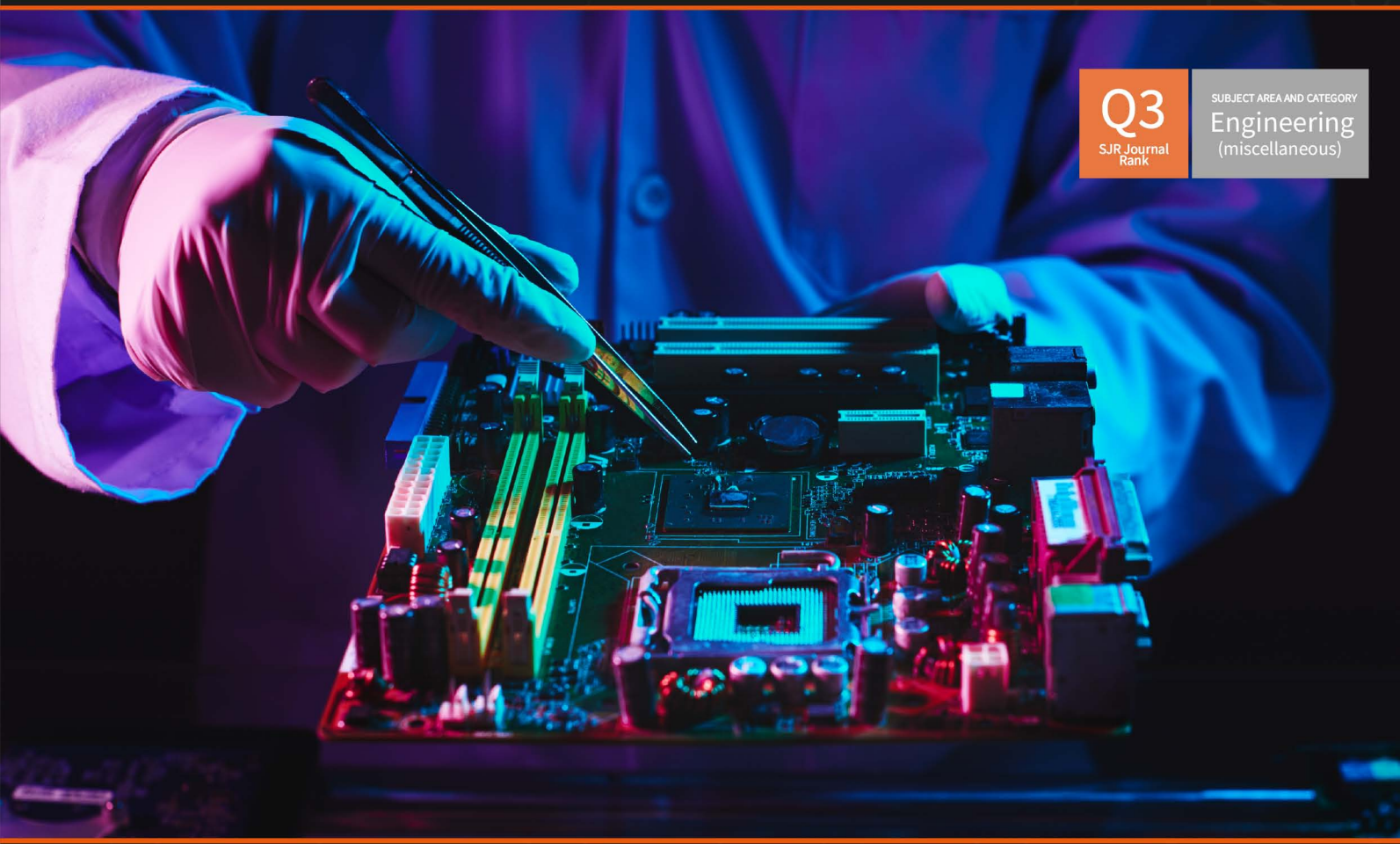
ISSUE: 35
SCOPUS
UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR

pISSN: 1390-650X
eISSN: 1390-860X
january / june 2026

INGENIUS

Revista de Ciencia y Tecnología



Indexed in: SCOPUS

Ingenius

Ingenius • Issue 35 • january-june 2026. Journal of Science and Tecnology of the Universidad Politécnica Salesiana of Ecuador. Publication dedicated to studies related to the Sciences of Mechanical Engineering, Electrical Engineering, Electronic Engineering, Mechatronic Engineering, Systems Engineering and Industrial Engineering.

Reviewers board

Editorial

Editorial Abya Yala (Quito-Ecuador),
Av. 12 de octubre N422 y Wilson,
Bloque A, UPS Quito, Ecuador.
Casilla 17-12-719 Teléfonos: (593-2) 3962800 ext. 2638
email: editorial@abyayala.org

Printing: 800 copies

Typographic system used in the composition of this document LaTeX.

Ingenius

Revista de Ciencia y Tecnologia

Issue 35

january – june 2026

ISSN impreso 1390-650X / ISSN electrónico 1390-860X

The administration of the journal is done through the following parameters:

The journal uses the academic anti-plagiarism system

The articles have an identification code (Digital Object Identifier)

The editorial process is managed through the Open Journal System

The articles of the present edition can be consulted in
http://revistas.ups.edu.ec/index.php/ingenius

Universidad Politécnica Salesiana del Ecuador

INGENIUS Journal, is indexed in the following Databases and scientific information systems:

SELECTIVE DATABASES

Scopus

Emerging Sources Citation Index

Google Scholar

SciELO Ecuador

BASE

Fuente Académica Plus EBSCOhost

SIS Scientific Indexing Services

EuroPub

REVIEWS EVALUATION PLATFORMS

MIAR

Quality Open Access Market

I2OR

SELECTIVE DIRECTORIES

latindex

Journal Seeker Research Bible

AcademicKeys Unlocking Academic Careers

Science Library Index

ULRICHSWEB™ Global Serials Directory

SELECTIVE SERIAL LIBRARY

redalyc.org

REDIB Red Iberoamericana de Innovación y Conocimiento Científico

refseek

SCIENTIFIC LITERATURE SEARCHERS OPEN ACCESS

DOAJ Directory of Open Access Journals

EXALEAD

WorldWideScience.org

Journals for Free

OTHER BIBLIOGRAPHICAL DATABASES

Dialnet

Journal TOCs The latest Journal Tables of Contents

PKP|INDEX

CATALOG OF INTERNATIONAL UNIVERSITY LIBRARIES

Editorial

Dear readers,

Scientific and technological progress is consolidated when knowledge is built through the integration of diverse contexts, disciplines, and perspectives. Issue 35 of our journal exemplifies this principle by bringing together scholarly contributions from Latin America, Asia, and the Middle East, thereby consolidating an international and multidisciplinary editorial space for the dissemination of scientific and technological knowledge.

