vessels," *Brodogradnja*, vol. 71, no. 2, pp. 75–89, Jun. 2020. <https://doi.org/10.21278/brod71205>
- [6] D. G. M. Watson and A. W. Gilfillan, "Some ship design methods." [Online]. Available: <https://trid.trb.org/View/60661>
- [7] D. G. M. Watson, *Practical ship design*, vol. 1. Oxford: Elsevier science LTD, 1998.
- [8] O. E. Surov, M. V. Kitaev, E. E. Solovieva, V. A. Veselov, and D. V. Tyuftyaev, "Analysis of main dimensions and characteristics of fishing vessels," *Russ. J. Water Transp.*, vol. 2, no. 72, pp. 41–53, Sep. 2022. <https://doi.org/10.37890/jwt.vi72.273>
- [9] H. Bin, L. Xiuwen, and R. Yuqing, "Statistics and analysis of main parameters of ocean fishing vessels in China," in *2020 5th Int. Conf. Electromech. Cont. Tech. Trans. (ICECTT)*, May 2020, pp. 594–598. <https://doi.org/10.1109/ICECTT50890.2020.00135>
- [10] A. Papanikolaou, *Ship design. Methodologies of preliminary design*, 1st ed. Dordrecht, Netherlands: Springer, 2014. <https://doi.org/10.1007/978-94-017-8751-2>
- [11] V. N. Acosta Pastor and V. A. Loarte Vicuña, "Determinación numérica de la estabilidad de embarcaciones pesqueras," *TECNIA*, vol. 10, no. 2, Dec. 2000. <https://doi.org/10.21754/tecnia.v10i2.460>
- [12] T. Stefano, R. van Anrooy, A. Gudmundsson, and D. Davy, *Classification and definition of fishing vessel types*, 2nd ed. Rome, Italy: FAO, 2023. <https://doi.org/10.4060/cc7468en>
- [13] Ministerio de la producción – Gobierno del Perú, "Embarcaciones pesqueras." [Online]. Available: <https://consultasenlinea.produce.gob.pe/ConsultasEnLinea/consultas.web/embarcacion>
- [14] L. J. C. Valerio Mena, V. N. Acosta Pastor, and V. A. Loarte Vicuña, "Determinación numérica de la resistencia al avance de buques," *TECNIA*, vol. 9, no. 2, pp. 61–68, Dec. 1999. <https://doi.org/10.21754/tecnia.v9i2.324>
- [15] N. Hamlin, "Ship geometry," in *Principles of Naval Architecture*, 1st ed., vol. I, E. Lewis, Ed. New Jersey, USA: Society of Naval Architects and Marine Engineers, 1988, ch. 1.
- [16] E. C. Tupper, *Introduction to Naval Architecture*, 3rd ed. Oxford, UK: Elsevier Butterworth-Heinemann, 2004.
- [17] Y. Okumoto, Y. Takeda, M. Mano, and T. Okada, *Design of ship hull structures*. Berlin, Heidelberg, Germany: Springer, 2009. <https://doi.org/10.1007/978-3-540-88445-3>
- [18] H. Schneekluth and V. Bertram, *Ship design for efficiency and economy*, 2nd ed. Oxford, UK: Butterworth-Heinemann, 1998.
- [19] R. Kiss, "Concept design," in *Ship Design and Construction*, R. Taggart, Ed. New York, NJ, USA: The Society of Naval Architects and Marine Engineers, 1980, p. 735.
- [20] C. B. Barrass, *Ship design and performance for masters and mates*. Oxford, UK: Elsevier Butterworth-Heinemann, 2004.
- [21] D. Doane and L. Seward, *Applied statistics in business and economics*, 3rd ed. New York, NJ, USA: McGraw-Hill, 2011.

Development of Auto-Injection Systems through the TRIZ Problem-Solving Method

Desarrollo de sistemas de autoinyección mediante el método de resolución de problemas TRIZ

Burhan Şahin ¹

ABSTRACT

This study presents a comprehensive approach to the redesign of auto-injection syringe systems, employing the TRIZ problem-solving framework along with Ishikawa analysis. The proposed design aims to address common challenges, including usability issues, high production costs, complex assembly procedures, and hygiene considerations. By leveraging the TRIZ methodology, this work successfully identified and addressed technical contradictions, leading to the development of an innovative auto-injection syringe. This design incorporates a recyclable polypropylene random copolymer, which not only reduces manufacturing costs but also promotes environmental sustainability. Replacing flexible springs with inexpensive rubber bands enhances the design's affordability and usability. This change lowers costs and improves user-friendliness, allowing patients to operate the system more easily while upholding performance standards. According to engineering validations carried out through static and dynamic simulations in NX Nastran, the design safely withstands up to 10 N of applied force, with its maximum stress levels remaining below 5.2 MPa, well within the material's 27.5 MPa yield strength. While prior studies have reported ergonomic or functional improvements, they often lack a systematic engineering approach to address design contradictions. This study fills that gap by uniquely integrating the TRIZ and Ishikawa approaches to develop an optimized, user-friendly, and sustainable autoinjector. As a result, our new design meets user needs and adheres to the industry's safety and efficacy standards. This research underscores the effectiveness of integrating the aforementioned methodologies to create practical and efficient solutions for patients requiring regular self-injection, thereby contributing to improved healthcare outcomes and a more sustainable medical device industry.

Keywords: syringe design, Ishikawa diagram, mechanical design and analysis, recyclable materials and sustainability, medical device innovation

RESUMEN

Este estudio presenta un enfoque integral para el rediseño de sistemas de jeringas autoinyectables, empleando el marco de resolución de problemas TRIZ junto con el análisis Ishikawa. El diseño propuesto busca abordar desafíos comunes, incluidos problemas de usabilidad, altos costos de producción, procedimientos de ensamblaje complejos y consideraciones de higiene. Aprovechando la metodología TRIZ, este trabajo logró identificar y resolver contradicciones técnicas, lo que condujo al desarrollo de una innovadora jeringa autoinyectable. Este diseño incorpora un copolímero aleatorio de polipropileno reciclable, lo cual no solo reduce los costos de fabricación, sino que también promueve la sostenibilidad ambiental. La sustitución de resortes flexibles por bandas elásticas económicas mejora la asequibilidad y la facilidad de uso del diseño. Este cambio reduce los costos y mejora la usabilidad, permitiendo que los pacientes operen el sistema con mayor facilidad sin comprometer los estándares de rendimiento. Según validaciones de ingeniería realizadas mediante simulaciones estáticas y dinámicas en NX Nastran, el diseño soporta hasta 10 N de fuerza aplicada de manera segura, con niveles máximos de tensión que se mantienen por debajo de los 5.2 MPa, muy por debajo del límite elástico del material, que es de 27.5 MPa. Si bien han reportado mejoras ergonómicas o funcionales, los estudios anteriores a menudo carecen de un enfoque sistemático de ingeniería para abordar contradicciones de diseño. Este estudio llena ese vacío al integrar de manera única los enfoques TRIZ e Ishikawa para desarrollar un autoinyector optimizado, fácil de usar y sostenible. Como resultado, nuestro nuevo diseño satisface las necesidades del usuario y cumple con los estándares de seguridad y eficacia de la industria. Esta investigación subraya la eficacia de integrar las metodologías mencionadas para crear soluciones prácticas y eficientes para pacientes que requieren autoinyecciones regulares, contribuyendo así a mejorar los resultados sanitarios y a una industria de dispositivos médicos más sostenible.

Palabras clave: diseño de jeringas, diagrama de Ishikawa, diseño y análisis mecánico, materiales reciclables y sostenibilidad, innovación en dispositivos médicos

Received: August 28th, 2024

Accepted: March 12th, 2025

¹ Mechanical engineer, Manisa Celal Bayar University, Manisa, Turkey. MSc and PhD in Production and Construction in Mechanical Engineering, Manisa Celal Bayar University, Manisa, Turkey. Affiliation: Chief engineer at Digital Mock-Up Management, Turkish Aerospace Industries, Ankara, Turkey. Email: burhan-03@hotmail.com



Introduction

The effective handling of various medical conditions often necessitates the use of injectable medications, which have become increasingly vital in modern healthcare. Pre-filled syringes are disposable plastic devices designed for the administration of single-dose vaccines and parenteral medications, specifically engineered to prevent needle reuse. The transition from metal or glass syringes to plastic has resulted in an increased demand for these syringes in the medical field [1]. As advancements in medical technology continue to evolve, a growing number of patients are opting to self-administer this type of treatment, resulting in a significant transformation in medication delivery methods [2]. Automated injection syringe systems have emerged as essential tools that not only enhance the convenience of self-administration but also improve the safety of these processes. The use of autoinjectors has been observed to significantly reduce pain at the injection site, with a substantial majority of patients favoring this method as a more satisfactory option [3]. By empowering patients to independently manage their treatment regimens, these systems can lead to improved adherence to prescribed therapies and, consequently, better therapeutic outcomes [4]. The applications of autoinjectors tested for patient use have been widely favored due to their ease of use and reduced discomfort [5]. Studies have also looked into the reliability of auto-injectors, exploring the frequent reasons for their failure. The findings indicate that these devices have a low failure rate, underscoring their benefits in terms of usability and maintenance, which positions them as highly dependable tools [6].

However, despite the numerous advantages offered by contemporary automatic injection systems, several challenges persist which can hinder their widespread adoption and effectiveness. Usability remains a prominent concern, as some devices are overly complex and difficult for patients to operate, which can lead to frustration and decreased compliance with treatment protocols [7]. In the literature, several studies have emphasized the increasing prevalence of autoinjectors, attributing this trend to their convenience and dependability. However, it has been noted that the high costs of these advanced systems may hinder widespread accessibility [8] and could create barriers for patients who would greatly benefit from the ability to self-administer their medications. In addition, certain products pose significant hygiene risks, particularly those that require the manual transfer of medication from disposable syringes to automatic devices. On the other hand, reusable devices pose a risk of infection and tend to be more expensive for manufacturing companies [9]. This process can introduce contamination, jeopardizing patient safety and undermining the very benefits these systems are designed to provide [10].

Medical devices also have significant adverse effects on the environment; disposable devices contribute to the increase in medical waste and are largely responsible for the inventory costs incurred by hospitals. The development

of biodegradable alternatives or the utilization of recyclable materials is crucial for mitigating the environmental impacts of single-use, non-recyclable, and non-biodegradable plastics within the healthcare sector [11]. Polymers (plastics) possess unique characteristics that allow for a wide range of applications in medical device technology, constituting the majority of single-use medical devices [12]. Additionally, the use of recyclable materials in medical device designs promotes sustainability and aligns with current environmental considerations, which are becoming increasingly important in today's healthcare landscape [13]. Studies have shown that the use of recyclable medical devices can boost environmental sustainability and increase efficiency in healthcare services [14].

Electromechanical autoinjectors have also been developed as suitable alternatives. These systems offer benefits such as customizable injection speeds, reliable medication delivery, and electronic record-keeping, which enhance adherence and sustainability through reusability. They can also reduce discomfort and anxiety during injections, improving the patient's experience [15]. Recent developments in autoinjector technology have integrated advantageous features such as precise dosing and user-friendly operation [16].

Research has also shown that integrating sustainability into medical device development has led to the creation of several approaches for assessing sustainability criteria. These approaches are designed to evaluate design concepts and improve the process overall [17].

Various methodologies are widely used in this style of product development. For example, utilizing the Ishikawa diagram for root cause analysis enables the systematic identification of the various problem factors contributing to the limitations of current devices. Furthermore, the TRIZ (theory of creative problem solving) supports systematic innovation, facilitating the development of creative and effective solutions to these challenges [18].

In the product development process, various creative problem-solving techniques are employed, such as brainstorming, design thinking, six sigma [19], PDCA (plan-do-check-act) [20], DMADV (define, measure, analyze, design, and verify) [21], 5M (machinery, manpower, material, measurement, method) [22], and morphological analysis. However, these methods often fall short in providing structured mechanisms to resolve specific contradictions, particularly within the engineering domain. While design thinking and brainstorming are fundamentally user-centered, they heavily rely on iteration and experiential learning. Moreover, DMAIC (define, measure, analyze, control) emphasizes statistical improvement over innovation. In contrast, TRIZ offers a solution-oriented and algorithmic framework based on inventive principle patterns, proving highly effective in addressing technical conflicts. The Ishikawa diagram complements this process by systematically identifying root causes and key focus areas across multiple domains.

The integration of the TRIZ and Ishikawa methods fills a significant gap in the literature by providing analytical depth and creative solutions. This integration is particularly crucial in the complex development of automatic injection systems, where safety, usability, cost, and sustainability intersect.

This study identifies shortcomings in current auto-injection systems which necessitate further investigation and innovation. Many systems lack intuitive design, making them challenging for users, especially for those with limited technical skills or anxiety. Research is needed to create more ergonomic designs and user interfaces for diverse populations, including children and the elderly. High production costs also restrict access to life-saving treatments. Exploring alternative materials, innovative manufacturing methods, and optimizing supply chains could yield more affordable solutions without sacrificing quality or safety. In addition, collaboration with manufacturers may enhance production efficiency and reduce costs.

The assembly of auto-injection systems is complex and labor-intensive, increasing production costs and the risk of human error. Streamlining this process through modular designs or automation could enhance efficiency and reliability. Hygiene is also a critical concern, as existing designs may not adequately address contamination risks. Investigating antimicrobial or self-sanitizing materials and packaging solutions that maintain sterility until use is essential for improving safety and reducing infection risks.

Our research hypothesis posits that the combination between the TRIZ problem-solving methodology and Ishikawa analysis can significantly enhance the development of an advanced auto-injection system, effectively addressing a series of existing challenges in this field. This research aims to systematically identify and address the shortcomings of existing auto-injection systems through the aforementioned approaches while also focusing on sustainability through the incorporation of recyclable materials.

Method

In this research, the Ishikawa method was used to identify problems, and TRIZ was employed to find their ideal solution. Furthermore, the NX software was used to design new CAD data, as well as NX Nastran to analyze the model and 3D printing technology to produce prototypes. These methods and tools are detailed in this section.

Ishikawa diagram method

Ishikawa is a prominent figure in quality management, best known for his development of the Ishikawa diagram, a tool for root cause analysis. This diagram categorizes the potential causes of a problem, providing a systematic approach to problem-solving. [23]. This type of diagram is also known as the *fishbone method*, and it is depicted in Fig. 2 [24].

TRIZ

TRIZ was initially developed by Genrich Altshuller and his colleagues in the mid-1940s. It is a systematic and practical methodology for designers/engineers to analyze inventive problems and further resolve them in a strategy-driven manner [25], and it is much less experience-dependent than many existing design methods, which rely too heavily on specific previous experience and thus limit potential innovation [26], [27], [28]. During the application of this method, while one desired product feature is improved, a different feature is bound to deteriorate in response. Altshuller developed a 40 x 40 matrix that describes the relationship between these features. TRIZ is a method that allows solving problems using this matrix. It is a logical and practical in comparison with other methods, as illustrated in Fig. 1.

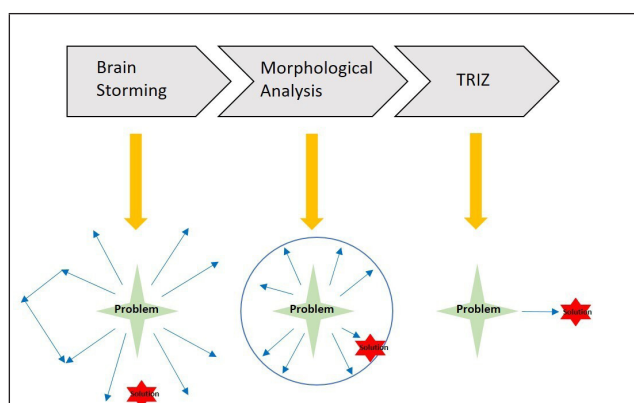


Figure 1. Comparison of three different problem-solving methods
Source: Authors

TRIZ is a convenient method to be used together with the Ishikawa diagram. This integration is depicted in Fig. 2.

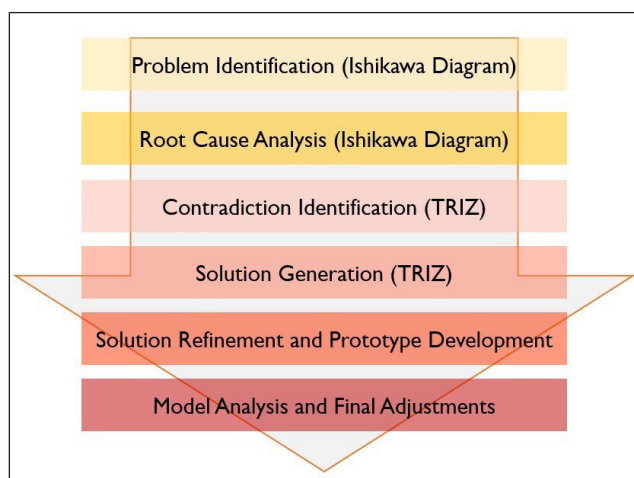


Figure 2. TRIZ-Ishikawa method
Source: Authors

In our approach, the application of the Ishikawa diagram and the TRIZ methodology is structured in a series of clear steps that ensure systematic problem identification and innovative solution development. The following steps outline the framework used in this study:

1. Problem identification (Ishikawa diagram). The initial phase aims to pinpoint and classify the fundamental causes of the challenges faced by the auto-injection system. Utilizing the Ishikawa diagram, the problem is dissected into essential categories, including design, materials, processes, user interaction, and safety, thereby offering a comprehensive view of the factors at play.
2. Root cause analysis (Ishikawa diagram). During this stage, the previously identified causes are assessed and ranked according to their importance. This targeted analysis sheds light on the most pressing issues, paving the way for the subsequent phases of the process.
3. Identification of contradictions (TRIZ). The TRIZ approach is employed to uncover the contradictions present in the system, e.g., the challenge of balancing usability with cost or enhancing functionality without exacerbating environmental concerns. The TRIZ matrix serves as a tool to map these contradictions, facilitating the creation of innovative solutions.
4. Solution generation (TRIZ). In this stage, creative solutions are formulated by applying the 40 inventive principles of TRIZ in order to solve the identified contradictions. The emphasis is on improving the design while preserving critical factors such as safety and cost efficiency.
5. Solution refinement and prototype development. The proposed solutions are polished and converted into detailed designs using CAD software (NX), followed by the creation of prototypes through 3D printing. This process enables swift testing and iterative enhancements based on user feedback.
6. Model analysis and final adjustments. The prototypes are subjected to performance evaluations using NX Nastran to verify compliance with engineering and safety standards. Necessary modifications are implemented, culminating in a fully optimized auto-injection system.

Problems and solutions

It is known that, in current practice, patients pay high costs to purchase autoinjectors. Therefore, one of the most important associated parameters is the need to ensure an engineering development that allows producing the product at a lower cost. To this effect, emphasis was placed on the use of cheaper materials in the design, on less costly and easier production methods, and on the use of minimum parts.

The main goal should be to design the most comfortable version possible for the user. In this context, another problem that we identified was the assembly of the product. In addition to being easy to assemble the product, for the sake of reusability, it is important for users to be able to perform re-assembly on their own. In this regard, the design must be reusable, in addition to complying with engineering tolerance calculations.

Since user comfort is the primary objective, an ergonomic design should be achieved wherein users will have no difficulties in applying the product themselves. They should be able to comfortably grasp the product and apply it with one hand, for which the location and shape of the trigger area is also important.

Since the market that the product appeals to is the healthcare sector, it is important for the designed product to meet hygiene standards in order to avoid any complications. In addition, the fact that the syringes used for treatment differ in size and are disposable represents hygiene issues for general-purpose products. In order for the design to be translated into practice, it must comply with certain engineering rules. In this regard, the static displacement of the main body, the properties of the materials with which the product will be designed, some dynamic calculations for the moving area, and the design parameters should not pose an issue in terms of engineering.

Finally, given the mounting environmental pollution issues, we aimed to make the product environmentally friendly. To this effect, we used a recyclable material and carried out the necessary medical certification procedures [29] to release the product to the market.

The Ishikawa diagram shown in Fig. 3 was used to identify the main problems.

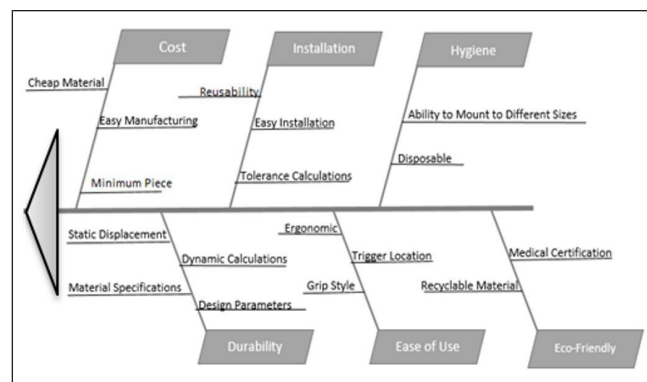


Figure 3. Ishikawa diagram

Source: Authors

In the matrix, note that there are six main problems, with their corresponding solution alternatives (Fig. 3). To find the best way to solve these problems, we used the TRIZ matrix, which is presented in Fig. 4.

Cost

As mentioned in the introduction, it is difficult to cover the cost of alternative devices that serve the studied purpose. Ahead of price hikes, three options have been identified among the procedures that need to be implemented for a satisfactory outcome: changing the device material for a cheaper one, making production easier and cheaper by using fewer parts, and adopting a design that can be

TRIZ Matrix									
Problems	Worsening Feature								Ideal Final Solution & How it is applied to the design / Integration
	Improving Feature	15-Duration of Action of Moving Objects	16-Duration of Action by Stationary Objects	21-Power	26-Quantity of substance / the matter	32-Ease of Manufacture	33-Ease of Operation	39-Productivity	
Durability	16-Duration of Action by Stationary Objects				3 31 35				Parameter Change - The material and the design was changed acc. to durability analyses
Cost	26-Quantity of substance / the matter	3 10 35 40	3 31 35						Partial Quality: Minimum number of parts is targeted.
Ease of Use	33-Ease of Operation			2 10 34 35					Prior Action: Prototyped and had been tried by a user.
Eco-Friendly (Recyclable)	34-Ease of Repair					1 10 11 35			Parameter Change: Materials were selected to be suitable for recycling
Hygiene	35-Adaptability or Versatility		2 16				1 15 16 34		Partial or Overwork: Extra ribs added.
Installation	38-Extent of Automation							5 12 26 35	Combining: Easy assembly and ergonomics are targeted.

Figure 4. TRIZ matrix

Source: Authors

produced at a lower cost. In this way, the cost of the product can be reduced, as well as its prices. These solutions are suggestions provided by the morphological analysis, but the TRIZ matrix was used to find the most appropriate of them. When using TRIZ, one feature improves while another deteriorates. Within the scope of the cost problem, the quantity of substance was chosen as the feature to be improved. In addition, the duration of object movement and the duration of the stationary objects' action were selected as the deteriorating features. In the examined 39 x 39 TRIZ matrix, solution number 3 (partial quality), located at the intersection of the mentioned features, was found to be the most suitable alternative. It was concluded that a model with fewer parts should be designed.