From a geographical perspective, this issue highlights valuable international collaboration. Contributions from Mexico are represented by the Autonomous University of Baja California; Brazil by the Pontifical Catholic University of Paraná; India by Lovely Professional University; and Saudi Arabia by the University of Jazan. Within Latin America, and particularly from Ecuador, research contributions originate from well-recognized institutions such as the Pontifical Catholic University of Ecuador (Esmeraldas Campus), the Technical University of Machala, the National University of Loja, the University of Cuenca, UTE University, and Universidad Politécnica Salesiana.

This geographical diversity reflects both the journal's international positioning and the strengthening of South–South scientific networks, as well as the integration of global perspectives into applied research.

Regarding the disciplines and thematic areas addressed, this issue is characterized by a strong focus on engineering, artificial intelligence, energy systems, computing, and technologies aimed at solving real-world problems. The published articles include studies on energy consumption optimization in remote oil camps, advanced computational analyses for jet engine optimization using computational fluid dynamics (CFD) and multidisciplinary approaches, as well as research on electrical power systems and short-circuit analysis in overhead distribution networks. A particularly relevant aspect of this issue is the application of machine learning and deep learning techniques across multiple domains. These methodologies are used in medical fracture classification, semantic segmentation of defects in photovoltaic panels, urban accessibility assessment, emission estimation in internal combustion engines, and the classification of agro-industrial products such as specialty Loja coffee. Collectively, these contributions highlight the transversal role of artificial intelligence as a driver of technological innovation and data-driven decision-making.

The scientific and professional contributions of the articles published in this issue are substantial. On the one hand, they advance theoretical knowledge through robust models, algorithms, and reproducible methodologies; on the other, they offer practical solutions with direct impact on strategic sectors such as energy, healthcare, industry, transportation, environmental management, and urban planning. This balance between methodological rigor and practical applicability is a fundamental pillar of the journal's editorial policy.

We invite researchers, academics, and professionals to use the articles published in this issue as references, to critically engage with their findings, and to cite them in future scholarly work, thereby contributing to the visibility, impact, and continuity of the scientific knowledge generated.

Finally, we extend our sincere appreciation to all authors who entrusted their manuscripts to our journal, as well as to the reviewers and evaluators whose expertise and ethical commitment ensure the scientific quality and academic integrity of each publication. Their collaboration reaffirms the journal's role as a respected platform for disseminating high-quality research in science and technology.

John Calle-Siguencia, PhD

Editor in Chief

Table of contents

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.01
pISSN: 1390-650X / eISSN: 1390-860X

Optimization of energy consumption in oil remote camps through energy management technologies

Optimización del consumo energético en campamentos remotos petroleros mediante tecnologías de gestión energética

Edwin Illescas[1], Edison Laz[1], Manuel Rogelio Nevarez Toledo[1], Miguel Alberto Dávila-Sacoto[1,*]

Received: 02-04-2025, Received after review: 21-05-2025, Accepted: 29-09-2025, Published: 01-01-2026

Abstract

Optimizing resource use in remote operational fields is a key strategy for enhancing efficiency and reducing operating costs. This study adopted a descriptive–experimental design with a quantitative orientation to analyze the energy consumption of a remote oil transportation camp, utilizing historical records collected since 2021. The primary objective was to achieve an energy consumption reduction of 15–30%. Using RETScreen, improvements in energy systems were simulated, resulting in an initial annual electricity consumption of 271,143 kWh and associated costs of USD 27,454. With the proposed photovoltaic system, designed to optimize energy use through advanced energy management technologies, annual consumption decreased to 185,878 kWh/year, substantially reducing dependence on the national interconnected grid and on polluting energy sources. Energy efficiency improved by 14.3% in air-conditioning systems and by 38.6% in electrical systems such as pumps, compressors, and lighting, resulting in an overall average energy savings of 30.9%. In terms of environmental impact, CO_2 emissions were reduced from 58.8 tCO_2 to 40.7 tCO_2, equivalent to the carbon sequestration of approximately 1.7 hectares of forest. The estimated return on investment is 10.7 years. These results demonstrate the technical, economic, and environmental feasibility of implementing energy-efficiency technologies to enhance sustainability in remote oilfield operations.

Keywords: remote oil camps, energy management, energy optimization, RETScreen, photovoltaic systems, management technologies.

Resumen

Optimizar los recursos en campos remotos es un criterio fundamental para la eficiencia y la reducción de costos operativos. El estudio tuvo un enfoque descriptivo-experimental, con orientación cuantitativa, en el que se analizó el consumo energético de un campamento remoto de transporte de petróleo, utilizando registros históricos desde 2021. El objetivo principal fue la reducción del consumo energético en un rango del 15% al 30%. Mediante el uso de RETScreen se simularon mejoras en los sistemas energéticos, obteniéndose como resultado un consumo eléctrico anual de 271 148 kWh/año, con gastos asociados de USD 27 454. Con la propuesta fotovoltaica, orientada a optimizar el consumo mediante tecnología de gestión energética, el consumo se redujo a 185 878 kWh/año, disminuyendo significativamente la dependencia del Sistema Nacional Interconectado y el uso de fuentes contaminantes. La eficiencia en sistemas de climatización mejoró en un 14.3%, mientras que en los sistemas eléctricos —bombas, compresores e iluminación— se alcanzó una mejora significativa de 38.6%, con un ahorro promedio de 30.9%. Respecto a las emisiones, estas se redujeron de 58.8 tCO_2 a 40.7 tCO_2, lo que equivale a la absorción de 1.7 hectáreas de bosque con un retorno de inversión estimado en 10.7 años. En conclusión, este proyecto resulta viable desde el punto de vista de la aplicación de tecnologías de eficiencia energética, al incorporar mejoras técnicas que consolidan una solución económicamente rentable y ambientalmente sostenible.

Palabras clave: campamentos remotos petroleros, gestión energética, optimización energética, RETScreen, sistemas fotovoltaicos, tecnologías de gestión.

[1,*]Pontificia Universidad Católica, Sede Esmeraldas, Ecuador.
Corresponding author ✉: madavila@pucese.edu.ec.

Suggested citation: E. Illescas, E. Laz, M. R. Nevarez Toledo and M. A. Dávila-Sacoto. "Optimization of energy consumption in oil remote camps through energy management technologies," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 9-20, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.01.

1. Introduction

The dynamism of the oil industry exerts a substantial influence on global economic and social development. Petroleum resources account for approximately one-third of the world's primary energy supply and contribute around 2.5% to the global GDP [1]. Consequently, crude oil transportation represents a critical component of the fossil fuel supply chain. Remote operational camps that support this infrastructure face persistent challenges related to energy efficiency, operational autonomy, and environmental sustainability. Most of these facilities rely on conventional power generation systems with high emission factors, which drive operating costs and significantly increasetheir carbon footprint.

The development of intelligent energy management technologies entails the integration of artificial intelligence (AI) and machine learning algorithms into SCADA and BEMS (Building Energy Management Systems) platforms [2, 3]. These systems enable the analysis of consumption patterns, the prediction of load curves, and the real-time optimization of energy use [4, 5], allowing them to adapt to stochastic disturbances such as demand fluctuations and variability in renewable energy sources [6]. In parallel, demand response programs (DRP) incentivize users to modify their consumption in response to network signals, such as elevated prices or targeted economicincentives [7, 8].

Following these principles, Energy Management Systems (EnMS), aligned with the International Organization for Standardization ISO 50001 standard, constitute an effective strategy for enhancing corporate energy efficiency and improving operational performance [9, 10]. These systems rely on the implementation of progressive energy policies, clearly defined objectives, and targeted actions aimed at optimizing energy performance [11].

The sustained growth in global energy consumption, along with the environmental impacts associated with conventional energy sources, has generated increasing international interest in more efficient energy management strategies. In this context, RETScreen Expert, developed by Natural Resources Canada, offers a powerful platform for assessing the technical and economic feasibility of energy projects. It enables the modeling of electrical and thermal systems and the calculation of key performance indicators such as energy savings, emission reductions, and economic returns. Collectively, these capabilities provide a solid foundation for data-driven decision-making to support an effective energy transition [12].

Among the renewable energy solutions that can be managed through the RET Screen platform, photovoltaic solar technology stands out for its ability to convert direct solar radiation into electricity using silicon-based panels. This technology is distinguished by its versatility, low maintenance requirements, and long service life (25–30 years). It produces no greenhouse gas (GHG) emissions during operation, and its successful implementation depends on prior assessments of solar irradiance, panel tilt, and overall system efficiency [13, 14].

In the lighting sector, LED luminaires powered by renewable energy sources represent an efficient and sustainable alternative, as they offer extended service life, lower emissions, and improved energy performance [15]. Although the initial investment cost is relatively high, a thorough techno-economic analysis can justify the expenditure by evaluating factors such as lighting quality, illuminance level (lux), and visual comfort [16].

Air conditioning systems account for a substantial share of energy consumption in buildings, and their optimization requires enhancing the coefficient of performance (COP) [17]. These systems contribute more than 34% of total energy demand and approximately 37% of CO_2 emissions [18], underscoring the need for decarbonization through reduced reliance on fossil fuels and hydrofluorocarbons (HFCs) [19].

The International Organization for Standardization ISO 50001:2018 standard, specifically Clause 6, establishes a framework for the implementation and continuous improvement of Energy Management Systems (EnMS), enabling reductions in energy consumption, operating costs, and greenhouse gas emissions, as well as the optimization of overall energy performance [20]. In this context, oilfield camps depend primarily on diesel and natural gas: the former entails high costs and considerable environmental impact, whereas the latter, although cleaner, is constrained by logistical challenges. This energy profile reflects the broader balance observed in building energy systems. Figure 1 illustrates the fundamental processes common to both air-conditioning systems and lighting installations.

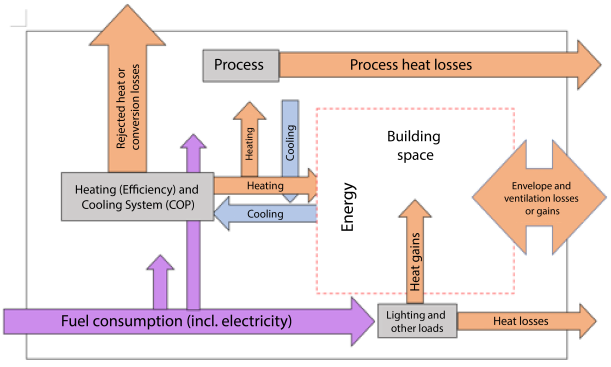


Figure 1. Energy balance for building systems

Energy use, encompassing both fuels and electricity, depends on the primary energy source and the performance of the heating and cooling system, which is typically quantified through the coefficient of performance (COP). A considerable share of this energy

is ultimately dissipated as waste heat.