Installation

Another problem is the assembly of the product. Designs that are not user-friendly and are extremely complex cause difficulties for the user in terms of cost and application. Current products are either complex or can be used only once. To avoid this, the design must be easy to assemble and suitable for multiple uses. In order to find a specific solution, we used the TRIZ matrix again. Here, increasing the level of automation entailed a loss in terms of productivity. As shown in Fig. 4, the most suitable solution is the combined method (number 5), which targets easy assembly and ergonomics are targeted.

Hygiene

The product type under study has been used for treating various diseases. In this context, hygiene is important for preventing undesirable complications. Nevertheless, the use of some current devices causes serious hygiene issues, as it includes a transfer process. To prevent these issues, disposable needles have been used, but they complicate self-administration. Therefore, we designed a system where

disposable needles can easily fit, aiming for compatibility with different syringe sizes. When we analyzed the features in the matrix, we concluded that the best alternative was solution 16, bearing in mind that, if it is difficult to fully obtain 100% a desired effect, one should seek to simplify the problem [30]. In this vein, by adding more than one rib to the system, it became possible to adapt to different syringe sizes, thereby ensuring the product's compliance with hygiene rules. The application of this solution is depicted in Fig. 5.

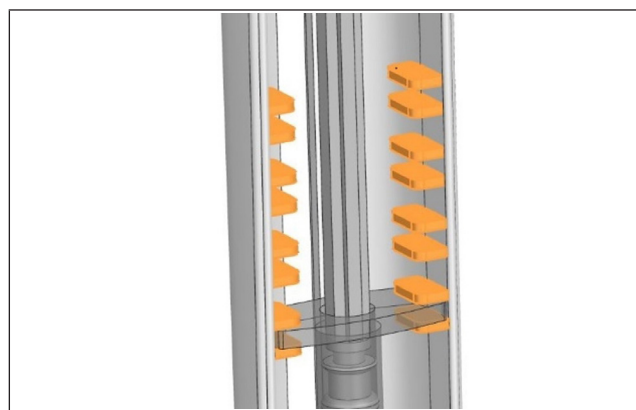


Figure 5. Rib region

Source: Authors

Durability

Durability (or strength) is one of the most important factors to consider when designing a new product, as it is important for the product to be usable in practice as well as in theory. To predict the strength of a product, static displacement, dynamic calculations, material specifications, and design parameters must be taken into account. These parameters are analyzed in detail in the solution section. With the help of the TRIZ matrix, we decided that the material's mechanical properties should be set according to durability analyses in order to increase product strength.

use, production efficiency, and recyclability. Mathematical and static analyses were performed to assess the strength and deformation of key components under operational forces, with calculations considering syringe and needle specifications. The results indicated that the spring and syringe retention areas could withstand forces of 10 N without structural failure, and that torsional and tensile loading up to 50 Nm and 20 N did not lead to plastic deformation. Additionally, the moving part of the system demonstrated elastic deformation within the material's limits. These findings confirm that the design, which includes fewer components and uses recyclable materials, is both functional and feasible for real-life applications, as demonstrated by the successful production of a prototype.

The solution found for the problems mentioned in methods section was to create a design with fewer parts that was easy to assemble and recyclable. Said design consisted of four parts: a main body, a moving part, a protective rubber cover, and a syringe.

In order to make the product easier to use, the design was made for the user to grasp and apply the product with one hand. This issue is of extra importance for self-treatment, especially in applications involving the arm. By placing the holding parts of the syringe (part number 8 in Fig. 6) in the compartments labeled as number 7, we aimed to accommodate syringes of different lengths and widths, with the help of the material used in production and the slit that forms the path of the part in motion (number 4).

Since syringes of different sizes can be mounted into the device, the user does not need to transfer any product, thus preventing the hygiene complications that may arise during this process. In addition, a protective rubber cover (number 11) was added to keep the tip of the needle sterile.

During the production of the model, polypropylene random copolymer (Bormed RG835MO) [32] was used to create a cheaper and recyclable design compared to other materials. The fact that the design consists of fewer parts not only facilitates production, it also makes assembly and application easier.

For increased recyclability, the entire main body of the product was made of a recyclable flexible plastic material. In addition, rubber bands were used – which are easier to recycle – instead of flexible springs, which are used by alternatives on the market. This led to more environmentally friendly and cheaper product. Since rubber bands are easier to access, they can be purchased at a lower cost than their counterparts in the market.

In order to increase the strength of the design, the parts with potential strength issues were mathematically and statically analyzed to observe their deformation under specific forces. For the corresponding calculations, the dimensions of the most commonly used syringes and needles were taken as reference [33]:

- Needle length: 4, 5, 6, 8, 10, and 12 mm
- Needle tip inner diameter: 29, 30, 31, and 32 G \approx 0.286, 0.255, 0.227, and 0.202 mm
- Syringe capacity: 0.3, 0.5, 1.0, and 2.0 ml
- Minimum needle inner diameter according to ISO standards (9626:2016) [34]: 0.114 and 0.125 mm for 31G; 0.089 and 0.105 mm for 32 G
- Rubber bands:
 - Length: 79 ± 1 mm
 - Height: 10 ± 1 mm
 - Width: 10 ± 1 mm [35]
 - Minimum tensile strength: 15.5 MPa
 - Minimum elongation at break: 650% [36]
- Density of insulin: 1.09 g/cm³

Calculations

Calculations were made based on the average values specified in the previous sections and using the law of conservation of energy, the final speed of the moving part when it hits the piston was found. A pressure value was found using terminal velocity and Bernoulli's principle. Using the pressure value found and Poiseuille's law for laminar flow formula, the net force applied on the piston was calculated.

Bernoulli equation (1)

$$P_1/\rho + (V_1^2)/2 + G_{z1} = P_2/\rho + (V_2^2)/2 + G_{z2} = \text{const.}$$

$$P_1 = 101.325 \text{ kPa}$$

$$\rho = 1.09 \text{ g/m}^3$$

$$V_1 = 0$$

$$G = 9.81 \text{ m/s}^2$$

$$z_1 = 0.55 \text{ mm}$$

$$\rho = 1.09 \text{ g/m}^3$$

$$V_2 = 15.66 \text{ [37]}$$

$$G = 9.81 \text{ m/s}^2$$

$$z_2 = 0$$

$$101.325 \text{ kPa}/(1.09 \text{ g/m}^3) + 0^2/2 + 9.81 \text{ m/s}^2 \times 0.055 \text{ m} = P_2/(1.09 \text{ g/m}^3) + \{(15.66 \text{ m/s})^2\}/2 + 9.81 \text{ m/s}^2 \times 0$$

$$P_2 = 133 \text{ Pa}$$

Poiseuille's law for laminar flow (2)

$$P = F/(n \times \pi \times (D/2)^2)$$

$$F = P \times n \times \pi \times (D/2)^2$$

$$\pi = 3.14$$

$$P_2 = 133 \text{ Pa}$$

$$D = 0.255 \text{ mm}$$

$$n = 1 [34]$$

$$z_2 = 0$$

$$F = 6.66 \text{ N}$$

With a safety factor $S = 1.5$

$$F \cdot S = 6.6 \cdot 1.5 = 10 \text{ N}$$

Needle forces must be at least 10 N

Force provided by rubber bands (3)

$$F = k \times \Delta x$$

$$\Delta x = x_2 - x_1$$

$$k = 38.8 \text{ N/m}$$

Based on the results, we concluded that the net force that should be applied to the piston should be 10 N, as a consequence of the tensile forces exerted by the rubber bands. Since two rubber bands were used to this effect, we sought to calculate the load per rubber band and the distance between the corresponding connection points. This was done using the elastic potential energy formula.

$$2 \times F = 10 \text{ N}$$

$$F = 5 \text{ N}$$

$$5 \text{ N} = 38.8 \text{ N/m} \times \Delta x \rightarrow \Delta x \approx 0.13 \text{ m}$$

The rubber bands were arranged in double layers. Moreover, the length loss at the connection points was considered.

$$\text{Connection points: } 2 \times 0.008 \text{ m}$$

$$0.065 \text{ m} - 0.008 \text{ m} = 0.057 \text{ m}$$

Considering these results, the distance between the connection points is 0.06 m.

Computer-aided analysis

In line with the calculated mathematical values, the parts that could experience strength issues were statically analyzed using NX Nastran. The mesh quality was calculated via the tunnel elastic and elasto-plastic deformation approach [38]. This study used the CTERA10 mesh type, with a standard element size of 1 mm. A maximum surface growth rate of 1.3 was selected, for a total of 1626 mesh elements across all components.

Analyzing the spring retention area: force = 10 N

Firstly, a static analysis of the spring retention area was performed. Since this area supports the rubber bands, its durability is very important. Considering the calculations presented in the previous section, when the system experiences 10 N of applied force, the maximum stress is 3.1

MPa. Since the strength of the material used in the design is 27.5 MPa [39], the spring retention area can withstand a force of 10 N. The analysis for this section is presented in Fig. 9.

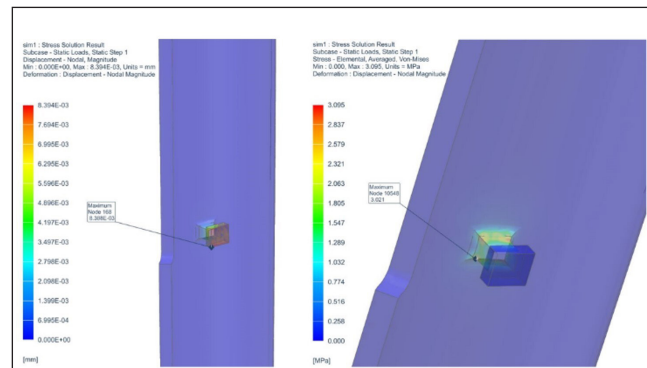


Figure 9. Static analysis of the spring retention area under a 10 N force
Source: Authors

Syringe retention (rib) area analysis: force = 10 N

Another area where the applied force is likely to have a negative effect on the product is the syringe retention (rib) area. As seen in Fig. 10, the 10 N deformation analysis yields a maximum value of 5.2 MPa. This is the maximum level of deformation that the material can withstand. It was determined that the rib region will not generate strength issues under the influence of a 10 N force, since the value obtained is below 27.5 MPa.

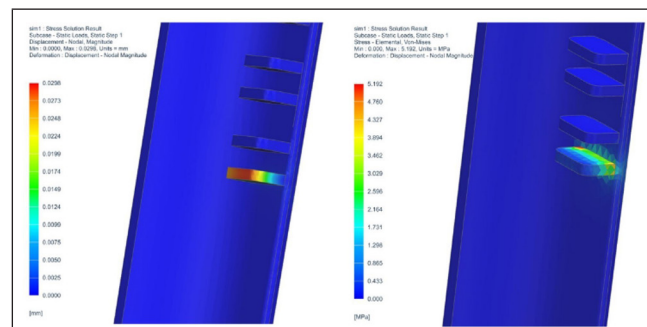


Figure 10. Static analysis of the syringe retention (rib) area under a 10 N force
Source: Authors

Syringe torsional analysis: torque = 50 Nm

When a torsional force of 50 Nm was applied to the model while considering the safety factor, the maximum stress level was 1.4 MPa (Fig. 11). Given the properties of the material used, the part will not undergo plastic deformation under 50 Nm of torsional force.

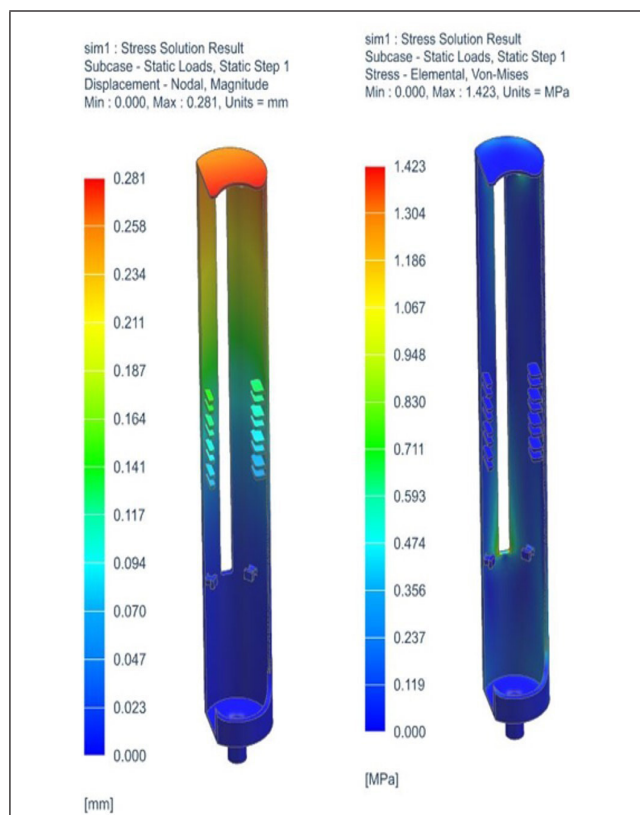


Figure 11. Static torsion analysis of the syringe under a 50 N torque
Source: Authors

Effect of rubber forces: force = 20 N

In order to determine the potential deformation effects arising from the tensile forces exerted by the rubber band, the analysis shown in Fig. 12 was conducted. According to the results, the model experiences 11.4 MPa under 20 N of tensile force. This value can be withstood by the material used.

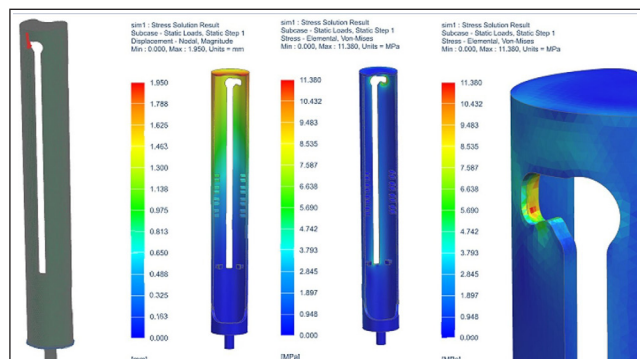


Figure 12. Static analysis of the effect of rubber forces (20 N)
Source: Authors

Moving parts analysis: force = 10 N

Finally, we analyzed the moving part, which was subjected to a maximum deformation of 10.3 MPa (under a 10 N force), as seen in Fig. 13. Considering the mechanical properties

of the material used, this part remained within the elastic deformation zone and was an obstacle for the prototype.

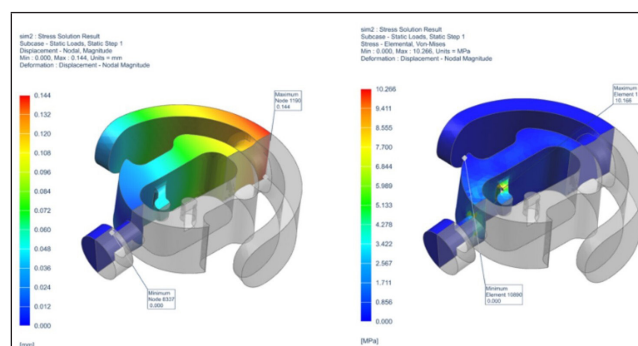


Figure 13. Static analysis of the moving part under a 10N force
Source: Authors

As a result of the solution suggestions and engineering calculations made in this research, a prototype of the design that is applicable in real life was produced. This prototype is presented in Figs. 14 and 15.



Figure 14. Unassembled prototype
Source: Authors

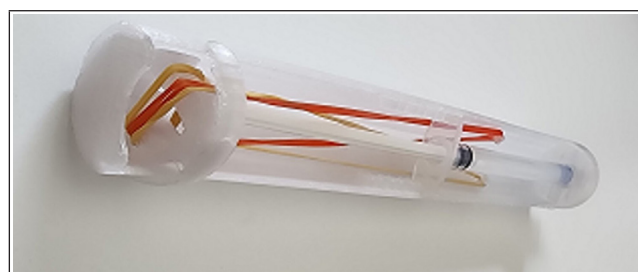


Figure 15. Assembled prototype
Source: Authors

Discussion

This research created an advanced auto-injection system by employing the TRIZ problem-solving methodology in conjunction with Ishikawa analysis to tackle issues related to usability, cost, assembly complexity, and hygiene. The proposed design incorporates polypropylene random

copolymer (Bormed RG835MO), which not only offers cost efficiency but also promotes recyclability, distinguishing it from the conventional materials typically utilized in similar devices.

A key benefit of this design is its significantly reduced number of components, which simplifies the manufacturing process and facilitates both assembly and application. This improvement is akin to other developments that advocate for streamlined designs to minimize complexity and enhance user experience [40]. The substitution of flexible springs with rubber bands further contributes to the design's environmental sustainability and cost-effectiveness, as rubber bands are more accessible and affordable than other alternatives on the market.

The literature has discussed some of the frequent mistakes made when using auto-injectors. These mistakes include not applying pressure to the injection site, overlooking the expiration date, failing to check the medication window, and not rotating the injection site for a second injection [41]. In this work, static and dynamic analyses performed using NX Nastran confirmed that our reengineered auto-injection system maintains adequate structural integrity. The analysis of the spring retention area indicated that the design can endure a force of 10 N, with a maximum deformation of 3.1 MPa, which is significantly below the material's strength threshold of 27.5 MPa. Additionally, evaluations of the syringe's retention area and torsional properties demonstrated that the system can withstand applied forces without experiencing plastic deformation, corroborating the findings of related studies on the mechanical characteristics of injection systems.

The emphasis on recyclability and environmental considerations in the field of medical devices reflects a growing commitment to sustainability. Studies have emphasized the need to incorporate environmentally friendly materials into the design of medical devices to support public health and reduce ecological damage [10]. The substitution of rubber bands for conventional springs not only advances this sustainability objective but also lowers manufacturing costs, thereby enhancing accessibility for a broader patient demographic.

Throughout the evolution of engineering systems, the advantages of employing the TRIZ methodology in crafting innovative solutions that simultaneously tackle various design challenges has become evident [16]. The methodical implementation of TRIZ principles in this work led to a design that fulfills functional specifications while also improving user convenience and promoting environmental sustainability.

Studies examining the use of various drugs have shown that autoinjectors facilitate faster injection times and minimal tissue effects, especially when dealing with higher-viscosity formulations [42].

It has been observed that the triggering of autoinjectors varies in terms of comfort and maximum force depending on the age and dexterity of the user. In addition, the applied forces are affected by the geometry of the device, as well as by the way it is held [43]. However, difficulties such as limited dexterity and anxiety related to injection may hinder an adequate application [44]. In this regard, it should be noted that this study did not test the system in groups of subjects in different age ranges.

Autoinjectors play an important role in the self-administration of drugs. However, the EU's understanding of a specific medical device development tool (MDDT) may cause inconsistencies regarding quality and regulatory compliance. Therefore, it is essential to conduct studies aimed at ensuring compliance with the standards set within this framework [45], [46].

Studies have shown that integrating sustainability into medical device development results in various useful approaches to evaluate sustainability criteria. These approaches help to evaluate design concepts and improve the overall design [47]. Various sustainability problems have been partially solved in this study.

The concept of *quality* should be understood in a broad context that includes factors such as material purity, biocompatibility, device reliability, recall rates, and acceptability for patients. Social, demographic, and economic factors also play an important role in determining the sustainability of medical devices. Careful selection, use, and waste management positively affect sustainability when environmental impacts are considered throughout the lifecycle of the device, which includes design, production, clinical application, and final disposal [12]. Studying the use of environmentally recyclable materials and the sustainability of waste management will contribute to addressing product development issues in the medical sector.

Despite the fact that the design was developed through the TRIZ methodology with a focus on customer requirements, concerns regarding the durability and long-term usability of our proposal may arise. From this viewpoint, the use of rubber bands could lead to a decline in functionality over time, necessitating user replacement. Additionally, the predominance of plastic in the structure increases the risk of breakage or deformation under unforeseen circumstances, such as impacts from falls or excessive weight. Future research aimed at establishing the minimum lifespan of the device through durability, storage conditions, and usage assessments will significantly contribute to advancements in the medical field. Furthermore, investigating the target demographic in relation to user experiences and preferences as well as market performance could provide valuable insights.

The results suggest that integrating the TRIZ and Ishikawa methodologies can substantially enhance both the design and operational efficacy of medical devices. Future works

should focus on assessing the long-term performance and user satisfaction associated with our revamped auto-injection system within clinical environments. Furthermore, applying these approaches to various medical device categories may result in widespread advancements throughout the healthcare sector.

Conclusion

This article provides a comprehensive investigation aimed at enhancing user experience in a cost-effective and practical way through two prominent methodologies: TRIZ and Ishikawa analysis. By leveraging these methodologies, our research established a robust framework for identifying specific improvement areas. In summary, after identifying the issues associated with using auto-injectors in the medical field through Ishikawa analysis, the TRIZ approach was employed to develop a matrix that served as a strategic problem-solving instrument.

The design process, guided by TRIZ to tackle the aforementioned issues, was marked by iterative improvements that incorporated feedback and insights from initial analyses. Rubber bands were selected as an alternative to metal springs, promoting both environmental sustainability and cost-efficiency, while thermoplastic materials were utilized for the main body of the device. To assess the practicality and reliability of the proposed design, static analyses were conducted through a series of engineering calculations. Both the static and dynamic analyses, which employed using NX Nastran, confirmed that the design maintains a sufficient structural integrity, with its spring retention area, syringe retention area, and torsional properties validating the system's capacity to endure applied forces without experiencing plastic deformation. Engineering validations through static and dynamic simulations showed that the design safely withstands up to 10 N of applied force, with the maximum stress levels remaining below 5.2 MPa, well within the material's 27.5 MPa yield strength. This provided essential insights into the structural integrity and operational capabilities of the proposed product, thereby affirming its viability for practical application.

The emphasis on utilizing recyclable materials highlights an increasing commitment to sustainability within the medical device sector. This design approach is in harmony with contemporary healthcare objectives, as it seeks to reduce both environmental impact and production expenses. The implementation of the TRIZ methodology played a crucial role in fostering innovative solutions that tackled various design obstacles, thereby improving functional attributes and user comfort. Additionally, it streamlined the assembly process by reducing the number of components involved, leading to enhancements in both cost efficiency and usability.

Nevertheless, some concerns persist regarding the system's durability, long-term usability, and overall lifespan. Although the economic benefits are clear, the dependence on rubber

bands necessitates frequent replacements, and the extensive use of plastic materials raises the potential for breakage or deformation under unforeseen circumstances. Future investigations should aim to establish the minimum lifespan of the device through rigorous durability testing, as well as to analyze user experiences and preferences to further optimize the design.

The proposed design offers notable enhancements in terms of user experience, sustainability, and cost-effectiveness, but it also has significant limitations. As mentioned above, the reliance on rubber bands, while economical and environmentally friendly, raises durability issues due to the need for frequent replacements. Additionally, the extensive use of thermoplastic materials, beneficial for manufacturing and recyclability, may lead to breakage or deformation under unexpected stress. Although static and dynamic analyses confirm structural integrity, these tests were conducted under controlled conditions, and real-world performance may involve unforeseen variables. Moreover, the long-term usability and lifespan of the device are still unverified, necessitating further durability assessments and user feedback in order to ensure reliability and practicality. These limitations underscore the need for continued research to enhance the design and evaluate its long-term effectiveness in diverse healthcare settings.

Future research should concentrate on assessing the long-term efficacy of the redesigned automated injection system in clinical environments, in addition to considering the application of these methodologies across a wider array of medical devices, facilitating substantial progress within the healthcare field. The results indicate that, by prioritizing user-centered solutions and employing systematic problem-solving techniques, effective treatment instruments can be developed which significantly enhance patient outcomes and satisfaction.

Acknowledgements

The author expresses gratitude to the editorial team of *Ingeniería e Investigación*, with particular thanks to José Daniel Gutiérrez Mendoza for their insightful comments, which contributed significantly to the enhancement of the manuscript.

CRediT author statement

The author handled the entire conceptualization process, from idea generation and research to workflow formulation and assessment. He also designed the component, wrote the manuscript, and conducted structural analyses, calculations, and prototype development.

Conflicts of interest

The author has no conflicts of interest to declare.

References

- [1] P. Gangane, N. Mahajan, U. Ahajan, and S. Hiranwar, "Delivering biologics in prefilled syringes: an innovation in parenteral packaging," *Int. J. Pharm. Res. Technol.*, vol. 10, no. 1, 2020. <https://doi.org/10.31838/ijprt/10.01.11>
- [2] S. Boccaletti et al., "Systematic literature review of asthma biologic self-administration enhanced by a patient perspective," *J. Allergy Clin. Immunol. Glob.*, vol. 3, no. 4, art. 100334, Aug. 2024. <https://doi.org/10.1016/j.jacig.2024.100334>
- [3] J. Borrás-Blasco, R. A. García, S. Cornejo-Uixeda, M. Matellanes-Palacios, and E. Casterá-Melchor, "Patient preference after switching guselkumab from prefilled syringe to an autoinjection pen in psoriasis and psoriatic arthritis patients," *Farm. Hosp.*, vol. 49, no. 3, pp. 160–163, Jul. 2024. <https://doi.org/10.1016/j.farma.2024.07.002>
- [4] H. Forcinio, "Innovations in prefilled biologics," *Pharm. Technol.*, vol. 48, no. 4, pp. 24–27, Apr. 2024. [Online]. Available: <https://www.pharmtech.com/view/innovations-in-prefilled-biologics>
- [5] S. Vermeire et al., "Preference for a prefilled syringe or an auto-injection device for delivering golimumab in patients with moderate-to-severe ulcerative colitis: a randomized crossover study," *Patient Prefer. Adherence*, vol. 12, pp. 1193–1202, Jul. 2018. <https://doi.org/10.2147/ppa.s154181>
- [6] P. Dostál, J. Taubel, U. Lorch, V. Aggarwal, and T. York, "The reliability of auto-injectors in clinical use: a systematic review," *Cureus*, vol. 15, no. 7, art. e41601, Jul. 2023. <https://doi.org/10.7759/cureus.41601>
- [7] B. Canaud et al., "Digital health support: current status and future development for enhancing dialysis patient care and empowering patients," *Toxins*, vol. 16, no. 5, art. 211, Apr. 2024. <https://doi.org/10.3390/toxins16050211>
- [8] A. Selvaraj, A. Kulkarni, and J. M. Pearce, "Open-source 3-D printable autoinjector: design, testing, and regulatory limitations," *PLoS ONE*, vol. 18, no. 7, art. e0288696, Jul. 2023. <https://doi.org/10.1371/journal.pone.0288696>
- [9] P. Kumar, S. Wang, and S. Kristensen, "EcoDesign for medical devices barriers and opportunities to eco-effective design of medical devices," MPhil thesis, Royal Coll. Art, London, UK, 2018. [Online]. Available: [https://researchonline.rca.ac.uk/4467/1/MPhil%20Thesis%20-%20Eco-Design%20for%20Medical%20Devices%20\(Deposit%20Version\).pdf](https://researchonline.rca.ac.uk/4467/1/MPhil%20Thesis%20-%20Eco-Design%20for%20Medical%20Devices%20(Deposit%20Version).pdf)
- [10] H. Singh et al., "Harnessing the foundation of biomedical waste management for fostering public health: strategies and policies for a clean and safer environment," *Discov. Appl. Sci.*, vol. 6, art. 89, Feb. 2024. <https://doi.org/10.1007/s42452-024-05735-2>
- [11] E. Moshkbid, D. E. Cree, L. Bradford, and W. Zhang, "Biodegradable alternatives to plastic in medical equipment: current state, challenges, and the future," *J. Compos. Sci.*, vol. 8, no. 9, art. 342, Sep. 2024. <https://doi.org/10.3390/jcs8090342>
- [12] J. Vienken and C. Boccato, "Do medical devices contribute to sustainability? The role of innovative polymers and device design," *Int. J. Artif. Organs*, vol. 47, no. 4, pp. 240–250, Apr. 2024. <https://doi.org/10.1177/03913988241245013>
- [13] Siemens, "Siemens NX | June 2023," *Siemens Digital Ind. Softw.*, 2023. Accessed: Nov. 13, 2024. [Online]. Available: <https://plm.sw.siemens.com/en-US/nx/news/la-test-version-june-2023>
- [14] B. Quronfuleh, D. Sleath, and S. Rahimifard, "Circular economy for medical devices: a case study of syringes," *Procedia CIRP*, vol. 122, pp. 449–454, 2024. <https://doi.org/10.1016/j.procir.2024.01.065>
- [15] A. Antalffy et al., "The adherence and outcomes benefits of using a connected, reusable auto-injector for self-injecting biologics: a narrative review," *Adv. Ther.*, vol. 40, no. 11, pp. 4758–4776, Sep. 2023. <https://doi.org/10.1007/s12325-023-02671-2>
- [16] U. Müller-Ladner, C. Edwards, and A. Erkens, "International survey to evaluate current options for subcutaneous injection of methotrexate (MTX) and a new button-free MTX auto-injector," *Patient Prefer. Adherence*, vol. 18, pp. 579–590, Mar. 2024. <https://doi.org/10.2147/ppa.s440818>
- [17] S. Karupppasamy, "A methodology to assess sustainability in medical device design," MS thesis, Patrick Power Library, Saint Mary's Univ., Halifax, NS, Canada, 2022. [Online]. Available: <https://library2.smu.ca/handle/01/31260>
- [18] I. Bukhman, *Technology for Innovation*, 2021. [Online]. Available: <https://books.google.com.tr/ks?hl=tr&lr=&id=Sro1EAAAQBAJ&oi=fnd&pg=PR5&dq=I.+Bukhman>
- [19] A. Baptista et al., "Applying DMADV on the industrialization of updated components in the automotive sector: a case study," *Procedia Manuf.*, vol. 51, pp. 1332–1339, 2020. <https://doi.org/10.1016/j.promfg.2020.10.186>
- [20] P. Neves et al., "Implementing lean tools in the manufacturing process of trimmings products," *Procedia Manuf.*, vol. 17, pp. 696–704, 2018. <https://doi.org/10.1016/j.promfg.2018.10.119>
- [21] T. Costa, F. J. G. Silva, and L. Pinto Ferreira, "Improve the extrusion process in tire production using Six Sigma methodology," *Procedia Manuf.*, vol. 13, pp. 1104–1111, 2017. <https://doi.org/10.1016/j.promfg.2017.09.171>
- [22] S. Elyoussoufi, M. Mazouzi, A. Cherrafi, and E. M. Tamasna, "TRIZ-ISHIKAWA diagram, a new tool for detecting influencing factors: a case study in HVAC business," in *Proc. IEOM Soc. Int.*, Oct. 2022, pp. 3673–3680. [Online]. Available: <https://ieomsociety.org/proceedings/2022istanbul/670.pdf>
- [23] Y. Akao, *Quality Function Deployment*, 2024. <https://doi.org/10.4324/9781003578833>
- [24] F. Pakdil, *Six sigma for students*. Cham, Switzerland: Springer Int. Publ., 2020. <https://doi.org/10.1007/978-3-030-40709-4>
- [25] D. D. Sheu, M.-C. Chiu, and D. Cayard, "The 7 pillars of TRIZ philosophies," *Comput. Ind. Eng.*, vol. 146, art. 106572, Aug. 2020. <https://doi.org/10.1016/j.cie.2020.106572>
- [26] K. Hmina et al., "TRIZ effect exploitation on engineering students thinking skills in product design," *FME Trans.*, vol. 52, no. 2, pp. 271–278, 2024. <https://doi.org/10.5937/fme2402271h>
- [27] V. Sojka, "Use of TRIZ, and TRIZ with other tools for process improvement: a literature review," *Emerg. Sci. J.*, vol. 4, no. 5, art. 17, Jan. 2020. <https://doi.org/10.28991/esj-2020-01234>

- [28] S. Guner and I. Kose, "Yaratıcı problem çözme tekniği TRIZ'in uygulama alanları: sağlık hizmetleri sektöründe kullanımı için öneriler," *J. Econ. Bus. Finance Res.*, vol. 2, no. 2, pp. 185–205, Aug. 2020. <https://doi.org/10.38009/ekimad.763538>
- [29] B. Karthika and A. R. Vijayakumar, "ISO 13485: Medical devices – quality management systems, requirements for regulatory purposes," in *Medical Device Guidelines and Regulations Handbook* P. S. Timiri Shanmugam, P. Thanagaraju, N. Palani, and T. Sampath, Eds. Cham, Germany: Springer, 2022, pp. 19–29. https://doi.org/10.1007/978-3-030-91855-2_2
- [30] M. Ghane *et al.*, "TRIZ trend of engineering system evolution: a review on applications, benefits, challenges and enhancement with computer-aided aspects," *Comput. Ind. Eng.*, vol. 174, art. 108833, Dec. 2022. <https://doi.org/10.1016/j.cie.2022.108833>
- [31] R. Reshmy *et al.*, "Promising eco-friendly biomaterials for future biomedicine: cleaner production and applications of nanocellulose," *Environ. Technol. Innov.*, vol. 24, art. 101855, Nov. 2021. <https://doi.org/10.1016/j.eti.2021.101855>
- [32] Borealis, "Polypropylene Bormed™ RG835MO polypropylene random copolymer," Jun. 2022. [Online]. Available: https://www.borealisgroup.com/storage/Datasheets/bormed/rg835mo/RG835MO-PDS-REG_WORLD-EN-V7-PDS-WORLD-53816-PDS_BORMED%20RG-835MO_7_28062022.pdf
- [33] S. Şen, "Insulin needles in adults and children," *Klin. Tip Bilim. Derg.*, vol. 7, no. 3, pp. 32–35, Apr. 2019.
- [34] International Organization for Standardization, "Stainless steel needle tubing for the manufacture of medical devices – requirements and test methods," ISO 9626:2016, 2016. [Online]. Available: <https://www.iso.org/standard/60480.html>
- [35] E. Pojer *et al.*, "The case for anthocyanin consumption to promote human health: a review: anthocyanins and human health," *Compr. Rev. Food Sci. Food Saf.*, vol. 12, no. 5, pp. 483–508, Sep. 2013. <https://doi.org/10.1111/1541-4337.12024>
- [36] International Organization for Standardization, "Rubber bands - general requirements and test methods," ISO 22843:2020, 2020. [Online]. Available: <https://www.iso.org/standard/74031.html>
- [37] E. Zijlstra *et al.*, "Impact of injection speed, volume, and site on pain sensation," *J. Diabetes Sci. Technol.*, vol. 12, no. 1, pp. 163–168, Oct. 2017. <https://doi.org/10.1177/1932296817735121>
- [38] K. N. Clayton *et al.*, "Measuring biotherapeutic viscosity and degradation on-chip with particle diffusometry," *Lab Chip*, vol. 17, no. 23, pp. 4148–4159, Jan. 2017. <https://doi.org/10.1039/c7lc00507e>
- [39] L. N. Equihua, E. Á. Cornejo, and Y. C. Sánchez, "Evaluation of tunnel elastic and elasto-plastic deformations with approximations obtained from 3D-FEM simulations," *Ing. Investig.*, vol. 43, no. 2, art. 7, 2023. [Online]. Available: <https://dialnet.unirioja.es/descarga/articulo/9070282.pdf>
- [40] K. T. Samenjo *et al.*, "Design of a syringe extension device (Chloe SED®) for low-resource settings in sub-Saharan Africa: a circular economy approach," *Front. Med. Technol.*, vol. 5, pp. 01–18, Sep. 2023. <https://doi.org/10.3389/fmedt.2023.1183179>
- [41] T. Serrecchia *et al.*, "Human factors validation for a rheumatoid arthritis auto-injector for the adalimumab biosimilar FKB327," *Int. J. Hum. Factors Ergonom.*, vol. 7, no. 2, art. 144, 2020. <https://doi.org/10.1504/ijhfe.2020.109561>
- [42] B. C. Roberts *et al.*, "Novel cannula design improves large volume auto-injection rates for high viscosity solutions," *Drug Deliv.*, vol. 29, no. 1, pp. 43–51, Dec. 2022. <https://doi.org/10.1080/10717544.2021.2018069>
- [43] J. Lange *et al.*, "Quantifying patient capabilities and setting the stage for future development: Insights from a sensor-augmented simulated use study with pen injectors," *Med. Devices Evid. Res.*, vol. 17, pp. 271–283, Jul. 2024. <https://doi.org/10.2147/mder.s478141>
- [44] K. Berman *et al.*, "Design development of the SMART-CLIC®/CLICWISE® injection device for self-administered subcutaneous therapies: findings from usability and human factor studies," *Adv. Ther.*, vol. 40, no. 7, pp. 3070–3086, May 2023. <https://doi.org/10.1007/s12325-023-02512-2>
- [45] M. L. Machal, "Framework for creating a qualified medical device development tool of autoinjectors," *Front. Med. Technol.*, vol. 5, art. 1281403, Nov. 2023. <https://doi.org/10.3389/fmedt.2023.1281403>
- [46] A. B. Singh, C. Khandelwal, and G. S. Dangayach, "Revolutionizing healthcare materials: innovations in processing, advancements, and challenges for enhanced medical device integration and performance," *J. Micromanuf.*, vol. 0, art. 0, Aug. 2024. <https://doi.org/10.1177/25165984241256234>
- [47] C. Boccatto and J. Vienken, "Do medical devices contribute to sustainability? Environmental, societal and governance aspects," *Int. J. Artif. Organs*, vol. 47, no. 4, pp. 229–239, Apr. 2024. <https://doi.org/10.1177/03913988241245015>

Development of a Solid Waste Collector Robot for Cleaning in Public Areas

Desarrollo de un robot recolector de residuos sólidos para limpieza en áreas públicas

Dick Díaz-Delgado¹ and Alexander Inga-Alva²

ABSTRACT

Solid waste management reduces pollution, protects public health, conserves ecosystems, and promotes recycling and circular economies in a sustainable manner. This study analyzes the pollution issues caused by inefficient solid waste management in urban areas, highlighting its negative impact on the environment. In this vein, the T5R-bot collection robot was developed, designed through 3D printing and implemented with an artificial vision model for object detection based on the single shot multibox detector (SSD) and the MobileNetV2 neural network architecture. This system allows for the autonomous identification and collection of up to 12 types of debris, achieving an accuracy of 98% and a mAP of 97.81%. The methodology included the mechanical design of the robot with a rocker-bogie mechanism, ultrasonic sensors for navigation, and a robotic arm with four degrees of freedom. The model was trained with a dataset collected from 2890 images, demonstrating high efficiency in detecting and collecting waste in contaminated environments in public areas. The results confirm the viability of the robot as a tool for improving solid waste management. In addition, the future integration of segregation and adaptive learning capabilities is proposed.

Keywords: IoT, robotics, waste management, deep learning, MobileNet

ABSTRACT

La gestión de desechos sólidos reduce la contaminación, protege la salud pública, conserva los ecosistemas y fomenta el reciclaje y las economías circulares de manera sostenible. Este estudio analiza la problemática de la contaminación causada por la gestión ineficiente de desechos sólidos en áreas urbanas, destacando su impacto negativo en el medio ambiente. En este orden de ideas, se desarrolló el robot recolector T5R-bot, diseñado mediante impresión 3D e implementado con un modelo de visión artificial para la detección de objetos basado en el *single shot multibox detector* (SSD) y la arquitectura de red neuronal MobileNetV2. Este sistema permite la identificación y recolección autónoma de hasta 12 tipos de residuos, alcanzando una precisión del 98 % y un mAP de 97.81 %. La metodología incluyó el diseño mecánico del robot con un mecanismo *rocker-bogie*, sensores ultrasónicos para navegación y un brazo robótico con cuatro grados de libertad. El modelo fue entrenado con un conjunto de datos recolectado a partir de 2890 imágenes, demostrando gran eficiencia en la detección y recolección de residuos en ambientes contaminados de áreas públicas. Los resultados confirman la viabilidad del robot como una herramienta para mejorar la gestión de residuos sólidos. Además, se propone la integración futura de capacidades de segregación y aprendizaje adaptativo.

Palabras clave: IoT, robótica, gestión de residuos, aprendizaje profundo, MobileNet

Received: August 16th 2024

Accepted: February 17th 2025

Introduction

Living sustainably, reducing the carbon footprint, and building a better planet for all are steps that humankind must take immediately. Today, carbon emissions into the environment are primarily generated by 1% of the world's population (wealthy individuals), who are the main contributors, in contrast to the world's poorest 50% [1]. Therefore, measures must be taken to reduce the overall consumption, such as using cleaner vehicles and adopting healthier diets, which could lower the environmental impact by 40-70% as of 2050, especially among the upper class. This is particularly important since, out of the 300 million tons of plastic produced worldwide, less than 70% is recycled [2]. Specifically, 79% ends up in the environment, 9% is recycled, and 12% is incinerated. Much of the solid waste is carried by rainwater through drains and sewers, which usually flow into larger rivers that end up in the seas [3]. This waste clogs sewers and ditches, causing floods that damage both homes and crops, contaminating them while endangering living beings and people's health.

German research confirms that 90% of the world's terrestrial plastic ends up in the sea, originating in the basins of ten large rivers, eight in Asia and two in Africa [4]. In the subtropical vortex of the North Pacific, American oceanographer Captain Charles Moore discovered an artificial island, which he described as a floating waste dump. There, he observed bottles, wrappers, containers, bags, and diapers, among other items, estimating that, by 2050, there could be more plastic than fish in the sea [5, 6]. The Caribbean is the world's second most polluted sea in terms of plastics and microplastics, as half of the urban

¹Computer engineer, Universidad César Vallejo, Tarapoto, Peru. MSc in Computer Engineering with a mention in Strategic Management, Universidad Internacional Iberoamericana de México, Campeche, México. Affiliation: Professor, Universidad César Vallejo, Tarapoto, Peru. E-mail: ddiazde@ucv.edu.pe

²Geographic engineer, Universidad Nacional Mayor de San Marcos, Lima, Peru. MSc in Software Engineering with a mention in Strategic Management, Communications, and Technology, Universidad Nacional Mayor de San Marcos, Lima, Peru. Affiliation: Professor, Universidad Nacional Mayor de San Marcos, Lima, Peru. E-mail: alexander.inga@unmsm.edu.pe



Attribution 4.0 International (CC BY 4.0) Share - Adapt

waste in this region ends up in open landfills, and 85% of its wastewater is discharged without treatment into the sea [7], causing harmful effects and health risks [8].

Urban solid waste management is an urgent global problem that is exacerbated by unregulated population growth. In Peru, despite the General Law on Solid Waste, enacted more than 15 years ago, the situation remains critical. 75% of the population lives in cities, generating 1-1.5 kg of waste per person on a daily basis, which has increased total waste production from 13 000 to 18 000 tons per day over the last decade. This has resulted in littering and pollution in urban areas and bodies of water. Addressing this issue requires a long-term public policy that integrates standards, incentives, innovative projects, technological solutions, decentralized technical assistance, and educational programs [9]. Municipalities are responsible for providing solid waste collection services to approximately 91.21% of the urban population across Peru’s 1867 districts. However, most lack an integrated management system with clear and effective implementation processes [10].

The expansion of deep learning algorithms in computer vision applications has led to state-of-the-art results in object detection systems through the automatic extraction of complex features from input images by means of self-learning. This allows computational models to process and interpret data at multiple levels of abstraction, mimicking the human brain’s ability to distinguish multi-modal information [11].

Machine learning (ML) and deep learning (DL) algorithms have also been widely adopted in mobile cleaning robots for tasks such as stain and waste detection and route planning. Additionally, 3D printing technology has facilitated the design and production of innovative components, contributing to the development of higher-quality products. In this context, our work focuses on developing an advanced object detection model by introducing improvements to the conventional MobileNetV2 architecture.

The main contributions of this study include an innovative algorithm capable of automatically identifying debris using advanced computer vision techniques, a model that achieves rapid convergence to reduce training times while enhancing detection accuracy through improved nonlinear feature extraction, the validation of a solid waste collection framework tested with an internally developed cleaning robot, and the proposal for a suspension system based on the rocker-bogie mechanism to enable safe traversal over obstacles exceeding twice the diameter of the wheels.

This document is structured as follows. First, the state of the art on object detection and solid waste collection methods is analyzed. The next section describes the methodology of the proposed approach. Then, the experimental results obtained are presented and discussed. Finally, the conclusions section summarizes the findings of this article.

Related works

Solid waste collection using mobile robots is an emerging field that addresses challenges stemming from rapid economic growth and population increase [12, 13]. While this research area has primarily focused on developing

countries, our analysis included studies from developed nations for a more comprehensive perspective. Table I lists the countries that have conducted research specifically on mobile robots optimized for solid waste detection and collection using ML and DL techniques.

Table I. Studies on mobile robots for waste collection by country

Item	Country	Studies
1	India	4
2	United States	3
3	Indonesia	2
4	Pakistan	1
5	Italy	1
6	Australia	1
7	Singapore	1
8	Malaysia	1
9	China	1
10	Germany	1
11	Bangladesh	1
12	United Kingdom	1
13	Romania	1
14	Cuba	1
15	Mexico	1
16	Brazil	1

Source: Authors

In Australia, research on waste management found that Internet of Things (IoT) and sensor network-based approaches are unsustainable for smart cities, proposing the use of swarm intelligence and the Internet of Vehicles (IoV) to enhance energy efficiency in data collection and waste traffic management [12]. In Italy, Industry 4.0 and IoT devices have been proposed to reduce operating costs and environmental impact by monitoring real-time waste levels in containers. Additionally, CO₂ emissions have been tracked from collection to waste separation using metaheuristic algorithms [13].