Thermal energy is transferred indoors, where it interacts with factors such as solar radiation, artificial lighting, and the operation of electrical equipment, resulting in internal heat gains. Conversely, thermal losses occur through the building envelope and ventilation, thereby increasing overall energy demand.

Industrial processes also contribute heat to the indoor environment; however, a portion of this energy is lost to the exterior, thereby reducing overall system efficiency. Similarly, heat losses through the roof and ventilation increase the building's thermal load and, consequently, its energy consumption.

In this context, optimizing energy consumption through advanced energy management technologies constitutes a key strategy for mitigating the environmental impacts associated with the use of non-renewable energy sources, particularly in remote operational camps [21]. This study examines the implementation of the RETScreen platform as a tool for enhancing energy efficiency and integrating renewable energy systems.

This research focuses on maximizing energy-use efficiency in oilfield camps through a comprehensive assessment of current consumption patterns, feasibility analysis using RETScreen, and the formulation of renewable energy solutions and short-term operational measures designed to enhance performance and reduce overall energy demand.

2. Materials and Methods

This study employs a descriptive–experimental design with a quantitative approach. Historical records of energy consumption for the 2021–2022 period were analyzed, encompassing 24 measurements collected at a crude oil transportation facility. The analysis compares a baseline scenario with an optimized case to assess the impact of energy efficiency measures.

The RETScreen Expert platform is employed to simulate operating conditions, model subsystems, and evaluate the feasibility of different energy efficiency measures. The analysis focuses on lighting, general services, and air-conditioning systems. Figure 2 illustrates the energy flow and the supply sources considered in the study.

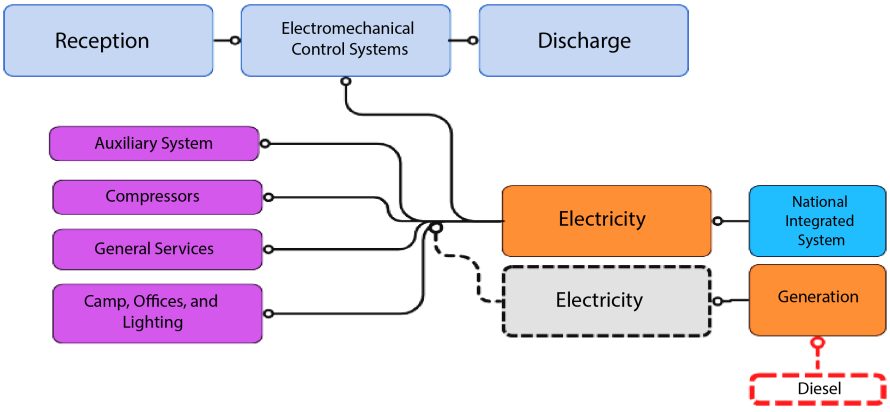


Figure 2. Schematic representation of electrical energy flow in the camp areas

The energy is supplied by the National Interconnected System (SNI), which feeds the fire protection systems, compressors, dining hall, offices, and camp lighting. The entire process is regulated by an electromechanical control system equipped with valves. As a backup, diesel generators are connected to the same network to provide power in the event of failures, ensuring operational continuity.

The energy performance assessment is structured around three methodological axes:

- Collection and analysis of operational data.
- Identification of improvement opportunities and strategic planning.
- Implementation and monitoring of results.

For the initial stage of model configuration, data on the camp's average energy consumption were entered, with particular emphasis on cooking, air conditioning, lighting, fluid compression systems, and energy-related costs. Additionally, the site's geographical characteristics, including latitude, longitude, climate zone, and terrain attributes, were incorporated into the analysis usingthe following georeferenced data:

- N 0° 58' 22.268", O 79° 40' 51.536"
- 0.9728521511737187, –79.68098222443405

To strengthen the proposal for incorporating renewable energy systems, environmental variables, includ-

ing air temperature, relative humidity, precipitation, atmospheric pressure, and soil temperature, were incorporated into the model. These parameters were integrated into the analysis to assess the thermal and energy performance of both conventional and renewable systems under the site's specific environmental conditions, as shown in Figure 3.

	Heating design temperature	19,6							
	Cooling design temperature	29,0							
	Soil temperature amplitude	9,7							
Month	Air temperature	Relative humidity	Precipitation	Global solar radiation	Atmospheric pressure	Wind speed	Soil temperature	Heating degree days	Cooling degree days
	°C	%	mm	kWh/m²/d	kPa	m/s	°C	°C-d	°C-d
January	24,9	83,7%	191,89	4,24	99,9	3,4	25,7	0	462
February	25,0	85,9%	231,28	4,43	99,9	2,8	25,6	0	420
March	25,1	86,5%	224,44	5,00	99,8	2,5	25,7	0	468
April	25,2	86,9%	244,20	4,80	99,8	2,8	25,8	0	456
May	25,0	86,1%	146,63	4,22	99,9	3,5	25,8	0	465
June	24,6	84,6%	66,60	3,73	99,9	3,8	25,5	0	438
July	24,5	81,5%	36,27	3,87	100,0	4,0	25,6	0	450
August	24,8	77,9%	26,97	4,06	100,0	4,2	26,1	0	459
September	24,9	76,4%	42,30	4,21	100,0	4,4	26,4	0	447
October	25,0	76,2%	44,95	3,98	100,0	4,5	26,4	0	465
November	24,9	75,9%	42,30	3,81	99,9	4,4	26,3	0	447
December	24,9	78,9%	88,04	4,07	99,9	4,1	26,1	0	462
Annual	24,9	81,7%	1.385,87	4,20	99,9	3,7	25,9	0	5.438

Figure 3. Monthly meteorological parameters for the proposed geolocation in 2025, based on data from NASA.

Through georeferencing, a graph of meteorological data representing monthly solar radiation and air temperature over a one-year period was generated, as shown in Figure 4.

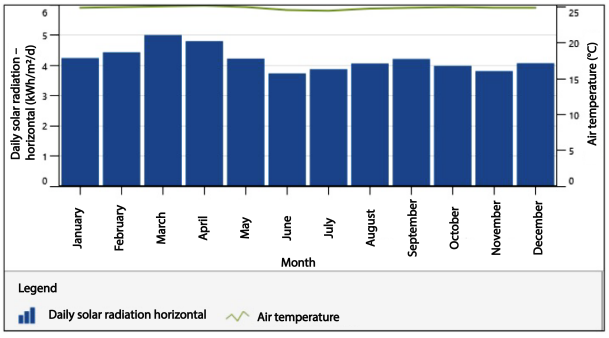


Figure 4. Comparison of monthly solar radiation and air temperature in 2025, based on data from NASA.

Additionally, electrical consumption data for the 2021–2022 period were collected. The analysis performed in RETScreen implemented demand management, energy storage, and optimization strategies in accordance with the modules structured within the software, aiming to reduce energy losses and improve cost efficiency. As a baseline reference, Figure 5 illustrates the comparison of monthly energy consumption (kWh) between2021 and 2022.

In 2021, total energy consumption reached 271,148 kWh (51.86%), whereas in 2022 it decreased to 251,622 kWh (48.04%). This reduction is attributed to social mobilizations that temporarily limited the operation of the camps.

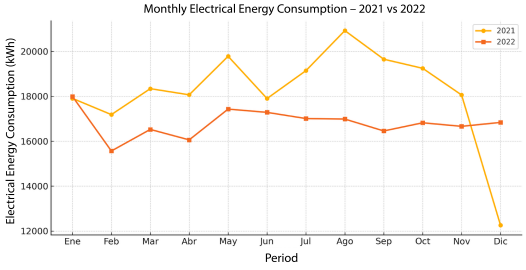


Figure 5. Comparison of the camp's total energy consumption between 2021 and 2022

Regarding fossil fuel consumption, records are available only for January–April 2021 and January–March 2022, as shown in Figure 6. These datasets enabled the analysis of the thermal impact associated with autonomous electricity generation, as well as its correlation with the efficiency of the evaluated energy system.

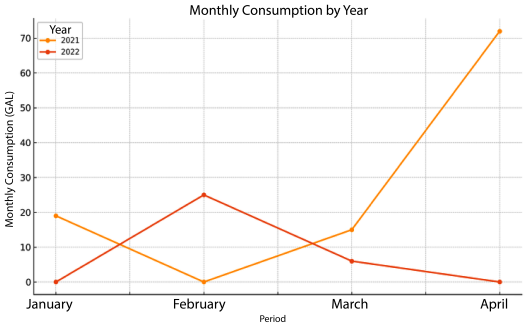


Figure 6. Comparison of fuel consumption during the 2021–2022 period.

Fuel consumption reflects fluctuations in the oil market, which influence costs during the analyzed period and constitute a key parameter for estimating the expenditures associated with fuel use.

Table 1 presents the unit fuel prices and their variation over time.

Table 1. Fuel consumption and cost under the baseline scenario

Year	Period	Monthly consumption (GAL)	Unit fuel cost combustible (USD)	Average unit cost (USD/GAL)
2021	January	19	1.58	12
	February	0	1.75	0
	March	15	1.94	8
	April	72	2.18	33
2022	January	0	2.47	0
	February	25	2.86	9
	March	6	3.39	2
	April	0	3.91	0

The reduction in SNI consumption was offset by fossil fuel generation. The camp's electrical data were entered into the software to establish the baseline scenario, considering generator systems, pumps, compressors, lighting, air-conditioning units, and other auxiliary equipment.

The energy demand (ED) was calculated as power × hours/day × days/year, adjusted for system efficiency, as expressed in Equation (1).

$$DE(kWh/year) = Power(kW)\cdot hours/day\cdot hours/year \quad (1)$$

The energy recording began with the loads associated with the camp's pumping system, as detailed in Table 2.

Table 2. Recorded data and energy performance analysis of the pumping systems in the camp

	Engine	Pump				Energy demand RETSCREEN
	Power consumption	***Nominal efficiency***	***Operating efficiency***	***Fluid load***	***Operating hours***	
	(kW)	***(%)***	***(%)***	***(HP)***	***(h/año)***	**(kWh/año)**
Evacuation pump	4.23	89.9	88.9	3.5	13	55
Diesel transfer pump	2.39	85.3	84.4	1.8	1	2.4
Recirculation pump for cleaning 1	0.96	79	78.2	0.69	26	25
Hydrocarbon separation pump	1.32	84	83.1	1	6	8
Recirculation pump for cleaning 2	1.81	84	83.1	1.4	208	377
Oxygenation pump	1.20	84.4	83.5	0.92	728	875
Storage pumps	0.62	82.2	81.3	0.23	728	449
		TOTAL				**1792**

The annual energy consumption was 1,792 kWh, calculated based on operating hours and pump power. Historical records reported a value of 1,790.89 kWh, thereby confirming the reliability of the established energy baseline.

In the RETScreen simulation, an air system with a screw compressor was modeled and configured as compressed air equipment within the software, as detailed in Table 3.

Table 3. Recorded data for electrical equipment

Fluid Compressor Equipment	
Parameter	**Value**
Compressor capacity, [ft^3/min]	20
System pressure, [PSIG]	165
Friction losses, [%]	10
Motor capacity, [kW]	7.4
Effective air demand, [ft^3/min]	10.5
Operating time, [h/year]	1196
Total electrical demand	**5315**

After recording the fluid compressor data, the electric generator was incorporated into the integrated energy system, whose technical characteristics are detailed in Table 4.

Table 4. Recorded data for power generation equipment

Power Generator Equipment	
Parameter	**Value**
Fuel	Diesel
Generated electrical energy, [kWh/year]	1370
Power, [kW]	110
Annual fuel consumption, [gal/year]	106.18

After recording the compressor, additional electrical loads were incorporated, including the water heating, refrigeration, kitchen, and computer systems, to complete the comprehensive modeling of electrical demand. This process was systematized by considering the connected load (kW), annual operating hours, and total system demand (kWh), in accordance with the guidelines of the RETScreen Expert electrical end-use module, see table 5.