In Pakistan, a smart system was designed to manage waste in streams and sewers using a sensor kit that sends data to the cloud when thresholds are exceeded, notifying collectors in real time to enhance operational efficiency, quality of life, and investment opportunities [14]. In Singapore, an innovative framework addressed the challenges of cleaning complex public spaces like hospitals and shopping malls by employing RGB-D sensors, CCTV, DL algorithms, and route planning to target only dirty areas. The results demonstrated a 90% accuracy, with a 15% reduction in navigation times and 10% lower battery consumption compared to conventional methods [15].

In Malaysia, an autonomous river-cleaning robot was developed using an improved Yolo model, achieving 89% accuracy in identifying five types of trash under various conditions [11]. In Germany, a waste collection robot focused on human-robot interaction (HRI) was designed with a modified double-diamond frame, successfully implementing and testing hardware functions to meet ease-of-use requirements [16]. In China, a trash rescue system was constructed with autonomous navigation and debris detection powered by DL. The system, equipped with ROS, LiDAR, a robotic arm, and YoloV4, operates in real time at 45 FPS with 88% detection accuracy [17].

In India, an autonomous robotic arm was developed for solid waste segmentation using a customized LeNet model. This AI technique classified cardboard and plastic, achieving

high accuracy in object capture and categorization. System performance was evaluated using the metrics presented in [18]. In Bangladesh, a sensor-controlled prototype was designed to detect metallic and non-metallic waste, monitor water levels in containers, and send real-time notifications displaying the current container status via a mobile app (or SMS when offline) [19]. In the UK, an automated recycling system integrated ML and IoT to sort and separate recyclable materials. In this approach, the devices installed in waste bins monitor real-time waste generation, process images to calculate waste rates, and provide suggestions to enhance recycling productivity [20].

The SIRAMAND project in Romania developed an autonomous mobile robot to collect solid urban waste (e.g., bottles and cans) in parks and streets, which follows a programmable trajectory and detects, collects, and deposits waste in containers, with remote control available in case of failure. The robot achieved 100% efficiency in both laboratory and outdoor tests [21]. In India, researchers introduced a mobile robot based on transfer learning to identify and segregate electronic waste. With a 96% sorting accuracy, this robot reduces risks to unskilled labor, cuts costs by 20% over five years, and minimizes human intervention [22]. Another Indian project developed a rocker-bogie robot using YoloV4-tiny, a camera, an ultrasonic sensor, GPS, Raspberry Pi, and Arduino. YoloV4-tiny demonstrated similar accuracy to YoloV4, albeit with faster detection times and reliable performance on various terrains [23].

Researchers in Indonesia developed a robotic boat controlled via an Arduino micro-controller and a PS2 joystick connected to an Android smartphone. This system includes a relay to regulate electric current for the DC motor and a camera to display results. Communication between the joystick and Arduino was efficient at distances of up to 25 m without barriers and 15 m with barriers. The prototype is capable of removing small trash items such as bottle caps, gallon containers, and 200 g drinking glasses [24].

Researchers in Cuba developed an efficient, low-cost embedded system for controlling and managing an autonomous naval vessel. The system demonstrated proper planning and route compliance, with a margin of error of 2 mm at each checkpoint and smooth operations during starts, stops, and alarms [25]. In the US, efforts have focused on promoting ecological and environmental literacy. One initiative, the Smart Trash Junior Robot, is a mechatronic trash can equipped with vision to identify recyclable objects, which is designed to teach elementary school children about recycling [26]. Additionally, a water-cleaning robot was developed in the US, achieving a 90.6% sorting accuracy. The robot successfully performed aquatic navigation, gripper handling, color recognition, and waste sorting tasks [27].

In Indonesia, researchers integrated mobile and stationary robotic systems for waste collection, demonstrating efficiency in collecting egg and bottle waste with ultrasonic sensors and a navigation system [28]. In Mexico, a waste collection robot with a volume capacity of 100 to 180 cm³ was developed for beaches. Using Bluetooth, an ATMEGA microcontroller, and ultrasound sensors, this robot collects three solid objects in seven minutes and operates autonomously for 60 minutes [29]. In the US, soft robotics was applied to recycling through an electric robotic gripper

that detects paper, metal, and plastic while being puncture-resistant. The gripper achieves 85% accuracy in stationary setups and 63% in simulated recycling pipelines [30].

In India, researchers proposed using IoT within containers to improve waste management by enabling comprehensive communication with collection carts [31]. In Brazil, a study analyzed the participation of waste collectors in Caxias do Sul's waste management system through observations, interviews, and waste characterization using two types of trucks. 12 samples were categorized as organic, plastic, paper, cardboard, metal, glass, Tetra Pak, and other types of waste, providing a diagnostic for municipal management [32]. In the United Kingdom, researchers classified various types of waste, including urban, construction, industrial, agricultural, commercial, market, hotel, and restaurant waste [33].

Methodology

IoT-based applications integrate various technologies to develop cohesive systems. This section outlines the steps taken to create the mobile cleaning robot, as well as the criteria used for evaluation.

Block diagram

Fig. 1 shows the block diagram of the T5R-bot prototype designed for solid waste collection. The robot is powered by a lithium battery (a *power bank*) managed by a BMS and connected to an Arduino Nano and a Raspberry Pi 4, which runs the object recognition and waste detection model. Motion is controlled by signals sent to the H-bridge, driving the DC motors to move forward or backward or rotate. An ultrasonic sensor calculates the distance, guiding the gripper to pick up and deposit objects into the container.

Mechanical design

The following activities were considered for the mechanical design.

3D printing

3D printing is an additive manufacturing technology that has significantly advanced in recent years across various disciplines [34]. The versatility of 3D printing materials stems from the wide variety of printer models and types, as well as from innovations in material processes, making it an essential component of multi-process systems to support the development of new materials and product requirements [35]. Figs. 2 and 3 show the design created in FreeCAD, which allowed for a rapid and efficient creation of parts. The design was then exported to UltiMaker Cura to configure its properties before 3D printing (Fig. 4).

Rocker-bogie mechanism

The rocker-bogie mechanism was developed by NASA in 1988 to easily adapt to unknown and rugged terrain during Mars expeditions, allowing for relatively smooth movements on rough surfaces [23]. An interesting feature of this mechanism is the differential constraint, which was used in the Mars Science Laboratory (MSL) rover to maintain the symmetry of the mobility system during the launch, cruise,

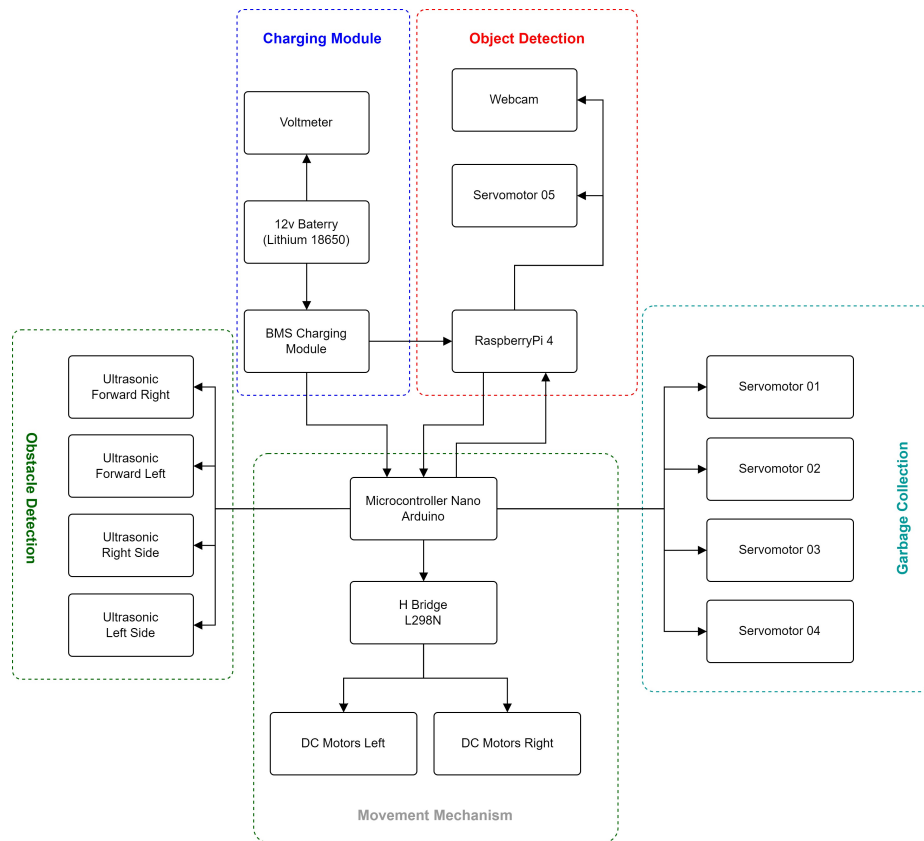


Figure 1. Block diagram

Source: Authors

entry, descent, and landing phases of the mission. This mechanism underwent almost three design cycles before being successfully completed [36]. To power our robot, six JGA25-370 DC gear motors were implemented, one for each leg of the mechanism. These motors are connected to an L298N driver, also known as an H-bridge, powered by 3000 mAh lithium batteries. A Raspberry Pi Model B computer supplies control signals to the H-bridge, allowing the wheels to move forward, backward, right, or left as needed.

Gripper

Handling objects can be dangerous for operators [37]. Gripping forceps, widely used in various research fields, provide reliability, precision, and safety. A study on automation in seedling transplantation achieved a 99% success rate [38]. Other research works highlight the development of ultra-compact and soft grippers capable of performing rapid, high-amplitude multidimensional maneuvers on microscale objects [39]. Various types of tweezers have been developed; among them, recent research introduced a compressible tweezer that adapts to the shape of objects, thereby improving grip efficiency [40].

The developed gripper features four degrees of freedom (DoF), as shown in Fig. 5. The model was 3D-printed, and its movement is powered by four servomotors installed at the joints. To ensure smooth operation, the maximum force required from these servomotors was calculated while considering several factors. The S3003 servomotor can lift 4.1 kg at a distance of 1 cm from its axis. Given that the effective distance from the first motor axis to the tip of the first link is 150 mm, and it operates at inclination angles

between 0° and 180°, the calculations using Eqs. (??) and (??) yield a required force of 0.615 kg.

$$\tau = F \cdot b \quad (1)$$

$$\tau = F \cdot r \cdot \sin \phi \quad (2)$$

Where:

τ : Torque or force of movement

b : Distance

F : Force

r : Radius

ϕ : Angle

Ultrasonic

Four ultrasonic sensors were installed: two at the front and one on each side of the robot. The front sensors can detect trash and stop the robot to allow the gripper to pick it up, while the side sensors are used to avoid obstacles.

Bucket

The design was created and 3D-printed, as shown in Fig. 6, with dimensions 380 x 320 x 75 mm. When the SSD MobileNetV2 algorithm running on the Raspberry Pi 4 Model B detects an object, the vehicle approaches it, picks it up with the gripper, and places it in the bucket. The recognition threshold was set to 90%.

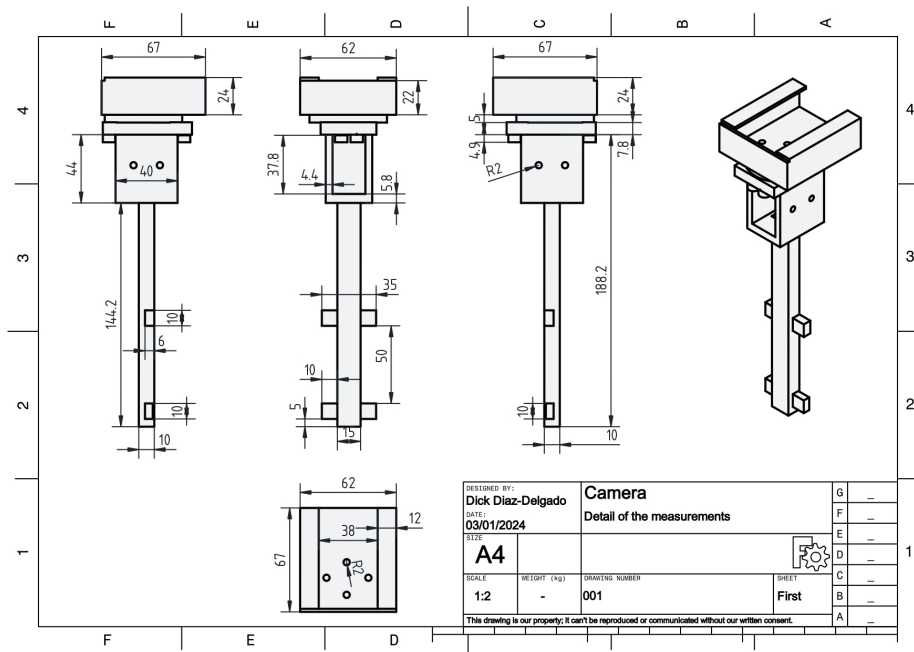


Figure 2. Tower measurements for the car camera

Source: Authors

Camera setup

The camera was mounted at the front of the robot using a 100 mm 3D-printed bracket, thereby enhancing the viewing angle and preventing obstructions to the pick-up gripper. Connected to an S3003 servo motor, the camera moves horizontally to detect and track solid waste. When debris is detected, it switches from search mode to tracking mode, maintaining the target in view.

Components list

The following components were used.

Raspberry Pi 4 Model B

This is one of the most basic and best-selling computers, featuring a Broadcom BCM2711 Quad-core Cortex-A72 1.8 GHz chip and SDRAM memory options of 1, 2, 4, or 8 GB depending on the model. It includes dual-band Wi-Fi (2.4 GHz and 5 GHz), Bluetooth 5.0, Gigabit Ethernet, two USB 3.0 ports, one USB 2.0 port, 40 GPIO pins, two micro-HDMI ports, and a Micro-SD slot for storage. It operates within a temperature range of 0 to 50 °C, with dimensions 94 x 70 x 26 mm and a weight of 66 g. Power over Ethernet (PoE) is supported but requires a separate HAT [41, 42, 43]. The SSD MobileNetV2 algorithm runs continuously via the webcam to recognize solid waste objects.

Arduino Nano

This compact and versatile micro-controller measures 18 x 45 mm, weighs 7 g, and is compatible with IoT components. It features 14 digital pins, eight analog pins, two reset pins, six power pins, and a mini-USB connection. Operating at 16 MHz with a power consumption of 19 mA, it is based on the ATmega328P. The power supply is 5 V, with an input range of 7 to 12 V, and it includes 1 or 2 KB of SRAM and 512 bytes or 1 KB of EEPROM depending on the model [19, 44, 45].

HC-SR04 ultrasonic sensor

This sensor measures distances using ultrasonic sound pulses and features four pins (Vcc, Trig, Echo, GND). It operates with a 5 V power supply, has a quiescent current of less than 2 mA, and measures 45 x 20 mm [46].

Servomotor S3003 (REES52)

This servomotor has a motion range of 0 to 360° and requires a pulse width of 0.3 to 2.3 ms to achieve its maximum angle. It has a load capacity of 4.1 kg [47, 48].

Dahua UZ3

This device has the following characteristics: a capacity of 2 megapixels, video resolution of 1920 x 1080 pixels (Full HD 1080p), and dimensions 97 x 57 x 33 mm. It captures the images to be analyzed by the Raspberry Pi.

Power bank

The power bank consists of nine 18650 lithium-ion batteries, each with a nominal voltage of 3.7 V, a maximum voltage of 4.2 V, and a minimum voltage of 2.9 V. Each battery transfers 3 A per hour and was connected in series and parallel to achieve a required voltage of 12 V and a current of 9 A. The system can operate at 3 Ah (ampere-hours, representing the battery capacity supplied to the system) for one hour. The batteries measure 18 x 65 mm and are cylindrical in shape [49].

BMS-3S charging module

This intelligent component enables the advanced control and management of a lithium battery storage system with a maximum output and charging power of 252 W, following the wiring diagram 0 V/4.2 V/8.4 V/12.6 V. Its main function is charge and discharge control, in addition to the collection,

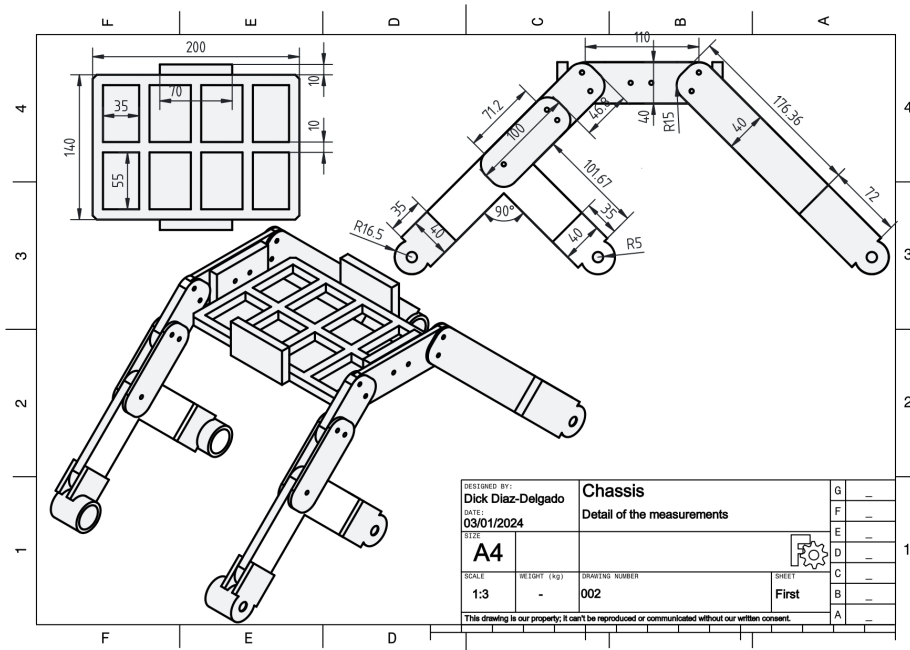


Figure 3. Car chassis measurements

Source: Authors

processing, and storage of real-time information on battery operation [50].

Mini digital voltmeter

This is a high-quality, durable component with a long service life. It features a 0.025% resolution core, ensuring a sufficient margin and high precision. It offers a direct power supply for testing, a low starting voltage, and a reverse connection protection function to prevent damage. The measurement range is between 2.4 and 32 V, and its dimensions are 30 x 11.7 x 9.2 mm [51].

L298N H-bridge

This module controls the direction and speed of two motors, operating at voltages from 3 to 35 V with a current capacity of 2 A. It includes a regulator that provides a 5 V output when powered within a range of 5 to 12 V. The module measures 43 x 43 x 27 mm, weighs 30 g, and delivers a power output of 25 W [52].

JGA25-370 motors, DC 12 V

These devices convert electrical energy into mechanical energy, or vice versa. They are widely used in nearly all human activities [53] and can function as generators or motors depending on the application [54, 55]. In the robot, the motors allow moving forward, backward, left, and right.

Electronic design

Fritzing was used to create the basic electronic design diagram by incorporating library components (Fig. 7).

Object detection model

Below are the implementation steps followed for the computer vision system using the MobileNet architecture and the SSD model.

Preparation and pre-processing data

To collect the dataset, photographs were taken in public places, such as parks, squares, sports fields, markets, roads, towns, and river slopes, among others. Additionally, free images were sourced from the Internet and the TACO (trash annotations in context) image collection [56], resulting in a total of 2890 images representing the four types of solid waste considered in this research. Common objects included bottles, cans, boxes, and cups captured against various backgrounds such as grass, sidewalks, and debris at different times of the day, with resolutions ranging from 3000 to 4500 pixels. These images were used to train the waste collection robot, with resized 300 x 300 pixel images to accelerate training and reduce processor load.

The images were manually annotated using Labellmg and MakeSense.AI, creating bounding boxes to classify waste into four categories: box, bottle, glass, or can. The annotations, including XY coordinates, were saved in the YOLO (.txt) and PASCAL VOC (.xml) formats. The dataset was divided into 80% for training (2312 images) and 20% for testing (578 images), totaling 2890 images.

Fig. 8 illustrates the working principle of the object detection system for solid waste using SSD MobileNetV2 and its processes. First, the dataset is collected for training and testing. Next, image pre-processing is performed, which includes tagging the images, converting the XML (eXtensible Markup Language) dataset to the CSV (comma-separated values) format, and then converting the CSV file to the TFRecord format. The final pre-processing step involves creating a tag map. In the image processing phase, the pre-trained MobileNetV2 SSD model is selected, and the *pipeline.config* file is configured. The model is then trained using a convolutional neural network (CNN) architecture, resulting in a classification, detection, and prediction model. The non-maximum suppression (NMS) process is applied to generate a model with a minimum intersection over union

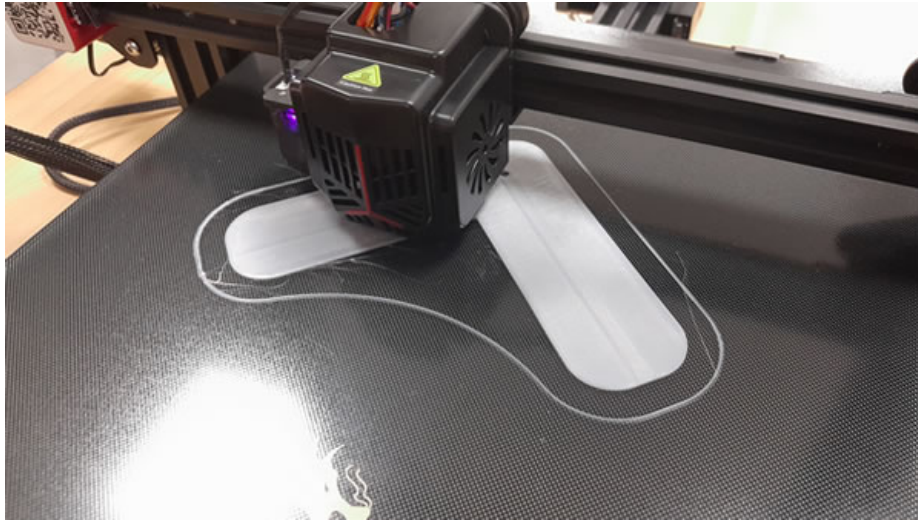


Figure 4. Printing parts with 3D technology

Source: Authors

(IoU) of 0.5. Once the model has been trained, it produces outputs that are converted into estimates of solid waste.

The main method calculates bounding boxes and object classes from images using the SSD model with MobileNet. Designed for fast detection and optimized for mobile devices like the Raspberry Pi, it efficiently detects objects and provides information about their position and class.