The lighting system modeling was conducted by entering data according to the camp's functional areas, following the same procedure applied to the general electrical loads. Because these components have a significant energy impact on overall energy consumption and exhibit substantial variability, they were parameterized in detail in RETScreen. Table 6 presents the total electrical demand associated with this category.

Table 5. Recorded data for electrical equipment in the kitchen area

	Quantity	**Operating hours (h/year)**	**Electrical load (W)**	**Utilization cycle (%)**
Water heater 1	1	8736	5000	100
Water heater 2	1	8736	3000	100
Refrigerator	3	8736	1000	100
Stove	1	2184	10 000	100
UPS room	1	8736	2880	100
Total electricity demand (kWh)		**143 096**		

Table 6. Electrical energy demand of the camp's luminaires, as provided by RETScreen

	Quantity	**Operating hours h/year**	**Electrical load (W)**	**Operating cycle (%)**
Industrial area	22	4368	350	100
Generator technical shelter	8	182	400	100
Security post	2	8736	40	100
DDV modular unit	2	3640	40	100
Operational monitoring room	2	4368	40	100
Automation room	6	8736	40	100
Control area restroom	1	3.64	58	100
Control storage	1	3.64	18	100
Operational modular unit	2	3640	40	100
Mechanical maintenance point	8	182	5	100
Electrical maintenance point	3	182	305	100
Seramin technical module	7	1820	18	100
Gym	2	1092	40	100
Tool station	3	364	5	100
Rooms	1	1820	305	100
Logistics warehouse	1	182	18	100
Camp common room	7	1820	40	100
Dining room	1	1820	40	100
Kitchen	1	6552	250	100
Restrooms	4	364	18	100
Total electricity demand (kWh/year)		**41151**		

For the registration of air conditioning systems, an average coefficient of performance (COP) of 3 was considered, corresponding to the three air conditioning units. These systems were subsequently integrated into the RETScreen platform, assuming a total cooling thermal load of 93,500 BTU/h, a utilization cycle of 100%, and 8,736 annual operating hours. The analysis determined an annual energy demand of 79,795 kWh, corresponding to the expected consumption of the air conditioning subsystem.

3. Results and Discussion

Table 7 summarizes the energy balance, indicating a total fuel consumption of 113 gal/year (equivalent to 4,557 kWh/year) and a total energy consumption of 271,148 kWh/year. The associated annual costs are USD 339 for diesel and USD 27,115 for electricity, assuming an average diesel price of USD 3/gal and an electricity rate of USD 0.10/kWh.

In this case, it was determined that reducing energy consumption in the oil transportation camp requires the implementation of a photovoltaic system designed to achieve savings of 15–30% of the camp's total energy demand. Accordingly, a fixed system was selected, consisting of 50 panels with an individual capacity of 600 W, yielding an approximate total power of 30 kW after accounting for panel and inverter losses. The remaining characteristics are presented in Table 8.

Table 7. Annual fuel and electricity consumption and associated costs in the baseline and proposed cases

Energy type		**Baseline case**		**Proposed case**	
Fuel type	**Fuel price (USD)**	**Fuel consumption**	**Fuel cost (USD)**	**Fuel consumption**	**Fuel cost**
Diesel	3	113 gal	339	113 gal	USD 339
Electricity	0.10	271 148 kWh	27 115	185 878 kWh	$ 18 588
Total			**27,454**		**$ 18 927**

Table 8. Components and key indicators of the photovoltaic system analyzed in RETScreen

Solar tracking mode	Fixed
Tilt angle	15°
Azimuth	180
Panel type	Mono - Si
Manufacturer	Jinico Solar
Model	Mono Si - JKM600N-78HL4-600W
Rated capacity	600W per panel
Number of units	50
Total capacity	30 kW
Nominal operating cell temperature	45° C
Efficiency	21,46%
Solar collector area	140 m^2
Miscellaneous losses	10%
Inverter efficiency	97% (as recommended by the software)
Inverter losses	1%

The premise established in this analysis is that, to maintain a constant consumption of the energy generated by the photovoltaic system, the tilt angle must be set at 15°. Regarding the azimuth orientation, it should be aligned according to the corresponding hemisphere. In this case, since Esmeraldas (Ecuador) is located in the Northern Hemisphere, the optimal orientation is toward the south, corresponding to an azimuth angle of 180°.

The premise also establishes that the initial, operating, and maintenance costs, along with the camp's energy savings, are summarized in Table 9.

Table 9. Installation and operation–maintenance costs and annual energy savings of the photovoltaic system

	Cost per kW (USD/kW-year)	**Total cost (USD)**
Initial cost	1176	35 294
Operation and maintenance cost	10.29	309
Energy saved		**36 103 kWh**

It should be noted that the components were selected based on the initial premise of reducing energy dependence on both the National Interconnected System (SNI) and conventional generation sources, such

as diesel generators, by 15–30%. Consequently, all parameters were automatically calculated by RETScreen, which provided recommendations regarding suppliers, models, capacity, efficiency, and losses, as well as economic indicators such as the panel cost per kilowattand the annual operation and maintenance cost.

3.1. Energy Consumption Optimization Measures

Focusing on the energy savings within the camp's energy system, a direct relationship was established between the input data and the incorporation of the photovoltaic system as a key element. Through this integration, the primary objective of the study was achieved, reducing overall energy consumption by 15–30% through the application of optimization principles. These results are summarizedin Table 10.

Table 11 indicates that, in the baseline case, 113 gallons of diesel (USD 339) and 271,148 kWh/year of electricity (USD 27,454) were consumed. In the proposed case, diesel consumption remains unchanged, while electricity consumption decreases to 185,878 kWh/year (USD 18,588), yielding energy savings of 85,270 kWh (31%) as a result of the photovoltaic system implementation. The distribution of savings by subsystem (heating, cooling, and electricity) is detailed in Table 11.