Structure model

Object detection is a highly competitive field dominated by models such as SSD (MobileNet), RCNN (Mask and Faster), and YOLO in various versions. While these models are accurate, they can be computationally expensive and require significant inference time. SSD, as its name suggests, performs multiple object detections in a single shot, unlike other two-stage models. This makes it much faster without compromising accuracy. SSD operates in two main steps: feature map extraction and object detection using convolutional filters [57]. The primary SSD models are SSD512, designed for high-resolution images (500×500), and SSD300, optimized for faster processing and used with low-resolution images (300×300). The SSD rendering architecture is illustrated in Fig. 9, where each feature layer generates a fixed set of detection predictions using convolutional filters.

MultiBox detector

The specific location information corresponds to the various outputs of the detector, which vary depending on the position, proportion, and size of objects. In SSD, the MultiBox loss function is employed, which consists of the weighted sum of the localization loss (L_{loc}) and the confidence loss (L_{conf}), as described in Eq. (3). In this context, N represents the number of matching predefined boxes. The localization loss refers to $smooth_{L1}$, calculated between the predicted (l) and actual box parameters (g), as shown in Eq. (4). Additionally, SSD adjusts the center coordinates (cx , cy), height (h), and width (w) of the detection boxes. The confidence loss, detailed in Eq. (7), corresponds to the softmax loss applied to the confidence scores (c) of various categories.

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (3)$$

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k smooth_{L1}(l_i^m - g_i^m) \quad (4)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^w \quad \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})/d_i^h \quad (5)$$

$$\hat{g}_j^w = \log(g_j^w/d_i^w) \quad \hat{g}_j^h = \log(g_j^h/d_i^h) \quad (6)$$

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad (7)$$

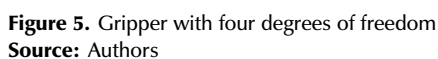
MobileNet

A key function of the backbone is to extract detailed features from the input image and pass them to the subsequent SSD layers. The MobileNet network configuration employs depthwise separable convolutions, as illustrated in Fig. 10, significantly reducing the network's size and yielding a lightweight model with lower computational complexity. Additionally, two hyperparameters (width and resolution) control the input/output of the convolutional layer and the resolution, providing better control over the trade-off between computational cost and network accuracy based on end-user requirements [57].

The integration of this framework with SSD enhances the performance of real-time applications using low-capacity or low-end devices such as mobile phones. After each layer in the main network structure, batch normalization and a rectified linear unit (ReLU) function are applied, except for the last fully connected layer. The final layer bypasses a nonlinear function, passing the data directly to a softmax layer for classification [58].

Implementation and operation

The implementation and operational procedures of the autonomous robot for solid waste collection are described below, along with the challenges encountered and potential solutions to enhance our proposal's functionality.



The Raspberry Pi 4 Model B was chosen to run the DL model. The weights of the MobileNet FPNLite SSD model were exported as a .ZIP file due to their compact size. Subsequently, the model was converted to TensorFlow Lite for deployment on the Raspberry Pi.

Training requires a large number of suitable, manually annotated images, which is labor-intensive and increases the training time as the dataset grows. For robot production, greater computational power is needed, which can be obtained through cloud services such as AWS, Azure, or IBM, leveraging multiple GPUs. Companies must pay for access to these services on a pay-as-you-go basis. Another option is the Coral USB Accelerator, which enhances processing and reduces object detection time.

The robot searches for solid waste using its camera while navigating obstacles with ultrasonic side and front sensors, pausing for 1-2 s to assess the surroundings. Its design allows it to overcome obstacles up to twice the diameter of its wheels. When debris is detected, the robot aligns itself and calculates the object's center along the X-axis of the camera (Fig. 11). It then aligns the object's center with the gripper and slowly approaches to pick up the debris. After selecting the object closest to the axis for collection, the robot continues scanning the area to complete its task.

Given that the average volume of objects is $720\,000\text{ mm}^3$, it is estimated that the repository can store up to 12 objects. Consequently, this number was established as the maximum limit to ensure a proper distribution within the container and prevent exceeding the weight limit of 0.615 kg . To this effect, the robot keeps track of the number of objects it collects and

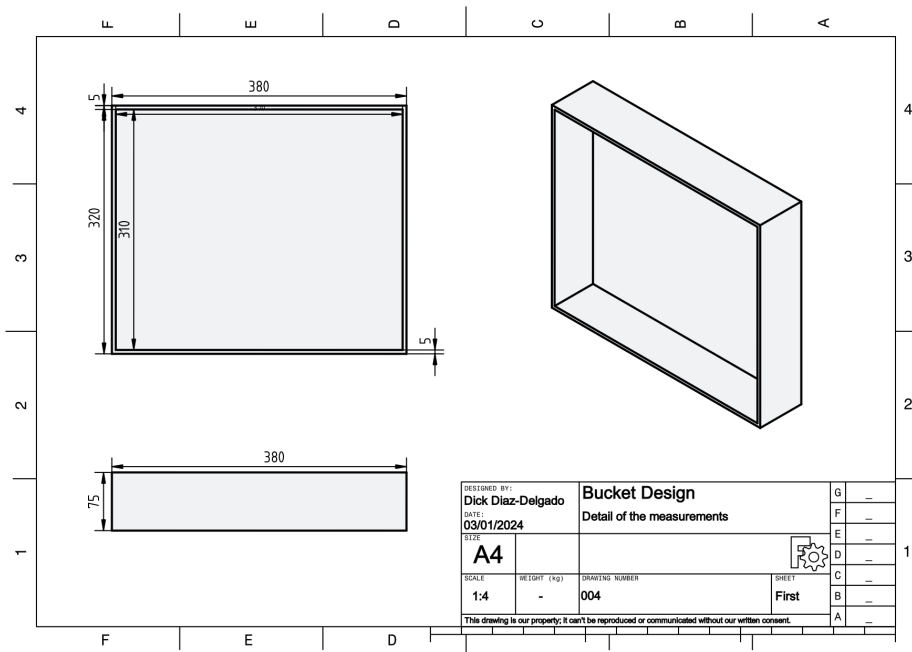


Figure 6. Perspective view of the design

Source: Authors

stores in the bucket, which allows it to determine when its task is complete.

Results

The decrease in losses (Figs. 12 and 13) suggests that, as the training period grows longer, the reliability of the model's predictions improves. The model exhibits a sharp decrease in losses during the initial training phase, followed by a gradual and continuous decline. The final loss achieved after 50 000 training steps was 0.1085.

Object detection model

While training was performed, the mAP (mean average precision) showed an increasing trend with occasional small decreases. The SSD MobileNetV2 model demonstrated a significant advantage, achieving a reduction in losses (*Lost*) during the initial training phase. A 80-20 split was chosen given its recognition as a standard in the field, ensuring a representative evaluation and allowing for direct comparisons against previous studies. This ratio provides an optimal balance between the size of the training set and the ability to effectively measure performance. The model's inference times are very fast and exhibit high accuracy, with mAP results from the test run reaching 94.84%. This is considered an ideal result, as the IoU threshold exceeds the expected 50% ($\text{IoU} \geq 0.5$) [23] (Table II).

Table II. Performance metrics

Nº	Class	0.5 IoU	0.95 IoU	mAP @ 0.5:0.95
1	Box	100%	52%	95.2%
2	Bottle	100%	80%	98%
3	Can	100%	80.56%	98.06%
4	Glass	100%	100%	100%
Total		100%	78.14%	97.81%

Accuracy test

In evaluating the model, the IoU and mAP metrics were key. IoU is a measure used to compute the accuracy of an object detection model on a specific dataset. Also known as the *Jaccard index*, it is calculated using Eq. (8).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

J, A, and B represent the Jaccard distance, the first set, and the second set, respectively. The IoU measures object detection performance by comparing the ground truth bounding box with the one generated by the model. Eq. (9) was used to calculate the IoU value.

$$\text{IoU} = \frac{(\text{Area of Overlap})}{(\text{Area of Union})} \quad (9)$$

The confusion matrix presented in Table III evaluates the performance of the object detection model in classifying solid waste into the following categories: box, bottle, can, and cup. Each column and row of the table represents the number of correctly identified cases (true positives, TP), misclassified cases (false positives, FPs), and undetected cases (false negatives, FNs). It also includes true negatives (TNs), although, in this case, there are always zero TNs because the problem does not include explicit negative examples.

For example, for the box category, the model correctly detected 48 boxes (TP), reported two errors by classifying other objects as boxes (FP), and left no boxes undetected (FN). The same analysis was applied to the other categories with similar results, reflecting an excellent model performance with accuracy and recall rates close to 100%. This analysis confirms the system's ability to reliably detect solid waste in different environments.

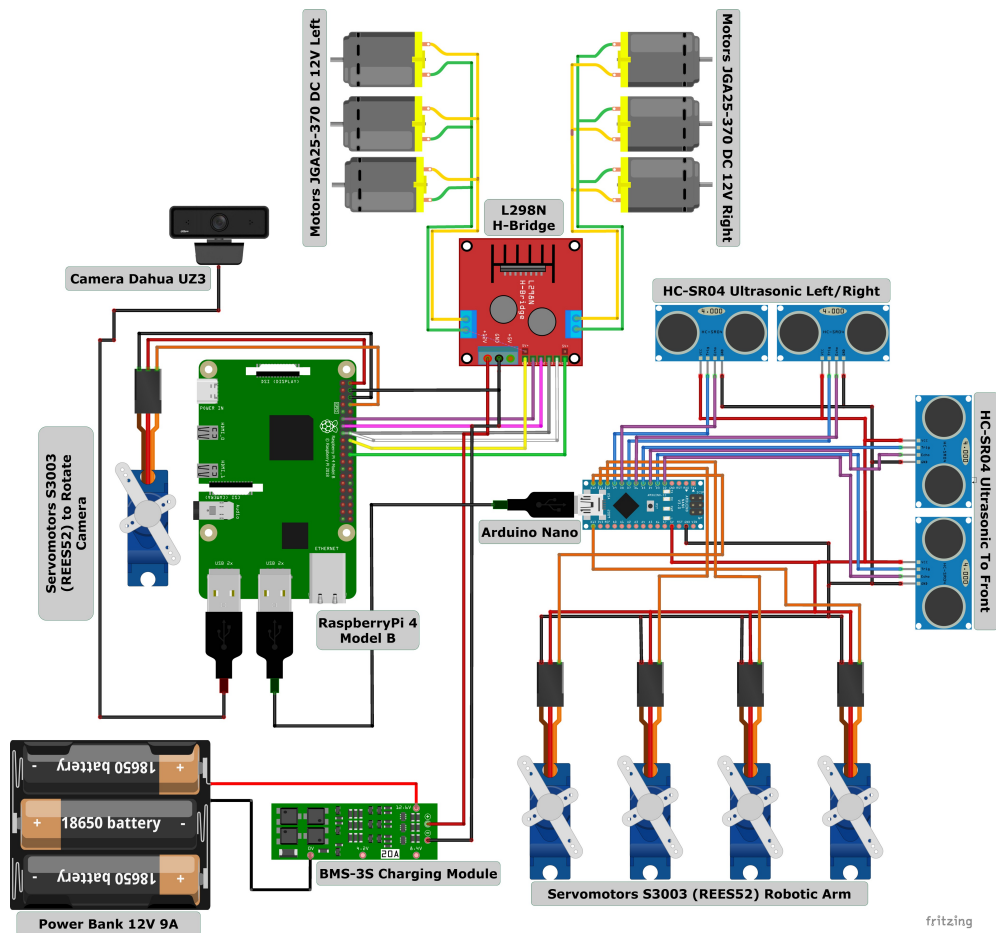


Figure 7. Circuit diagram

Source: Authors

The evaluation metrics used for the calculation of performance parameters were accuracy (10), precision (11), recall (12), and the F1-score (13).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

To calculate the average precision (AP) and the mAP, Eqs. (14) and (15) were used.

$$AP = \int_0^1 P(R) dR \quad (14)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (15)$$

Table IV shows the precision, recall, and F1-score values calculated to evaluate the model's effectiveness and quality.

Table III. Confusion matrix

Nº	Type	TP	FN	FP	TN
1	Box	661	0	28	0
2	Bottle	674	0	14	0
3	Can	812	0	14	0
4	Glass	688	0	0	0

Source: Authors

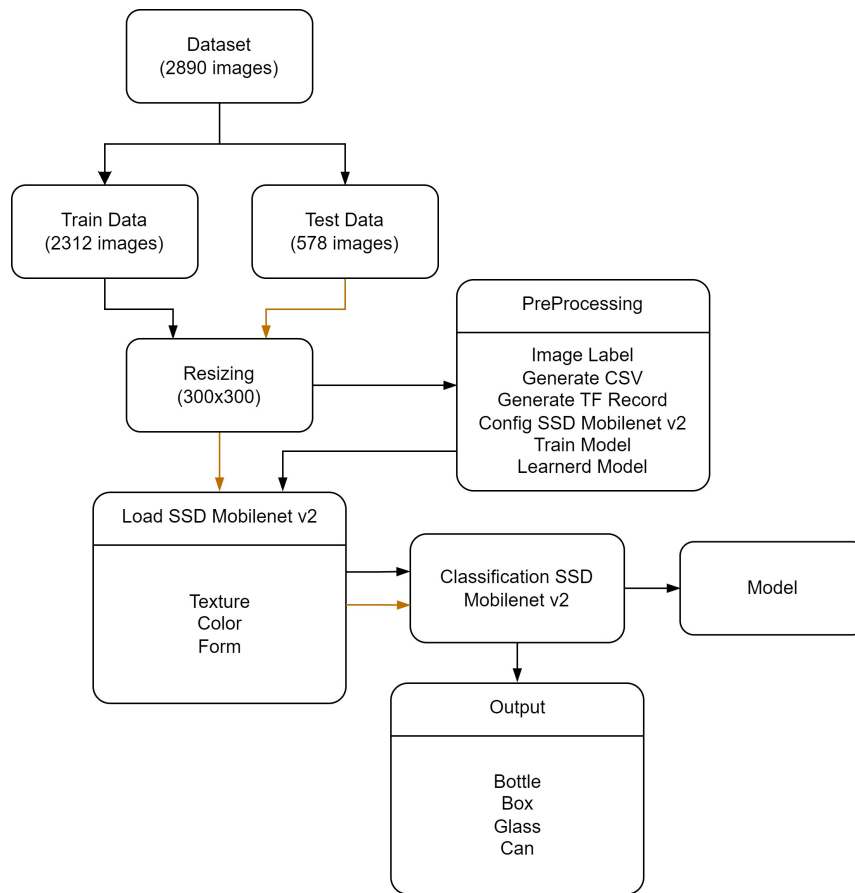
Table IV. Accuracy test results

Nº	Type	Precision	Recall	F1-Score
1	Box	98%	100%	99%
2	Bottle	96%	100%	98%
3	Can	98%	100%	99%
4	Glass	100%	100%	100%

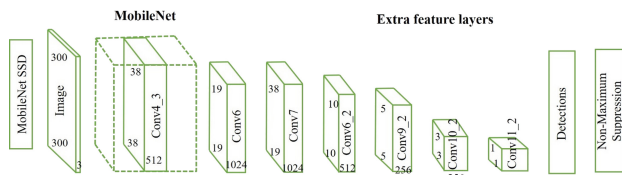
Source: Authors

Table V presents the image detection results, including the overall values for accuracy, recall, and the F1-score. The MobilenetV2 SSD algorithm achieved a 94% accuracy in solid debris detection, enhancing detection applications by expanding the dataset.

Fig. 14 shows the object detection of the SSD MobileNetV2 model, which was integrated into the Raspberry Pi through the webcam. The movement and search algorithm allowed the robot to operate in an endless loop to locate and collect

**Figure 8.** General architecture

Source: Authors

**Figure 9.** MobileNetV2 model [57].

Source: Authors

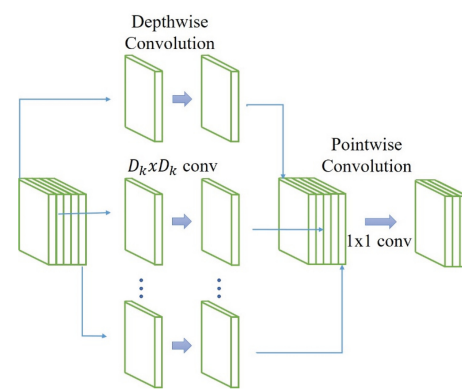
Table V. Percentage results

Source: Authors		
Precision	Re-call	F1-Score
98%	100%	99%

waste, influencing detection by up to 20%. The robot also featured an emergency switch to stop its operation in case of unforeseen events.

Regarding the process of grabbing and placing objects on the tray, the robot takes 7 s per object once it focuses on the target (Fig. 15).

As reported in the specialized literature, [11] achieved a mAP of 89% by training the YoloV4 model. Additionally, [23] achieved a value of 97.1% in the same model, 83.3% in the Mask-RCNN model, and 95.2% in another instance of Mask-RCNN. For comparison, SSD MobileNetV2 achieved a 98% precision and a 97.81% mAP (Table VI).

**Figure 10.** Depthwise convolutional layer architecture [57]

Source: Authors

Discussion

The implementation of the T5R robot using the MobileNetV2 SSD architecture for solid waste detection and collection yielded promising results, achieving an accuracy of 98% and a mAP of 97.81%. These outcomes surpass those reported in previous studies using models such as YOLOv4 and Mask-RCNN, which achieved accuracies of 89 and 83.3%, respectively, as cited by [11, 23]. This performance underscores the model's effectiveness in hardware-constrained environments (e.g., deployment on a Raspberry Pi) while maintaining low computational costs without sacrificing accuracy.

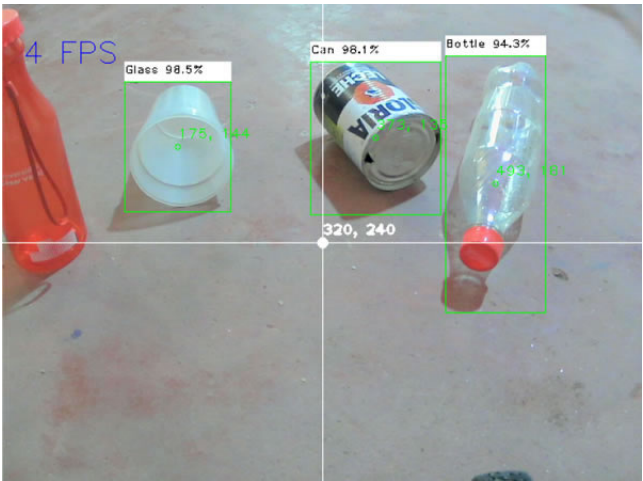


Figure 11. Object detection
Source: Authors

Table VI. Model comparison

Source: Authors

Research	Algorithm	mAP
[11]	YoloV4	89%
	YoloV4	97.1%
[23]	Mask-RCNN	83.3%
	YoloV4-tiny	95.2%
Ours	MobileNetV2	97.81%

The results obtained in controlled environments are promising. However, it is essential to evaluate the robot’s effectiveness in real-world and diverse scenarios. Factors such as variations in waste type, climatic conditions, and the distinctive characteristics of urban environments can significantly affect system performance. These variables may pose challenges that are not evident in controlled settings, emphasizing the importance of additional testing. Conducting pilot tests in various locations would not only help identify areas for improvement; it would also allow adapting the system to specific contexts. This approach would ensure that the robot meets the standards necessary for large-scale deployment while maintaining efficiency and functionality under practical conditions. Such evaluations are critical to validating the system’s feasibility and its effective contribution to waste management in complex urban environments.

Conclusions

In this work, a dataset of 2890 images collected under various scenarios and conditions was consolidated for the training and testing of solid waste detection. The T5R robot was built from scratch using 3D printing technology and ABS, featuring a suspension system inspired by the rocker-bogie automobile mechanism, which allows for safe movement in work areas. The robot employed computer vision and a versatile clamp to autonomously collect waste on different terrains. Its MobileNetV2 SSD architecture operated in TFLite for efficient execution on the Raspberry Pi, identifying solid waste through an advanced approach. This facilitated fast convergence, reduced training times,

improved accuracy to 98%, and allowed reaching a detection threshold of 90% and a mAP of 97.81%.

The T5R robot is capable of storing and detecting up to 12 objects, including bottles, boxes, cans, and glasses, in just 7 s, highlighting its effectiveness in contaminated environments. The innovation represented by the T5R underscores the feasibility of collection robots as an effective solution to the challenges of waste management in urban environments.

The development of the T5R robot represents a significant advancement in sustainable solid waste management. By automating waste collection, it reduces dependence on human labor for repetitive and potentially hazardous tasks, thereby minimizing the health risks associated with such activities. Moreover, the deployment of this technology in urban environments has the potential to make a positive impact by lowering contamination levels and increasing recycling rates. These advancements could play a pivotal role in facilitating the transition to a circular economy, fostering more sustainable and responsible waste management practices.

Recommendations and future developments

Currently, the robot only picks up objects, but, in the future, waste segregation and categorization capabilities will be added. We also plan to integrate LiDAR technology in order to improve navigation. Additionally, the DL model will be optimized to avoid overfitting and detect moving obstacles, such as people and vehicles. Future versions of the robot will use reinforcement learning to acquire knowledge of the environment and make more intuitive decisions. It would also be advisable to incorporate techniques such as cross-validation or other partitioning schemes to verify and strengthen the generalization of the results.

Additionally, the robot’s automatic classification capabilities could be enhanced through the use of deep neural networks specialized in waste segmentation, enabling a more efficient and precise handling of various waste types.

Acknowledgments

We would like to thank the Faculty of Engineering and Architecture of Universidad César Vallejo (Tarapoto, Peru) for their support.

Research funding

This research was supported by the Teaching Research Fund of Universidad César Vallejo, with project number P-2022-185.

Conflicts of interest

The authors declare no conflict of interest.

Data Availability

All data related to this research study can be requested from the corresponding author.