Table 10. Electricity and fuel savings in the camp

Fuel		Baseline case		Proposed case		Savings	
Energy type	**Price of fuel (USD)**	**Fuel consumption**	**Fuel cost (USD)**	**Fuel consumption**	**Fuel cost (USD)**	**Fuel saved**	**Amount saved (USD)**
Diesel	3	113 gal	339	113 gal	339	0 gal	0
Electricity	0.10	271 148 kWh	27 115	185 878 kWh	18 588	85 270 kWh	8527
Total		**27 454**			**18 927**		**8527**

Table 11. Energy and percentage savings for each subsystem

	Heating (kWh)	**Cooling (kW)**	**Electricity (kWh)**	**Total (kWh)**
Baseline case	4557	79 795	191 353	275 705
Proposed case	4557	68 396	117 482	190 435
Energy saved	**0 kWh**	**11 399 399 kW**	**73 871 kWh**	**85 270 kWh**
Percentage saved (%)	***0%***	***14.3%***	***38.6%***	***30.9%***

With the incorporation of the photovoltaic system, the cooling subsystem achieves 14.3% energy savings, while the electrical subsystem (pumps, compressors, and lighting) achieves 38.6%, resulting in an overall average reduction of 30.9%. By eliminating dependence on the SNI and diesel generators, the economic benefit is substantial, with an estimated return on investment of 10.7 years. Figure 7 compares the results forboth cases.

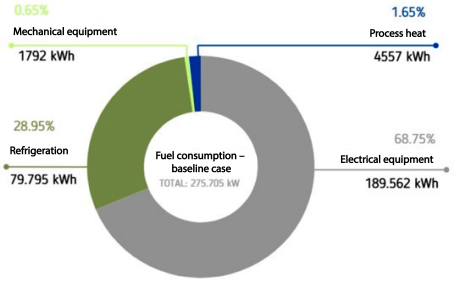


Figure 7. Baseline energy consumption by fuel type for the different equipment categories

Electrical equipment, such as generators, accounts for the largest share of fuel consumption in the camp facilities, followed by refrigeration and air-conditioning systems. Mechanical equipment represents a smaller fraction, while process heat corresponds to thermal losses associated with combustion. The proposed case is illustratedin Figure 8.

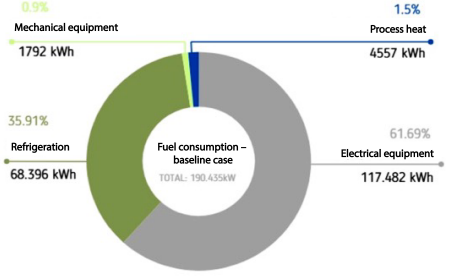


Figure 8. Proposed energy consumption by fuel type for the different equipment categories

Table 12. Energy consumption in the proposed case, as modeled in RETScreen

Proposed case energy consumption		
Type of equipment	**Energy consumption kWh**	**Percentage value**
Electrical equipment	146 479 kWh	64.7%
Refrigeration	68 396 kWh	30.2%
Mechanical equipment	7106 kWh	3.1%
Process heat	4557 kWh	2%
Photovoltaic electricity generation	–36 103 kWh	-

A significant reduction in energy consumption is evident with the proposed implementation of the photovoltaic system, contributing to an annual decrease of 36,103 kWh, as summarizedin Table 12.

3.2. Associated Costs

Regarding the costs associated with implementing the photovoltaic system, an increase in the initial investment is observed; however, a significant reduction in fuel expenses is achieved, representing a clear economic benefit. Although the return on investment occurs in the long term, the project offers advantages in the optimization of sustainable and environmentally friendly energy systems, as detailedin Table 13.

Table 13. Expenses considered for implementation of the photovoltaic system

Costs associated with implementation	
Incremental initial costs [USD]	35 294
Fuel cost savings [USD]	8527
Incremental operation and maintenance savings [USD]	–309
Payback period (years)	10.7

The initial investment is relatively high, and the operation and maintenance costs present a negative value, indicating that they should be considered part of the expenses associated with the operation of the photovoltaic system.

3.3. Environmental Analysis

In the environmental impact analysis, parameters such as greenhouse gas (GHG) emission factors were considered, using a value of 0.213 tCO_2/MWh as defined by the program for Ecuador. As shown in Figure 9, GHG emissions correspond to a gross annual reduction of approximately 18.1 tCO_2, equivalent to the carbon sequestration capacity of about11.7 hectares of forest.

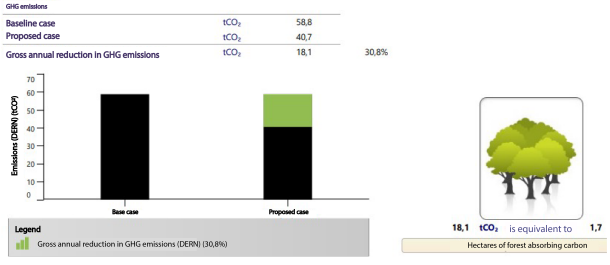


Figure 9. Gross annual reduction in GHG emissions

In the baseline case, emissions totaled 58.8 tCO_2, while in the proposed case they were reduced to 40.7 tCO_2.

In addition to the analysis of the photovoltaic system implementation, several immediately applicable measures are proposed. These alternatives are suggested because the system's implementation cost may represent a limitation if a comprehensive preliminary assessment is not conducted. Accordingly, optimization options were identified and evaluated using the RETScreen platform.

3.4. Energy Management Measures in the Oilfield Camp

Kitchen Area

A baseline scenario and an improvement proposal were developed for the water heating system (8,000 L/day), based on historical records indicating that an electric generator served as the energy source for heat production.