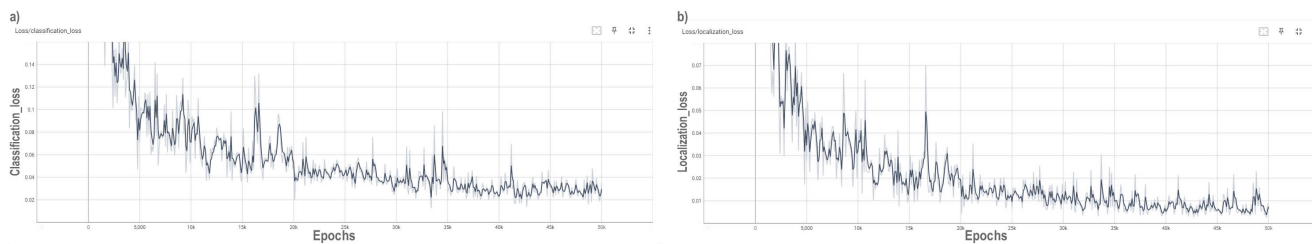


Figure 12. Model results regarding a) classification loss and b) localization loss

Source: Authors

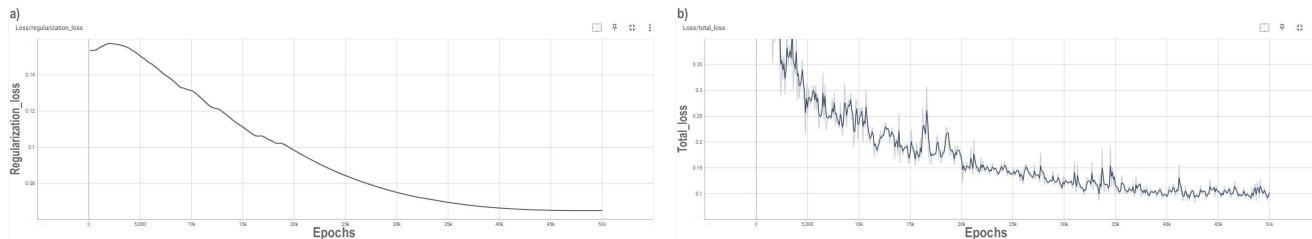


Figure 13. Model results regarding a) regularization loss and b) total loss

Source: Authors

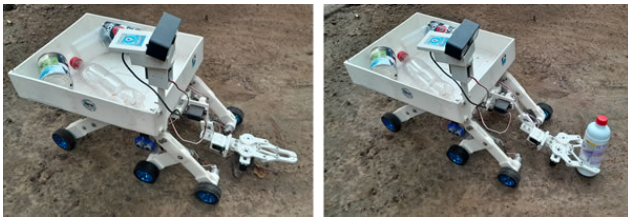


Figure 15. Solid waste collection

Source: Authors

writing (review and editing). *Alexander Inga-Alva:* conceptualization, methodology, investigation, validation, writing (original draft), writing (review and editing).

References

- [1] X. Li, D. Shi, Y. Li, and X. Zhen, "Impact of carbon regulations on the supply chain with carbon reduction effort," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 6, pp. 1218–1227, Jun. 2019. <https://doi.org/10.1109/TSMC.2017.2741670>
- [2] V. Meiler, J. Pfeiffer, L. Bifano, C. Kandlbinder Paret, and G. Fischerauer, "Approaches to detect microplastics in water using electrical impedance measurements and support vector machines," *IEEE Sens. J.*, vol. 23, no. 5, pp. 4863–4872, Mar. 2023. <https://doi.org/10.1109/JSEN.2023.3236375>
- [3] T. Mato and S. Noguchi, "Microplastic collection with ultra-high magnetic field magnet by magnetic separation," *IEEE Trans. Appl. Supercond.*, vol. 32, no. 4, pp. 1–5, 2022. <https://doi.org/10.1109/TASC.2021.3135796>
- [4] C. Schmidt, T. Krauth and S. Wagner, "Export of plastic debris by rivers into the sea," *Environ. Sci. Technol.*, vol. 51, no. 21, pp. 12246–12253, 2017. <https://doi.org/10.1021/acs.est.7b02368>
- [5] G. Castañeta, A. F. Gutiérrez, F. Nacaratte, and C. A. Manzano, "Microplastics: A contaminant that grows in all environmental areas, its characteristics and possible risks to public health from exposure," *es, Rev. Boliv. Química*, vol. 37, pp. 160–175, Dec. 2020, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=9001749>
- [6] G. Vázquez-Rodríguez, "Los microplásticos textiles (o la increíble historia de cómo tu suéter termina en el salero)," *Ciencia*, vol. 70, no. 1, pp. 56–63, 2019. https://www.revistaciencia.amc.edu.mx/images/revista/70_1/PDF/Microplasticos.pdf

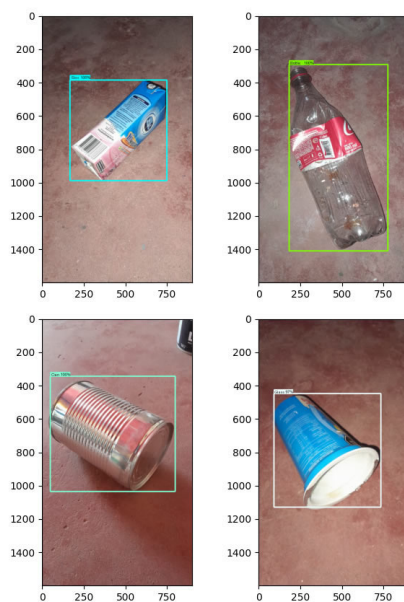


Figure 14. MobileNetV2 SSD results

Source: Authors

CRedit author statement

Dick Díaz-Delgado: conceptualization, methodology, software, formal analysis, investigation, data curation,

- [7] D. Aldana Aranda, E. Díaz, and V. Castillo Escalante, "El caribe y su contaminación por microplásticos," *Ciencia*, vol. 73, no. 2, pp. 8–13, 2022. https://revistaciencia.amc.edu.mx/images/revista/73_2/PDF/03_73_2_1429_Microplasticos_Caribe.pdf
- [8] P. Taboada-Gonzalez, Q. Aguilar-Virgen, and S. E. Cruz-Sotelo, "Manejo y potencial de recuperación de residuos sólidos en una comunidad rural de México," *Rev. Int. Contam. Ambient.*, vol. 29, pp. 43–48, 2013, [Online]. Available: <https://www.revistascca.unam.mx/rica/index.php/rica/article/view/43517>
- [9] S. A. Mahfoodh, M. Aouad, I. Maki, R. Hdeib, and H. Mohamed, "Towards sustainable municipal solid waste management: Practical solutions to contain underground toxic gases from landfills," in *Proc. ASU Int. Conf. Emerg. Technol. Sustain. Intell. Syst. (ICETIS)*, 2024, pp. 444–450. <https://doi.org/10.1109/ICETIS61505.2024.10459548>
- [10] J. Bartra Gómez and J. M. Delgado Bardales, "Gestión de residuos sólidos urbanos y su impacto medioambiental," *Cienc. Latina Rev. Cient. Multidiscip.*, vol. 4, no. 2, pp. 993–1008, Dec. 2020. https://doi.org/10.37811/cl_rcm.v4i2.135
- [11] N. A. Zailan, M. M. Azizan, K. Hasikin, A. S. Mohd Khairuddin, and U. Khairuddin, "An automated solid waste detection using the optimized yolo model for riverine management," *Front. Public Health*, vol. 10, 2022. <https://doi.org/10.3389/fpubh.2022.907280>
- [12] G. K. Ijamaru, L.-M. Ang, and K. P. Seng, "Swarm intelligence internet of vehicles approaches for opportunistic data collection and traffic engineering in smartcity waste management," *Sensors*, vol. 23, no. 5, art. 2860, 2023. <https://doi.org/10.3390/s23052860>
- [13] G. Rahmanifar, M. Mohammadi, A. Sherafat, M. Hajiaghahi-Keshteli, G. Fusco, and C. Colombaroni, "Heuristic approaches to address vehicle routing problem in the iot-based waste management system," *Expert Syst. Appl.*, vol. 220, art. 119708, 2023. <https://doi.org/10.1016/j.eswa.2023.119708>
- [14] J. U. Rahman, A. Khan, J. I. Bangash, A. Khan, D. A. Ramli, and S. Khan, "An efficient smart streamlet management system using internet of thing," in *Proc. 26th Int. Conf. Knowledge-Based Intell. Inf. Eng. Syst. (KES2022)*, *Procedia Comput. Sci.*, vol. 207, pp. 1743–1753, 2022. <https://doi.org/10.1016/j.procs.2022.09.232>
- [15] B. Ramalingam, A. V. Le, Z. Lin, Z. Weng, R. E. Mohan, and S. Pookkuttath, "Optimal selective floor cleaning using deep learning algorithms and reconfigurable robot htetro," *Sci. Rep.*, vol. 12, art. 15938, 2022. <https://doi.org/10.1038/s41598-022-19249-7>
- [16] A. Vazhapilli Sureshbabu, N. M. Martins Pacheco, L. I. Duran Noy, and M. Zimmermann, "Design of an autonomous trash-picking service robot focussed on human-robot interaction," *Proc. Des. Soc.*, vol. 2, pp. 2523–2532, 2022. <https://doi.org/10.1017/pds.2022.255>
- [17] W. Yang, J. Che, L. Zhang, and M. Ma, "Research of garbage salvage system based on deep learning," in *Proc. Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS)*, vol. 12260, SPIE, 2022, art. 1226014. <https://doi.org/10.1117/12.2637385>
- [18] P. R. Kshirsagar et al., "Artificial intelligence-based robotic technique for reusable waste materials," *Comput. Intell. Neurosci.*, vol. 2022, art. 2073482, 2022. <https://doi.org/10.1155/2022/2073482>
- [19] M. A. Rahman, F. Hasan, I. Akter, U. A. J. Sutapa, M. R. Shahriar, and K. M. S. Ibne Sayed, "Smart trash distribution, and recycling processes using iot sensing & mobile application," in *Proc. 2nd Int. Conf. Comput. Advancements (ICCA)*, 2022, pp. 227–232. <https://doi.org/10.1145/3542954.3542988>
- [20] X. Chen, "Machine learning approach for a circular economy with waste recycling in smart cities," *Energy Rep.*, vol. 8, pp. 3127–3140, 2022. <https://doi.org/10.1016/j.egyr.2022.01.193>
- [21] A. Stan, M. Mărgăritescu, A. M. E. Rolea, A. C. Dinu, V.-M. Zafiu, and D. M. Cotorobai, "Construction and control of an autonomous mobile robot for urban waste collection," *Int. J. Mechatronics Appl. Mech.*, no. 10, pp. 24–31, 2021.
- [22] A. S. Madhav, R. Rajaraman, S. Harini, and C. C. Kiliroor, "Application of artificial intelligence to enhance collection of e-waste: A potential solution for household weee collection and segregation in India," *Waste Manage. Res.*, vol. 40, no. 7, pp. 1047–1053, 2022. <https://doi.org/10.1177/0734242X211052846>
- [23] M. Kulshreshtha, S. S. Chandra, P. Randhawa, G. Tsaramirsis, A. Khadidos, and A. O. Khadidos, "Oatcr: Outdoor autonomous trash-collecting robot design using yolov4-tiny," *Electronics*, vol. 10, no. 18, art. 2292, 2021. <https://doi.org/10.3390/electronics10182292>
- [24] Y. Irawan, H. Fonda, Yulisman, and Mardeni, "Garbage collecting ship robot using arduino uno microcontroller based on android smartphone," *Int. J. Eng. Trends Technol.*, vol. 69, no. 6, pp. 25–30, 2021. <https://doi.org/10.14445/22315381/IJETT69I6P204>
- [25] A. Sánchez Moya and B. Rodríguez Rueda, "Desarrollo de plataforma autopiloto para robot subacuático," *Ing. Electron., Autom. Comun.*, vol. 42, pp. 89–104, Apr. 2021. <https://doi.org/10.14445/22315381/IJETT69I6P204>
- [26] M. Arnett, Z. Luo, P. K. Paladugula, I. S. Cardenas, and J.-H. Kim, "Robots teaching recycling: Towards improving environmental literacy of children," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Cambridge, U.K., 2020, pp. 615–616. <https://doi.org/10.1145/3371382.3379462>
- [27] X. Dong, "Research and design of marine trash classification robot based on color recognition," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 514, no. 3, art. 032043, May 2020. <https://doi.org/10.1088/1755-1315/514/3/032043>
- [28] A. P. P. Prasetyo, Rendyansyah, S. D. Siswanti, S. Nurmaini, and Abdurahman, "Garbage collector robot (gacobot) design for dry waste distribution," *J. Phys. Conf. Ser.*, vol. 1500, no. 1, art. 012103, Apr. 2020. <https://doi.org/10.1088/1742-6596/1500/1/012103>

- [29] G. A. García-Rodríguez, B. A. Aguilar-Figueroa, P. A. Lucho, R. M. Woo-García, and F. López-Huerta, "Diseño e implementación de un robot recolector de residuos sólidos en playas," *J. CIM Rev. Digit.*, vol. 8, no. Num.1, pp. 1220–1227, Oct. 2020. <https://doi.org/10.5281/zenodo.6527784>
- [30] L. Chin, J. Lipton, M. C. Yuen, R. Kramer-Bottiglio, and D. Rus, "Automated recycling separation enabled by soft robotic material classification," in *Proc. 2nd IEEE Int. Conf. Soft Robot. (RoboSoft)*, 2019, pp. 102–107. <https://doi.org/10.1109/ROBOSOFT.2019.8722747>
- [31] H. N. Saha, S. Auddy, S. Pal, S. Kumar, S. Pandey, R. Singh, A. K., Singh, S. Banerjee, D. Ghosh, and S. Saha, "Waste management using internet of things (iot)," in *Proc. 8th Annu. Ind. Autom. Electromech. Eng. Conf. (IEMECON)*, 2017, pp. 359–363. <https://doi.org/10.1109/IEMECON.2017.8079623>
- [32] M. D. M. Poletto, P. R. Schneider, and A. J. Vania E. Zattera, "Urban solid waste management in Caxias do Sul/Brazil: Practices and challenges," *J. Urban Environ. Eng.*, vol. 10, no. 1, pp. 50–56, 2016. <https://doi.org/10.4090/juee.2016.v10n1.50-56>
- [33] E. Amasuomo and J. Baird, "The concept of waste and waste management," *J. Manage. Sustainability*, vol. 6, no. 4, pp. 88–96, 2016. <https://doi.org/10.5539/jms.v6n4p88>
- [34] P. Andrés-Cano, J. A. Calvo-Haro, F. Fillat-Gomà, I. Andrés-Cano, and R. Perez-Mananes, "Papel del cirujano ortopédico y traumatólogo en la impresión 3d: Aplicaciones actuales y aspectos legales para una medicina personalizada," *Rev. Esp. Cir. Ortop. Traumatol.*, vol. 65, pp. 138–151, 2021. <https://doi.org/10.1016/j.recot.2020.06.014>
- [35] J. -Y. Lee, J. An, and C. Kai Chua, "Fundamentals and applications of 3d printing for novel materials," *Appl. Mater. Today*, vol. 7, pp. 120–133, 2017. <https://doi.org/10.1016/j.apmt.2017.02.004>
- [36] E. Jordan, "Mars science laboratory differential restraint: The devil is in the details," in *Proc. 41st Aerosp. Mechanisms Symp.*, pp. 263–277. [Online]. Available: <https://ntrs.nasa.gov/citations/20130003152>
- [37] J. W. Ugalde Vicuña, "Diseño de una pinza para el agarre de objetos," *E-IDEA J. Eng. Sci.*, vol. 1, no. 2, pp. 54–68, Nov. 2020. <https://revista.estudioidea.org/ojs/index.php/esci/article/view/67>
- [38] A. Khadatkhar, A. P. Pandirwar, and V. Paradkar, "Design, development and application of a compact robotic transplanter with automatic seedling picking mechanism for plug-type seedlings," *Sci. Rep.*, vol. 13, no. 1883, pp. 2045–2322, 2023. <https://doi.org/10.1038/s41598-023-28760-4>
- [39] J. Yan, Y. Zhang, Z. Liu, J. Wang, J. Xu, and L. Yu, "Ultracompact single-nanowire-morphed grippers driven by vectorial lorentz forces for dexterous robotic manipulations," *Nat. Commun.*, vol. 14, no. 3786, pp. 2041–1723, 2023. <https://doi.org/10.1038/s41467-023-39524-z>
- [40] K. C. Bingham, M. Hessler, S. Lama, and T. Deemyad, "Design and implementation of a compliant gripper for form closure of diverse objects," *Appl. Sci.*, vol. 13, no. 17, pp. 2076–3417, 2023. <https://doi.org/10.3390/app13179677>
- [41] G. Lambropoulos, S. Mitropoulos, C. Douligeris, and L. Maglaras, "Implementing virtualization on single-board computers: A case study on edge computing," *Computers*, vol. 13, art. 54, 2024. <https://doi.org/10.3390/computers13020054>
- [42] R. R. Ibarra García et al., "Vehículo aéreo no tripulado empleando hardware Raspberry Pi 3 para aplicación en monitoreo remoto," *Pistas Educ.*, vol. 40, no. 130, pp. 1687–1707, Nov. 2018. <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/1634>
- [43] S. Karthikeyan et al., "A systematic analysis on RaspberryPi prototyping: Uses, challenges, benefits, and drawbacks," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14397–14417, Aug. 2023. <https://doi.org/10.1109/JIOT.2023.3262942>
- [44] P. Patel, N. Gupta, and S. Gajjar, "Real-time voice recognition system using tiny ML on Arduino Nano 33 BLE," in *Proc. IEEE Int. Symp. Smart Electron. Syst. (iSES)*, pp. 385–388. <https://doi.org/10.1109/iSES58672.2023.00085>
- [45] Z. Ksira et al., "A novel embedded system for real-time fault diagnosis of photovoltaic modules," *IEEE J. Photovoltaics*, vol. 14, no. 2, pp. 354–362, 2024. <https://doi.org/10.1109/JPHOTOV.2024.3359462>
- [46] N. I. Ramli, M. I. M. Rawi, and F. N. N. Rebutan, "Integrated smart home model: An IoT learning inspired platform," *Int. J. Web-Based Learn. Teach. Technol.*, vol. 3, no. 1–14, Mar. 2022. <https://doi.org/10.4018/IJWLTT.20220501.0a1>
- [47] M. Giriraj and P. Anvesh, "Controlling of servomotors according to pitch, yaw, and roll motions of accelerometer," in *Proc. Int. Conf. Energy Eff. Technol. Sustain. (ICEETS)*, 2016, pp. 886–889. <https://doi.org/10.1109/ICEETS.2016.7583873>
- [48] K. Singh, M. Mehndiratta, and M. Feroskhan, "Quadplus: Design, modeling, and receding-horizon-based control of a hyperdynamic quadrotor," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 1766–1779, 2022. <https://doi.org/10.1109/TAES.2021.3133314>
- [49] V. Quintero et al., "Baterías de ion litio: Características y aplicaciones," *Rev. I+D Tecnol.*, vol. 17, no. 1, Jan. 2021. [Online]. Available: <https://portal.amelica.org/ameli/jatsRepo/339/3392002003/index.html>
- [50] H. M. O. Canilang, A. C. Caliwag, and W. Lim, "Design of modular BMS and real-time practical implementation for electric motorcycle application," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 2, pp. 519–523, 2022. <https://doi.org/10.1109/TCSII.2021.3093937>
- [51] C. Stansbury, "A general description of D-C digital voltmeters," *Trans. Amer. Inst. Electr. Eng., Part I: Commun. Electron.*, vol. 80, no. 5, pp. 465–470, 1961. <https://doi.org/10.1109/TCE.1961.6373002>
- [52] L. Chen, J. Guo, B. Zhu, and Z. Zhang, "Electronic nonlinearity of full-bridge PWM inverter for zero power PEMS system," *IEEE Access*, vol. 10, pp. 37670–37677, 2022. <https://doi.org/10.1109/ACCESS.2022.3165307>

- [53] D. Y. Jaimes and F. Fajardo, "Caracterización de motores DC de imán permanente mediante un sistema motor-generador," *Rev. Bras. Ensino Fis.*, vol. 44, art. e20220199, 2022. <https://doi.org/10.1590/1806-9126-RBEF-2022-0199>
- [54] A. A. A. Ismail and A. Elnady, "Advanced drive system for DC motor using multilevel DC/DC buck converter circuit," *IEEE Access*, vol. 7, pp. 54167–54178, 2019. <https://doi.org/10.1109/ACCESS.2019.2912315>
- [55] S. J. Chapman, *Máquinas Eléctricas*. México DF: Mexico: McGraw-Hill, 2012.
- [56] P. F. Proença and P. Simões, "TACO: Trash annotations in context for litter detection," arXiv:2003.06975, 2020. [Online]. Available: <https://doi.org/10.48550/arXiv.2003.06975>
- [57] D. M. Chilukuri, S. Yi, and Y. Seong, "A robust object detection system with occlusion handling for mobile devices," *Comput. Intell.*, vol. 38, no. 4, pp. 1338–1364, 2022. <https://doi.org/10.1111/coin.12511>
- [58] Y. Li, H. Huang, Q. Xie, L. Yao, and Q. Chen, "Research on a surface defect detection algorithm based on MobileNet-SSD," *Appl. Sci.*, vol. 8, no. 9, 2018. <https://doi.org/10.3390/app8091678>

Understanding How Software Project Managers Learn Using Kolb's Learning Style Inventory

Entendiendo cómo aprenden los gestores de proyectos de *software* mediante el inventario de estilos de aprendizaje de Kolb

Mauricio Hidalgo¹, Laura M. Castro², and Hernán Astudillo³

ABSTRACT

Teaching project managers involves conveying complex concepts in an accessible manner while fostering strategic thinking and promoting problem-solving skills using specific methodologies. Additionally, continuing education requires techniques for meaningful learning that balance theory with practice and build upon professional experience. However, unlike other software development roles, little is understood about how software project managers (SPMs) can learn effectively. Therefore, this research aimed to identify teaching techniques that better align with SPMs' learning styles. Kolb's learning styles test was administered to 27 project managers to determine their preferences in this regard. The surveyed SPMs leaned towards abstract conceptualization, exhibiting a balance between active experimentation and reflective observation, with the *thinking* style best reflecting these preferences. Building on this, some SPM-centered teaching guides were adapted from existing knowledge on teaching professionals with Kolb's *thinking* learning style.

Keywords: Software engineering education, project managers, continuing education, learning styles, Kolb's learning style inventory

RESUMEN

Enseñar a los gestores de proyectos implica transmitir conceptos complejos de manera accesible, fomentando el pensamiento estratégico y promoviendo habilidades de resolución de problemas mediante metodologías específicas. Además, la educación continua requiere técnicas de aprendizaje significativo que brinden un balance entre la teoría y la práctica y construyan sobre la experiencia profesional. Sin embargo, a diferencia de otros roles en el desarrollo de *software*, existe un conocimiento limitado sobre la manera en que los gestores de proyecto de *software* (SPMs) aprenden de manera efectiva. Por ende, esta investigación tuvo como objetivo identificar las técnicas de enseñanza que mejor se alineen con los estilos de aprendizaje de los SPMs. Se aplicó la prueba de estilos de aprendizaje de Kolb a 27 gestores de proyecto para determinar sus preferencias en este aspecto. Los SPMs encuestados se inclinaron hacia la conceptualización abstracta, presentando un balance entre la experimentación activa y la observación reflexiva. El estilo *thinking* fue el que mejor representó estas preferencias. Con base en esto se adaptaron algunas guías de enseñanza centradas en SPMs a partir del conocimiento existente sobre la enseñanza para profesionales con el estilo de aprendizaje *thinking* de Kolb.

Palabras clave: Educación para la ingeniería de *software*, gestores de proyecto, educación continua, estilos de aprendizaje, inventario de estilos de aprendizaje de Kolb

Received: July 17th 2024

Accepted: January 15th 2025

Introduction

A software project manager (SPM) is responsible for leading a team of developers and ensuring that projects are completed on time and within the desired scope. Leadership, technical knowledge, and experience allow the SPM to manage the team in an ever-evolving technological landscape, and continuous education serves as an *open door* to further develop these skills. Didactic approaches have been designed to meet specific academic needs [5], and traditional software engineering curricula can benefit from integrating specific software project management knowledge, such as the Software Extension to the PMBOK Guide, in order to better align with industry requirements [21]. While a simple theoretical approach may seem clear, facilitating the acquisition of new tools, methodologies, and best practices through continuous education is not without challenges. Software project management education (SPME) is essential for providing students with appropriate competencies to effectively serve as professionals in the software industry [12]. Although the teaching model based

on course objectives is a traditional standard for software project management courses [35], key insights for SPMs extend beyond these objectives:

- Determining the problem, evaluating design alternatives, and using appropriate process models [9].
- Planning and estimating, measuring and controlling, leading and communicating, and managing risk [11].
- Using digital communication channels to develop knowledge collection and sharing [2].
- Leadership, team building, human resource policies, and project learning tools [4].

¹PhD in Informatics Engineering. Affiliation: Academic researcher, Universidad Finis Terrae, Chile. E-mail: mhidalgo@uft.cl

²PhD in Computer Science. Affiliation: Professor, Universidade da Coruña, Spain. E-mail: lcastro@udc.es

³PhD in Information and Computer Science. Affiliation: Research professor, ITISB-Universidad Andrés Bello, Chile. E-mail: hernan.astudillo@unab.cl



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Continuous education, learning, and improvement are key principles in software project management, which are aimed at ensuring software success and longevity [34]. The collaborative, complementary, sustained, and simultaneous learning principles of continuous education for SPMs can yield significant benefits [3], but, after all, one challenge remains: *How can a meaningful teaching and learning process for SPMs be achieved?*

For swift reference, this paper is organized as follows. The **Context and research question** section presents the context, a description of Kolb's learning style inventory (KLSI), and the research objectives and question. The **Related work** section reviews related works in the field of SPME. **Research method: survey** outlines the research methodology, including the use of KLSI to assess the learning preferences of SPMs. The **Results** section presents the results of a survey applied within the framework of this research, as well as the identified learning style of the SPM sample, and it proposes tailored educational strategies. The **Experimental validation** section presents the results obtained for the experimental validation of the proposed learning strategies, which were tested in a postgraduate course over four sessions, with a focus on assessing their relevance for the students. In addition, **Threats to validity and limitations** discusses the study's limitations and threats to validity { considering external, internal, conclusion, and construct validity. Finally, the **Discussion, Conclusions**, and **Future work** sections discuss the implications of our results for improving SPME, present the conclusions of this work, and suggest directions for future research, respectively.

Context and research question

The role of an SPM is multifaceted and evolves with technological advancements and industry demands. This profession requires a blend of organizational prowess, strategic planning, and adept communication to oversee and steer the development of intricate software projects. The skills required for an SPM include:

- Adaptability to new technologies, methodologies, and changing project requirements [27].
- Strong leadership, technical, and analytical skills, as well as decision-making abilities, to understand and manage the complexities of software development projects [17, 28].
- Interpersonal and communication skills for managing teams and stakeholders and ensuring project requirements [27, 30].
- Soft skills, including negotiation, conflict resolution, team motivation, and emotional intelligence [10, 22, 30].

The above-presented { non-exhaustive { list reveals several challenges that can be addressed through innovative educational strategies for teaching SPMs and fostering a more practical and immersive learning environment, bridging the gap between theoretical concepts and industrial applications. However, without clarity on the learning styles

and preferences of SPMs, the selection of these techniques becomes a complex issue.

Regarding the educational methodology, Kolb [18] defined the concept of *learning* as the process through which knowledge is built by transforming experience: knowledge arises from the merging of grasping and reshaping experience. He introduced the experiential learning cycle (Fig. 1), a recursive process finely tuned to both the learning context and the subject under study [18].



Figure 1. The experiential learning cycle

Learning styles delineate how individuals traverse the learning cycle, guided by their inclination towards four distinct learning modes [18]: experiencing (concrete experience, or CE), reflecting (reflective observation, or RO), thinking (abstract conceptualization, or AC), and acting (active experimentation, or AE). However, over time, empirical and clinical research has demonstrated that these initial four learning modes can be further elaborated into a nine-style typology [18] known as *Kolb's learning style inventory* (Fig. 2).

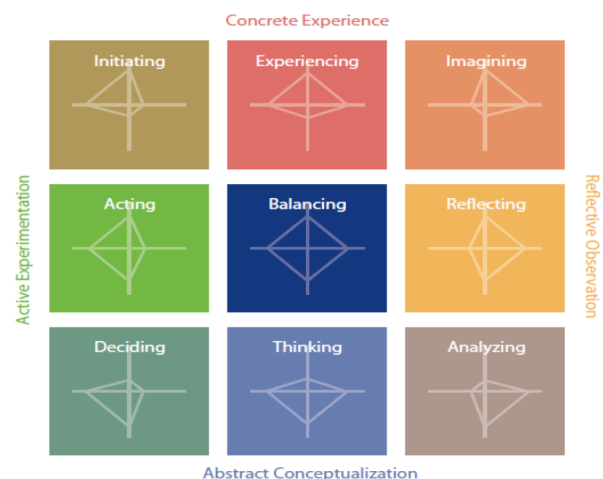


Figure 2. The learning styles in KLSI

KLSI expands upon the original model by establishing the nine learning modes detailed below.

- **Initiating.** Defined by the capacity to take action in response to experiences and situations, this style involves both AE and CE.

- **Experiencing.** Marked by a deep engagement with experiences to derive meaning, this style combines CE while balancing both AE and RO.
- **Imagining** is characterized by the ability to envision possibilities through observation and reflection on experiences, blending CE with RO.
- **Reflecting.** Distinguished by the ability to link experiences with ideas through extended reflection, this style integrates RO with both CE and AC.
- **Analyzing** is characterized by the ability to structure and systematize ideas through reflection, combining RO with AC.
- **Thinking.** Defined by the capacity for focused, abstract, and logical reasoning, this style uses AC while balancing both AE and RO.
- **Deciding.** Defined by the ability to use theoretical models and frameworks to make decisions and determine courses of action, this style combines AC with AE.
- **Acting.** Characterized by a strong drive for goal-oriented actions that integrate both people and tasks, this style emphasizes AE, with a balance of CE and AC.
- **Balancing.** Defined by adaptability, this style involves weighing the benefits of action vs. reflection, as well as experience vs. thinking. It balances all four learning modes: CE, AC, AE, and RO.

Considering the skills required and the conceptual model presented by KLSI, software engineering education (SEE) stands to benefit from identifying the predominant learning style among SPM practitioners. This approach can help to align instructional methods with the preferred learning styles, thereby enhancing skills development. Therefore, our objective and its corresponding research question are as follows:

Research objective. To identify the predominant learning styles among SPMs using KLSI and propose tailored educational strategies that align with these preferences. By understanding how SPMs learn, we aim to bridge the gap between theoretical knowledge and practical application, thereby enhancing the effectiveness of training programs in SEE.

Research question. In a given sample of SPMs, what are the common teaching and learning preferences?

Rationale. Understanding the learning and teaching preferences of SPMs allows educators to develop tailored activities and strategies that effectively meet unique needs and learning styles within a given practitioner group.

Why Kolb's model? Kolb's experiential learning theory (ELT) was chosen for this study due to its comprehensive framework, which bridges learning styles with practical application. This is essential for SEE. While other models (e.g., VARK* or Honey-Mumford[†]) focus on specific aspects related to learning preferences [such as sensory modalities or adaptations of Kolb's framework], Kolb's model emphasizes a cyclical process of learning through experience, reflection, conceptualization, and

experimentation. This iterative approach aligns closely with the dynamic and practical nature of software project management, where professionals must continuously adapt theoretical concepts to evolving project environments. Moreover, KLSI offers an empirically validated tool [18] to categorize learning preferences, providing a strong basis for designing educational strategies tailored to engineering roles.

Related work

The continuous evolution of software development techniques demands skilled project managers [1]. In higher education, graduate and post-graduate curriculum guides have been developed to meet industry demands, emphasizing the need for better software project management skills [6]. New teaching models based on professional certification within engineering education can improve the achievement of course objectives and the overall training of software engineering professionals [35]. Additionally, the use of e-learning platforms and digital communication channels can enhance the continuous improvement of project managers' competencies, reducing onboarding time and improving performance [2].

In the industry, SPMs must master both management and technical skills to effectively lead and control software development projects [17, 24]. A systematic review of SPME highlights the need for the continuous evaluation and improvement of teaching methods and tools to ensure quality education and meet industry demands [12]. In terms of teaching methods, serious games can effectively teach software engineering and project management concepts, engaging students in a dynamic and safe learning environment [6]. Iteratively refined courses that include evidence-based readings, quizzes, in-class activities, and ambitious projects can address common challenges in teaching software project management [25]. Additionally, tools like ProMES are used to train project managers on basic techniques such as CPM, PERT, and RACI, enhancing their understanding through pedagogical approaches [16]. Furthermore, project management software packages can reduce issues related to management and control in educational software projects [33].

In summary, the effective teaching of software project management necessitates a blend of methods and interactive learning tools. Despite this, there is limited evidence on the impact of learning styles on practitioners in SEE, pointing to a critical gap in enhancing educational strategies. While previous research has explored various aspects of SPME, most studies focus on general methodologies or specific pedagogical tools without addressing participants' individual learning preferences.

For instance, [12] conducted a systematic review of teaching methods in project management education but did not examine the role of learning styles in their effectiveness.

*The VARK model (visual, aural, read/write, kinesthetic) assesses four sensory modalities that people use to learn.

[†]The four different ways in which people prefer to learn that Honey and Mumford have identified relate to a different stage in the learning cycle. These are the Activist, the Reflector, the Theorist, and the Pragmatist. In this model, Mumford and Honey describe the learning styles as a continuum that one moves through over time.

Similarly, [25] proposed an iterative course based on theoretical objectives but lacked adaptation to specific learning styles.

Our work advances this field by applying KLSI to identify and characterize the predominant learning style among SPMs. This allows designing evidence-based educational strategies that align with their preferences, incorporating both practical and theoretical approaches. Unlike studies such as [20], which highlight the absence of learning style integration in engineering education, our research incorporates learning preferences and teaching strategies to enhance SPM training programs.

Research method: survey

This section summarizes the survey employed as our research method, its implementation, its inclusion and exclusion criteria, its underlying assumptions, and the profile of its participants.

Kolb's learning style test survey

To guide our research, we drew upon insights from [29] and [13] regarding survey application and results analysis. Given the nature of our study, we utilized the KLSI test for data collection. This test comprises 12 sentence completion questions (Table I), where participants rank provided suffixes on a scale of 1 to 4, with 4 indicating the highest preference.

Table I: Kolb's test applied

PREFIX	SUFFIX
Q1. When I learn...	I prefer to rely on my sensations and feelings
	I prefer to look and pay attention
	I prefer to think about ideas
	I prefer to do things
Q2. I learn best when...	I trust my hunches and feelings
	I listen and observe carefully
	I trust my logical thoughts
	I work hard to get things done
Q3. When I am learning...	I have strong feelings and reactions
	I am reserved and calm
	I seek to reason about things that are happening
	I feel responsible for things
Q4. I learn through...	Feelings
	Observations
	Reasoning
	Actions
Q5. When I learn...	I am open to new experiences
	I take into account all related aspects
	I prefer to analyze things by breaking them down into their component parts
	I prefer to do things directly

Continued on next page

Table I: Kolb's test applied (Continued)

Q6. When I am learning...	I am an intuitive person
	I am an observant person
	I am a logical person
	I am an active person
Q7. I learn best through...	Relationships with my peers
	Observation
	Rational theories
	The practice of the topics covered
Q8. When I learn...	I feel involved in the topics covered
	I take my time before acting
	I prefer theories and ideas
	I prefer to see the results through my own work
Q9. I learn best when...	I rely on my intuitions and feelings
	I rely on personal observations
	I take into account my own ideas about the subject matter
	I personally try out the task
Q10. When I am learning...	I am open-minded
	I am a reserved person
	I am a rational person
	I am a responsible person
Q11. When I learn...	I get involved
	I prefer to observe
	I prefer to evaluate things
	I prefer to take an active attitude
Q12. I learn best when...	I am receptive and open-minded
	I am careful
	I analyze ideas
	I am practical

Source: Authors

The test structure facilitates the assessment of participants' inclinations toward each of the four learning modes and allows categorizing each individual into a specific learning style [18].

Inclusion/exclusion criteria

Our research established the following inclusion/exclusion criteria for the participants:

Inclusion criteria. To qualify for the study, the respondents had to meet the following criteria:

- Hold the position of SPM within the industry.
- Have no prior experience with the KLSI test.

Exclusion criteria. To minimize potential biases, individuals would be deemed ineligible for participation if they met any of the following criteria:

- Lack of proficiency in reading English or Spanish.
- Work directly with any of the researchers.

Assumptions. Our research was based on the following assumptions:

- **Assumption 1.** The KLSI test, designed for adults, applies to professionals without necessitating adjustments or modifications.
- **Assumption 2.** The sample will offer comprehensive insight into learning styles, thereby facilitating the identification of representative trends.

Test execution

For our research, the execution of the survey considered the following aspects:

Test administration. The test was conducted with voluntary participation and anonymously collected responses. Invitations were distributed to 83 individuals currently employed as SPMs through professional networks, inviting them to participate in the study. The participants were provided with a link to access.

Characterization of the survey sample. The test was administered over 60 days, from the second week of January 2024 to the first week of March 2024. Ultimately, a sample of 27 participants who met the criteria was obtained, representing individuals from Chile, Costa Rica, Argentina, Colombia, and Spain. From the total number of invitations sent, 27 individuals completed the survey, representing a response rate of 32.5%. The distribution of participants by country is as follows:

- **Chile:** eight participants (29.6%)
- **Costa Rica:** six participants (22.2%)
- **Argentina:** five participants (18.5%)
- **Colombia:** four participants (14.8%)
- **Spain:** four participants (14.8%)

This distribution reflects the participation rate by country, with Chile representing the largest proportion of the sample.

Results

This section presents the results of the applied survey, indicating the learning modes and characterizing the learning style of the SPMs. It provides appropriate teaching/learning strategies to serve as a baseline for educators and researchers in SEE.

SPM sample and representativeness

The SPM group, consisting of 27 respondents, included 1 junior, 2 mid-senior, and 24 senior practitioners. The test results for each individual are presented in Table ???. To determine the validity of the mean as a representative measure of each learning mode for the sample, we employed the Student's t-test [31] and defined the following hypotheses:

$H_0 \rightarrow$ The mean of each learning mode is not representative of the SPM sample[‡].

$H_1 \rightarrow$ The mean of each learning mode is representative of the SPM sample.

According to Student's t-distribution table (two-tailed), with a significance level of 0.05 (corresponding to 95% confidence) and 26 degrees of freedom ($27 - 1 = 26$), the critical value for the sample is ± 2.056 . Utilizing the scores from Table ??, and taking the median as our hypothetical population mean, Table III presents this reference value alongside the sample mean and the standard deviation.

Table II. Kolb's test applied

Experience	Total_CE	Total_RO	Total_AC	Total_AE
Junior	20	29	35	36
Mid-senior	25	34	31	30
Mid-senior	23	35	28	34
Senior	21	30	36	33
Senior	26	44	27	23
Senior	20	36	40	24
Senior	24	40	27	29
Senior	21	34	35	30
Senior	26	25	40	29
Senior	22	30	36	32
Senior	27	33	33	27
Senior	22	32	40	26
Senior	25	21	34	40
Senior	22	38	26	34
Senior	19	38	39	24
Senior	31	19	25	45
Senior	32	26	37	25
Senior	22	21	33	44
Senior	21	33	37	29
Senior	36	34	27	23
Senior	20	30	39	31
Senior	21	23	33	43
Senior	31	27	29	33
Senior	23	40	30	27
Senior	34	19	37	30
Senior	24	34	40	22
Senior	26	25	37	32

Source: Authors

Table III. Mean, median, and standard deviation of the sample

	CE	RO	AC	AE
Mean	24.59	30.74	33.74	30.93
Median	23.00	32.00	35.00	30.00
Standard deviation	4.58	6.75	4.88	6.38

Source: Authors

To determine the t-statistic value of each learning mode, the following formula is employed:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where

- \bar{x} represents the sample mean,

[‡]By *representative of the sample*, we imply that the mean accurately reflects the central tendency of the sample's learning modes.

- μ represents the hypothetical population mean (median),
- s represents the sample standard deviation, and
- n represents the sample size.

The results of the t-statistic value for each mode are shown in Table IV.

Table IV. T-statistic value for each learning mode

Learning mode	t-statistic value
Concrete experience (CE)	1.80
Reflective observation (RO)	-0.97
Abstract conceptualization (AC)	-1.34
Active experimentation (AE)	0.75

Source: Authors

Finally, since the statistical value of each learning mode falls within the acceptance interval determined by the critical value of Student's t-distribution (± 2.056), the H_0 hypothesis is rejected, and H_1 is accepted. This indicates that the sample mean is representative of each learning mode in the sample.

Learning style of the SPMs

Based on the results of the Student's t-test, and considering the sample mean as a representative measure of the population for each learning mode, it can be stated that the AC mode shows the highest inclination, suggesting a preference for theoretical analysis. However, there is a balance between AE and RO, indicating that the SPMs also have a strong inclination towards the practical application of learning and towards considering different perspectives before taking action. CE indicates a preference { though less pronounced when compared to the other modes { for learning through practical experiences. Overall, while the modes are not entirely balanced, there is a notable presence in all four modes, leading SPMs to be categorized under the *thinking* learning style, as depicted in Figure 3.

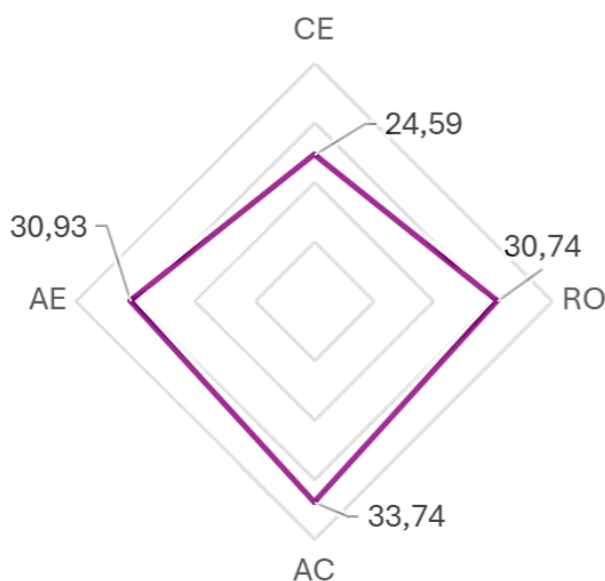


Figure 3. SPMs learning style characterization

The thinking learning style

The *thinking* learning style emphasizes AC while maintaining a balance between AE and RO. This style is characterized by a strong capacity for disciplined engagement in abstract reasoning, mathematics, and logic [18]. Individuals with the thinking style typically enjoy working with numbers and engaging in mental activities that require abstract reasoning and analytical skills. They often prefer to work with quantitative rather than qualitative information [18]. Moreover, they excel in planning and goal-setting, tend to express emotions in a controlled manner, and prefer precise and concise communication [18]. The learning strengths of individuals with the *thinking* style include logical analysis, rational decision-making, and proficiency in analyzing quantitative data.

Learning preferences and teaching strategies for SPMs

Individuals with the *thinking* style learn best in well-structured learning environments that feature clear directions and defined learning agendas. They thrive in settings where they can design experiments, manipulate data, or engage in abstract reasoning [18]. However, they typically prefer to work independently and require ample time to contemplate ideas thoroughly, which can pose challenges in collaborative settings and in remaining open to new ideas [18].

Learning preferences

This subsection explores the main learning preferences of the SPMs within the context of the *thinking* learning style from KLSI and its application in the software industry.

Analysis and reflection. SPMs tend to prefer detailed analysis and reflection on problems and solutions. To cater to this preference, consider the following:

- case study analysis,
- study of past project successes and failures, and
- reflective discussions on various software development approaches and methodologies [20].

Modeling and conceptualization. Creating models and conceptualizing abstract ideas are essential for SPMs in software processes. To meet these preferences, consider

- visualizing concepts using flowcharts and UML [19] and
- software process modeling using BPMN tools [32].

Problem-solving and critical thinking. SPMs often enjoy intellectual challenges and exercises that allow them to utilize their critical thinking and problem-solving skills. To cater to this preference, consider:

- complex case studies requiring deep analysis,
- problem-solving debates on controversies in the software industry [7], and

- practical exercises from the SPM perspective to identify and resolve software development issues.

Experimentation. SPMs often prefer learning through research and practical experimentation. To address this preference, consider

- practical projects for comparing methodologies and
- laboratory environments for experimenting with various tools and software development techniques [15].

Learning strategies

This subsection outlines the recommended teaching techniques/strategies for SPMs featuring the *thinking* learning style [18].

Problem-based learning (PBL). This technique involves presenting students with complex problems related to software development and guiding them through the problem-solving process. The expected learning outcome is to develop critical thinking and problem-solving skills by applying theoretical concepts in practical situations [23].

Collaborative learning. Given the learning challenges associated with the thinking style and the collaborative nature of an SPM's work, fostering collaboration among students through group projects and class discussions can be highly beneficial. The expected learning outcome involves sharing knowledge and experiences and fostering teamwork to collectively tackle complex challenges [14].

Experiential learning. A valuable strategy for SPMs includes gaining practical experience through projects and, ideally, internships in related industries. The expected learning outcome, particularly from internships, aims to apply theoretical knowledge in practical, unfamiliar settings, in addition to developing adaptability skills [8].

Case studies and simulations. Utilizing real case studies and simulations of software projects helps students to understand the complexities and challenges of software development in practical contexts. The expected learning outcome focuses on analyzing successful and failed cases in the software industry to derive lessons on best practices, as well as on the pitfalls to avoid [26].

Summary. Based on the identified learning styles of the sampled SPMs, Table ?? summarizes the key learning preferences observed, along with the teaching strategies recommended to enhance their learning experience, considering the learning mode to be strengthened. Furthermore, this study highlights the learning preferences of SPMs and proposes tailored educational strategies to optimize their development. Notably, PBL and collaborative learning emerge as particularly effective approaches for SPMs, given their inclination towards AC and AE. These strategies foster critical thinking, problem-solving, and teamwork, which are essential skills for effective project management in the software industry. Table ?? provides a summary of these strategies and their expected outcomes in terms of skills development. Based on the learning preferences identified herein, we recommend the integration of PBL and collaborative learning as core pedagogical strategies. These strategies align well with the

learning style of SPMs, promoting the acquisition of key project management skills.

Experimental validation

To validate the effectiveness of the proposed learning strategies, we conducted an experimental application [13] within a postgraduate course on software project management. The aim was to assess how strategies such as PBL, collaborative learning, and case study analysis could enhance student engagement and learning. The experiment involved 16 students who voluntarily participated in the experience. These strategies were designed to align with the learning preferences identified through KLSI, particularly focusing on students who exhibited a preference for AE and RO. By integrating these strategies into the course, we aimed to engage students whose learning styles were most conducive to these interactive, collaborative, and hands-on approaches. The effectiveness of the learning strategies was explored throughout these sessions, allowing for an assessment of both engagement and learning outcomes.

Experimental setup

The experiment took place over four sessions in the postgraduate course. A total of 16 students voluntarily chose to participate in this experience, and they were randomly divided into two groups:

- **Experimental group** (eight students). These individuals engaged in learning activities based on the proposed strategies.
- **Control group** (eight students). These individuals participated in traditional lecture-based learning with no integration of the experimental strategies.

Session details

The experiment consisted of three main sessions, followed by a fourth session to gather feedback from the students.

1. Session 1: problem-based learning (PBL)

- *Experimental group.* The students were given a real-world project failure scenario and worked in teams to identify the causes and propose recovery solutions.
- *Control group.* The students attended a lecture on risk management and were assigned individual readings about the common causes of project failure.

2. Session 2: collaborative learning

- *Experimental group.* The students worked in pairs to develop a project plan for a new software product, including its scope, timeline, and resource management. They then presented their plans to the class for feedback.
- *Control group.* The students worked individually to create a project plan based on theoretical readings, which they then submitted for review.

3. Session 3: case study analysis

Table V. Recommended pedagogical strategies by learning mode

Learning mode	Recommended pedagogical strategy
Concrete experience	Practical projects, laboratory environments, experiments comparing methodologies
Reflective observation	Case study analysis, reflective discussions on project approaches and methodologies
Abstract conceptualization	Theoretical analysis, conceptual modeling tasks, problem-solving debates on software development
Active experimentation	Practical exercises to apply methodologies and real-time problem-solving in software development

Source: Authors

Table VI. Educational strategies and expected learning outcomes

Educational strategy	Expected learning outcome
Problem-based learning (PBL)	Enhances critical thinking, problem-solving skills, and the ability to apply theoretical knowledge in practical settings.
Collaborative learning	Fosters teamwork, communication skills, and the ability to tackle complex problems through group discussions and shared knowledge.

Source: Authors

- Experimental group. The students analyzed the case of a software project and discussed the factors that contributed to its success. They applied these lessons to their project plans developed in Session 2.
- Control group. The students read a case study individually and summarized the key points. This was followed by a brief group discussion.

4. **Session 4: feedback and survey.** After the completion of the activities, all students completed a survey to express their opinions on the learning strategies used during the sessions. The results were then explained to the students in a follow-up session, where they discussed their experiences and provided further insights into their perceptions of the activities.

Survey and results

After completing the activities, all 16 students filled out a survey to provide their opinions on the learning strategies used during the sessions. The survey consisted of five questions designed to assess student engagement, satisfaction, and the effectiveness of the learning activities. The questions were:

1. **How engaged did you feel during the activities?**
 - 1 (Not engaged) to 5 (Very engaged)
2. **How useful did you find the learning activities for developing project management skills?**
 - 1 (Not useful) to 5 (Very useful)
3. **To what extent do you think the learning activities helped you understand project management concepts?**
 - 1 (Not at all) to 5 (A great deal)
4. **How confident are you in applying the strategies learned during the activities to real-world projects?**
 - 1 (Not confident) to 5 (Very confident)

5. Overall, how satisfied are you with the learning activities in this course?

- 1 (Not satisfied) to 5 (Very satisfied)

The average survey results (Table ??) are divided into two groups: the experimental group and the control group.

The survey results indicate that the experimental group, which participated in the activities designed with PBL, collaborative learning, and case study analysis, reported significantly higher levels of engagement, usefulness, and satisfaction compared to the control group. The experimental group also showed greater confidence in applying the strategies learned during the course. In contrast, the control group, which followed traditional lecture-based learning, reported lower satisfaction and engagement levels. This suggests that the interactive and practical nature of the activities implemented in the experimental group had a more positive impact on their learning experience.

These results align with the learning preferences identified through KLSI in the previous study. The experimental group, which likely exhibited preferences for AE and RO (common traits for the *thinking* learning style), benefited from the hands-on, collaborative, and case study-based strategies. These strategies are well-suited for learners who prefer to engage in practical activities and analyze outcomes through reflection and collaboration. On the other hand, the control group, which received traditional lecture-based instruction, may not have fully engaged with their preferences, leading to lower levels of satisfaction and engagement.

Threats to validity and limitations

This section discusses the potential threats to the validity of our study and outlines how they were mitigated. We followed the standard categories: conclusion, internal, construct, and external validity.

Conclusion validity

Conclusion validity examines whether the observed relationship between the treatment (KLSI) and the results

Table VII. Comparison of experimental and control group responses

Question	Experimental group	Control group
Engagement level	4.625	2.250
Usefulness of learning activities for project management skills	4.750	3.125
Effectiveness in understanding project management concepts	4.875	2.625
Confidence in applying strategies to real-world projects	4.625	2.375
Overall satisfaction	4.875	2.375

Source:

Authors

(the identified learning styles and educational strategies) is statistically significant. The use of the Student's t-test, with a confidence level of 95%, ensures the robustness of the conclusions. While the sample size of 27 participants limits the detection of smaller effects, the statistical methods employed mitigate this threat and provide a solid foundation for the findings.

Internal validity

Internal validity aims to determine whether the relationship between the treatment and the outcome is causal. Potentially confounding variables, such as the participants' prior exposure to management training or cultural influences, were minimized by ensuring that the respondents had no prior experience with KLSI. Additionally, the consistency in the responses across participants supports the causal interpretation of the findings.

Construct validity

In this case, construct validity focused on whether KLSI accurately captured the theoretical constructs it aimed to measure. The KLSI is an empirically validated tool for identifying learning styles in adults, making it suitable for this study. Although the tool was not specifically tailored to SPMs, its broad applicability ensures reliable measurement.

External validity

External validity examines the extent to which the results of this experimental study can be generalized beyond the specific sample to other contexts, populations, or learning environments. In this experiment, the learning strategies, *i.e.*, PBL, collaborative learning, and case study analysis, were tested in a postgraduate SPM course with 16 volunteers. While this experiment was conducted in a controlled setting with a specific group of postgraduate students, the strategies used are expected to have broader applicability in similar educational contexts, particularly in learners whose preferences align with active and experiential learning styles, as identified through KLSI.

The experimental design, which incorporated interactive, collaborative, and hands-on activities, was well-suited for students with a preference for AE and RO. Given that these strategies are grounded in active learning principles, they can likely be applied in other postgraduate or professional development programs in SPM and other related fields, particularly those focused on developing both technical and soft skills.

However, the study's external validity is somewhat limited by the specific demographic of the sample, which comprised only 16 students from a particular academic background

and geographic region. Although the diverse backgrounds and professional experiences of the participants allowed for a broad exploration of the strategies' effectiveness, future studies could enhance generalizability by including larger samples that also stem from different academic settings or industries. For example, expanding the sample to include students from other engineering disciplines or professionals with varying levels of experience in project management could provide further insights into the applicability of these strategies in different contexts. Additionally, replicating the experiment in other cultural or geographic regions could help to determine whether these strategies are equally effective across diverse educational systems.

The successful application of the proposed strategies in this experimental context suggests that they could be adapted and implemented in various educational and professional development settings, but further, more diverse research is needed to fully assess their applicability in other environments.

Limitations

While this study provides valuable insights, certain limitations should be acknowledged:

- *Sample size.* The relatively small sample size limits the detection of subtler effects. However, the identified patterns are consistent and meaningful.
- *Geographical scope.* The participants were predominantly from Spanish-speaking countries, which may affect applicability to other cultural contexts.
- *Self-reported data.* The reliance on self-assessment in KLSI may introduce minor biases, although the tool's design minimizes this risk.

By addressing these threats and limitations, this study provides a robust foundation for understanding the learning styles of SPMs and offers an adaptable framework for diverse educational contexts.

Discussion

This article explored the use of Kolb's test to determine the learning style that best represents a sample of SPMs according to KLSI [18]. Our research revealed that SPMs tend to have a strong preference for theoretical analysis and the practical application of concepts, with a particular inclination towards AC and AE, as well as a balance between RO and CE. This suggests that this group of SPMs leans towards the *thinking* learning style. A statistical

analysis showed that this result is representative of the sample. Thus, it is understood that SPMs consider multiple perspectives before making decisions [2], implying the need for an educational approach that combines theoretical and practical methods [16, 33].

The findings of this study on SPMs' learning styles, particularly their inclination towards the *thinking* learning style, align with the specialized literature, which emphasizes the importance of both technical and management skills in effective project management [17, 24]. The emphasis on AC as a dominant learning mode among SPMs corroborates the notion that project managers benefit from strong analytical and problem-solving skills, as supported by studies such as [12] and [35].

Moreover, our research expands the current understanding by incorporating KLSI, highlighting a specific need for tailored educational strategies that address the unique learning preferences of SPMs. This finding aligns with the work of [17], which underscores the necessity of continuous improvement in teaching methods to meet industry demands [25]. Additionally, the balanced representation of RO and AE in our sample indicates a versatile learning approach, which is important for adapting to the dynamic nature of software project environments [6].

On the other hand, and contrary to the limited evidence on the consideration of learning styles in SEE [20], our results provide empirical support for the integration of learning style-based strategies. This is particularly relevant for enhancing the pedagogical approaches used in SPM training programs, as suggested by [16]. Our findings suggest that incorporating varied teaching methodologies, such as PBL and collaborative learning, can better cater to the diverse learning preferences observed among SPMs, thus potentially improving the efficacy of SEE programs [33].

Furthermore, the importance of practical experiences and hands-on learning, as indicated by the significant scores in the AE mode, aligns with previous research advocating for experiential learning in SEE [15]. This study not only reaffirms the value of such approaches but also emphasizes the need for educational curricula to evolve continually to accommodate the changing requirements of the software industry [12].

Addressing this gap could lead to optimized educational strategies and teaching methods adapted to the specific needs of SPMs [16, 33]. In this vein, we suggested a set of strategies that align with the identified learning preferences. These strategies include PBL, collaborative learning, experiential learning, and case studies. Their effectiveness was experimentally validated, with the experimental group achieving higher levels of engagement (4.625 vs. 2.250), perceived usefulness (4.750 vs. 3.125), and overall satisfaction (4.875 vs. 2.375) compared to the control group. These results support the implementation of these techniques to foster a more dynamic and adaptable learning environment that better responds to the current challenges of the software industry [1].

In summary, this research offers insights that can enhance the design and implementation of SEE programs. The alignment of our findings with the existing literature supports the need for an interdisciplinary approach that integrates both technical and managerial training, tailored to the

specific learning preferences of SPMs. Future research could further explore the impact of these educational strategies on the performance and career development of SPMs, thereby contributing to a more robust and responsive educational framework in software project management.

Conclusions

In this article, Kolb's test was used to identify the predominant learning styles in a group of 27 software project managers based on Kolb's learning style inventory. The results revealed a strong preference for theoretical analysis and the practical application of concepts, with a particular inclination towards abstract conceptualization and active experimentation. Additionally, a balance was noticed between reflective observation and concrete experience, suggesting that these SPMs lean towards the *thinking* learning style.

A statistical analysis, conducted using Student's t-test, demonstrated that the results are representative of the sample with a 95% confidence interval. This validates the robustness of our conclusions and suggests that the results could be generalized (with a larger sample) to a broader population of SPMs, at least within Spanish-speaking contexts.

Moreover, the geographical diversity of the participants reinforces the reproducibility of our findings, indicating that the identified learning preferences are not significantly influenced by the geographical location of the SPMs. The use of neutral prefixes in the KLSI test also helped to mitigate potential biases in the responses, ensuring the objectivity of the results.

Based on these findings, it could be argued that an educational approach combining theoretical and practical methods can improve the training of SPMs by aligning with their learning preferences. In this vein, strategies such as problem-based learning, collaborative learning, experiential learning, and the use of case studies and simulations are recommended. Implementing these strategies will not only improve the effectiveness of SPM training but also foster a more dynamic and adaptable learning environment that is better able to respond to the current challenges of the software industry.

In summary, this study not only provides a detailed insight into the learning preferences of SPMs; it also offers a practical framework for optimizing educational strategies in software project management. The findings are statistically significant and reproducible, underscoring the relevance and applicability of the results in educational and professional practice.

Future work

For future work, we plan to expand the sample size and include a more diverse group of SPMs to validate the generalizability of the learning preferences identified in this study. Additionally, we aim to conduct intervention-based research to implement and evaluate specific educational strategies such as PBL and collaborative learning, providing empirical evidence of their effectiveness in enhancing SPM training and project management performance, with a

particular focus on their integration into frameworks like PMBOK or Agile.

Acknowledgements

This work has been partially supported by Universidad Finis Terrae through the "Laboratorio de Investigación Avanzada en Ciencias de Datos en Derecho" - LIACDD UFT.

CRedit author statement

All authors have contributed to the current work, which constitutes part of Mauricio Hidalgo's doctoral research work, supervised by Dr. Astudillo and Dr. Castro.

Conflicts of interest

The authors have no competing interests to declare that are relevant to the content of this article.

Data availability

All the research data is provided in the article.

Statement on Artificial Intelligence (AI)

During the elaboration of this work, the authors used ChatGPT to improve the writing. After using this tool, the authors reviewed and edited the content as necessary and took full responsibility for the content of the publication.

References

- [1] S. Al-Saqqah, S. Sawalha, and H. AbdelNabi, "Agile software development: Methodologies and trends," *Int. J. Interact. Mobile Tech.*, vol. 14, no. 11, pp. 246-270, 2020. <https://doi.org/10.3991/ijim.v14i11.13269>
- [2] S. Atanasijevic, M. Zahar, D. Rancic, T. Atanasijevic, and M. Đorđević, "Creating an educational framework for project managers at a software company: A sample approach," in *Proc. Int. Sci. Conf. - Sinteza 2023*, 2023, pp. 199-205. <https://doi.org/10.15308/Sinteza-2023-199-205>
- [3] S. Atanasijevic, M. Zahar, D. Rančić, I. Vulić, and T. Atanasijevic, "Enabling digital growth through continuous education of project managers: A framework for collaborative, complementary, sustained, and simultaneous learning in software engineering organizations," *Vojnotehnicki Glasnik*, vol. 71, no. 4, pp. 1172-1197, 2023. <https://scindeks-clanci.ceon.rs/data/ipdf/0042-8469/2023/0042-84692304172A.pdf>
- [4] K. Ayas, "Learning through projects: Meeting the implementation challenge," *J. Eur. Ind. Train.*, vol. 22, no. 3, pp. 89-98, 1998.
- [5] C. Boiangiu, A. C. Firculescu, and I. Bucur, "Teaching software project management: The independent approach," *Romanian Economic Business Review*, vol. 10, no. 1, pp. 11-22, 2016. <https://ideas.repec.org/a/rau/journl/v10y2016i1p11-22.html>
- [6] C. Caulfield, D. Veal, and S. Maj, "Teaching software engineering project management { A novel approach for software engineering programs," *Math. Models Methods Appl. Sci.*, vol. 5, art. 87, 2011. <https://doi.org/10.5539/mas.v5n5p87>
- [7] P. Chen and Y.-C. Chan, "Enhancing creative problem solving in postgraduate courses of education management using project-based learning," *Int. J. High. Educ.*, vol. 10, no. 6, art. 11, 2021. <https://doi.org/10.5430/IJHE.V10N6P11>
- [8] S. Christov and M. E. Hoffman, "Experiential learning of software project management and software development via course collaboration," in *Proc. 50th ACM Tech. Symp. Comput. Sci. Educ.*, 2019, pp. 160-166. <https://doi.org/10.1145/3287324.3287457>
- [9] A. Davis, "Fifteen principles of software engineering," *IEEE Softw.*, vol. 11, pp. 94-96, 1994.
- [10] A. Ellahi, Y. Javed, M. F. Jan, and Z. Sultan, "Determining the effect of software project managers' skills on work performance," *Int. J. Inf. Technol. Proj. Manag.*, vol. 15, no. 1, pp. 1-20, 2023. <https://dx.doi.org/10.4018/IJITPM.333620>
- [11] R. Fairley, *Managing and leading software projects*. Hoboken, NJ, USA: Wiley, 2009.
- [12] M. Farooq, U. Omer, and R. Tahseen, "Software project management education: A systematic review," *J. Inf. Syst. Oper. Manag.*, vol. 9, no. 3, pp. 102-119, 2021. <http://dx.doi.org/10.21015/vtse.v9i3.702>
- [13] M. Felderer and G. H. Travassos, "Contemporary empirical methods in software engineering," *Contemp. Empir. Methods Softw. Eng.*, 2020. <https://api.semanticscholar.org/CorpusID:221346139>
- [14] R. Garcia, C. Treude, and A. Valentine, "Application of collaborative learning paradigms within software engineering education: A systematic mapping study," in *Proc. 55th ACM Tech. Symp. Comput. Sci. Educ.*, Portland, OR, USA, 2024, pp. 366-372. <https://doi.org/10.1145/3626252.3630780>
- [15] G. Gonçalves, R. Meneses, J. Faria, and R. M. Vidal, "A living lab for professional skills development in software engineering management at U.Porto," in *Proc. 2020 IEEE Global Eng. Educ. Conf. (EDUCON)*, 2020, pp. 1079-1085. <https://doi.org/10.1109/EDUCON45650.2020.9125099>
- [16] G. Gregoriou, K. Kiriopoulou, and C. Kyriklidis, "Project management educational software (ProMES)," *Comput. Appl. Eng. Educ.*, vol. 21, pp. 46-59, 2013. <https://doi.org/10.1002/cae.20450>
- [17] J. Jurison, "Software project management: The manager's view," *Commun. Assoc. Inf. Syst.*, vol. 2, art. 17, 1999.
- [18] D. Kolb and A. Kolb, *The Kolb Learning Style Inventory 4.0: Guide to Theory, Psychometrics, Research & Applications*, 2013. [Online]. Available: <https://learningfromexperience.com/research-library/the-kolb-learning-style-inventory-4-0/>
- [19] S. Kumari, "Visual modeling: 'Unlocking ideas and enhancing understanding: The power of visual modeling,'" *Int. J. Eng. Technol.*, vol. 12, no. 2, pp. 20-25, 2023. <https://doi.org/10.14419/ijet.v12i2.32334>

- [20] M. Mantyla, M. Jørgensen, P. Ralph, and H. Erdogmus, "Guest editorial for special section on success and failure in software engineering," *Empir. Softw. Eng.*, vol. 22, pp. 2281-2297, 2017. <https://doi.org/10.1007/s10664-017-9505-5>
- [21] A. Moreno, M. I. Sánchez Segura, F. Medina-Domínguez, L. Peters, and J. Araujo, "Enriching traditional software engineering curricula with software project management knowledge," in *Proc. 2016 IEEE/ACM 38th Int. Conf. Softw. Eng. Companion (ICSE-C)*, 2016, pp. 404-411. <https://ieeexplore.ieee.org/document/7883326>
- [22] S. H. Ow, "Improving soft skills of university students through software development team projects," *Asian Soc. Sci.*, vol. 4, art. 135, 2009.
- [23] B. Pérez and A. Rubio, "A project-based learning approach for enhancing learning skills and motivation in software engineering," in *Proc. 51st ACM Tech. Symp. Comput. Sci. Educ.*, 2020, pp. 309-315. <https://doi.org/10.1145/3328778.3366891>
- [24] L. Peters and A. Moreno, "Educating software engineering managers - revisited: What software project managers need to know today," in *Proc. 2015 IEEE/ACM 37th IEEE Int. Conf. Softw. Eng.*, vol. 2, 2015, pp. 353-359. <https://doi.org/10.1109/ICSE.2015.168>
- [25] P. Ralph, "Re-imagining a course in software project management," in *Proc. 2018 IEEE/ACM 40th Int. Conf. Softw. Eng. Softw. Eng. Educ. Train. (ICSE-SEET)*, 2018, pp. 116-125. <https://doi.org/10.1145/3183377.3183379>
- [26] A. H. Reed and M. Angolia, eds., *The value of simulation for learning project management*. Hershey, PA, USA: IGI Global Scientific Publishing, 2021. <https://doi.org/10.4018/978-1-5225-3471-6.CH002>
- [27] G. Ruhe and C. Wohlin, *Software project management in a changing world*. Berlin, Germany: Springer, 2014.
- [28] W. Scacchi, "Managing software engineering projects: A social analysis," *IEEE Trans. Softw. Eng.*, vol. SE-10, no. 1, pp. 49-59, 1984. <https://doi.org/10.1109/TSE.1984.5010198>
- [29] M. Shaw, "Software engineering education: A roadmap," in *Proc. Conf. The Future of Software Engineering*, 2000, pp. 371-380.
- [30] A. Sukhoo, A. Barnard, M. Eloff, J. Van der Poll, and M. Motah, "Accommodating soft skills in software project management," *Issues Inform. Sci. Inf. Technol.*, vol. 2, pp. 691-703, 2005. <https://doi.org/10.28945/860>
- [31] H. W. Thompson, R. Mera, and C. Prasad, "A description of the appropriate use of Student's t-test," *Nut. Neurosci.*, vol. 1, no. 2, pp. 165-172, 1998.
- [32] O. L. Vega-Márquez, J. Chavarriaga, M. Linares-Vásquez, and M. E. Sánchez, "Requirements comprehension using BPMN: An empirical study," in *Empirical Studies on the Development of Executable Business Processes*, 2019. https://doi.org/10.1007/978-3-030-17666-2_5
- [33] V. Stiehl, M. Danei, J. Elliott, M. Heiler, and T. Kerwien, "Effectively and efficiently implementing complex business processes: A case study," in *Empirical studies on the development of executable business processes*, D. Lubke and C. Pautasso, Eds. Cham, Switzerland: Springer, 2019, pp. 33-57. https://doi.org/10.1007/978-3-030-17666-2_3
- [34] B. Wu, "Modeling software maturity: A knowledge accumulation and continuous improvement approach," in *2012 3rd Int. Conf. Intell. Cont. Info. Proc.*, 2012, pp. 739-743.
- [35] H. Zhou, Y. Zhu, J. Wan, and Q. Wang, "Exploration of teaching mode of software project management based on engineering education professional certification," *J. Educ. Human. Soc. Sci.*, vol. 6, pp. 160-164, 2022. <https://doi.org/10.54097/ehss.v6i.4418>