A 5 °C reduction in temperature was proposed for the hot water system, lowering the setpoint from 45.5 °C to 40.5 °C. From a thermodynamic perspective, this decrease implies that less energy is required to reach the target temperature. The decision to limit the reduction to 5 °C, rather than 10 °C, is based on maintaining a balance between energy efficiency and user thermal comfort. In practical terms, a 5 °C decrease ensures that the energy required for water heating remains directly proportional to the temperature difference, without compromising user comfort or system performance.

Based on actual energy consumption data, the estimated annual savings are approximately 17,000 kWh, representing a 24.3% improvement in thermal performance. In the baseline case, consumption was 70,180 kWh/year, whereas in the proposed scenario, under the suggested operating conditions, it decreased to 53,147 kWh/year. This reduction results from both the temperature adjustment and the optimization of the usage cycle.

As a complementary measure, a 25% reduction in the utilization cycle is proposed, which would decrease the overall thermal demand and, in turn, lower the frequency of preventive and corrective maintenance.

For future research, a heat recovery analysis could be conducted on extraction systems through the implementation of extractor hoods designed to preheat water or supply secondary thermal processes within the camp.

Camp Lighting Fixtures

To optimize the lighting systems across the different areas of the camp, the replacement of obsolete technologies, such as fluorescent lamps with electronic ballasts, with LED luminaires is proposed. In this analysis, total system losses of 25% were considered for both the baseline and the proposed cases.

From a technical perspective, a 50% reduction in electrical demand was observed, decreasing from 41,151

kWh/year to 20,855 kWh/year, thereby improving the energy efficiency of the lighting systems. This improvement is attributed to the higher efficacy of the luminaires: while conventional lights with electronic ballasts operate at 70 lm/W, LED luminaires reach 85.5 lm/W, allowing the same illuminance levels (500 lux) to be maintained with a lower connected load.

It is important to note that lighting systems must be adapted to an appropriate optical design, considering both directional and light dispersion characteristics. LED luminaires, which emit light directionally, minimize dispersion losses and help maintain the required illuminance level (500 lux) with fewer fixtures and a lower installed load. In contrast, fluorescent lamps with ballasts emit light omnidirectionally, requiring a greater number of fixtures. Consequently, LED technology reduces unit power demand from 150 W to 75 W per lamp, decreasing the total connected load from 4,688 W (T8 fluorescents) to 2,344 W (LEDs).

For future research, it is recommended to conduct a more comprehensive analysis of lighting efficiency in critical areas, incorporating illuminance levels evaluated through specialized software such as DIALux. Furthermore, a comparative assessment between conventional LED luminaires and those equipped with presence sensors or automatic dimming is suggested to determine their additional energy-saving potential.

Air Conditioning Systems

The air conditioning system, composed of three units with a total capacity of 93,500 BTU/h, offers a significant opportunity to enhance efficiency by increasing the coefficient of performance (COP) from 3.0 to 3.5. From a thermodynamic standpoint, the COP expresses the ratio of useful cooling energy delivered to the energy consumed; thus, increasing the COP reduces overall energy consumption while maintaining the same cooling capacity.

With a COP of 3.5, the energy demand of the air conditioning system decreases from 79,795 kWh/year to 68,396 kWh/year, representing a 14.28% reduction. This behavior aligns with the inverse proportional relationship between energy consumption and the COP.

To achieve this COP value, the following strategies are proposed:

- Periodic preventive and corrective maintenance, including coil cleaning, verification of high and low pressures, and leak detection when applicable.
- Improvements in the airtightness of air-conditioned areas to reduce the thermal load and, consequently, the system's operational demand.
- Implementation of smart controls, such as programmable thermostats and occupancy sensors, to optimize operating cycles.
- Integration of renewable energy technologies (e.g., aerothermal systems or solar thermal support for domestic hot water) to reduce compressor workload when applicable.

A 30 kWp photovoltaic system comprising 50 panels of 600 Wp each is proposed, achieving an estimated 30% reduction in displaceable electrical consumption. The system can be expanded through additional modules or complemented by replacing lighting fixtures with LED luminaires in critical areas. In addition, regular maintenance of the air conditioning system or replacement with inverter-type units using R410 refrigerant is recommended.

In the future, improving the distribution of the photovoltaic load by prioritizing critical equipment and incorporating protective devices (e.g., remotely controlled load switches and, when feasible, energy storage systems) will enhance overall system resilience.

For future research, it is recommended to evaluate the operating cycles of air conditioning and extraction systems by establishing real operating parameters, such as periods of peak thermal load and opportunities to reduce operation without compromising comfort, and to assess the compatibility of inverter-based equipment with low-GWP (Global Warming Potential) refrigerants.

4. Conclusions

The central focus of this study was to evaluate the technical, economic, and environmental impacts of the implemented measures. Accordingly, the primary objective was to achieve a reduction of between 15% and 30% in energy consumption in a remote oilfield camp.

This analysis was conducted using the RETScreen platform, through which improvements were proposed for the air conditioning, lighting, and general service systems. Based on historical records, the initial electricity consumption was 271,148 kWh/year, with an associated cost of USD 27,454. In the proposed scenario, which incorporates a photovoltaic system to optimize energy use, consumption decreased to 185,878 kWh/year, representing a 31% reduction and annual cost savings of USD 8,527.

The photovoltaic generation system, with an installed capacity of 30 kWp (not 30 kWh/year), reduced dependence on the National Interconnected System (SNI) and on polluting energy sources.

Regarding energy efficiency distribution, the air conditioning systems achieved a 14.28% improvement, while the electrical systems, including pumps and lighting, registered the highest reduction, at 38.6%. The

overall assessment indicates an average saving of 30.9% across the entire system.

From an environmental perspective, CO_2 emissions decreased from 58.8 tCO_2 to 40.7 tCO_2, equivalent to the carbon sequestration capacity of approximately 1.7 hectares of forest. The estimated return on investment of 10.7 years reinforces the proposal's technical, economic, and environmental viability as a sustainable and replicable solution.

Contributor Roles

- **Edwin Illescas:** Conceptualization, formal analysis, writing – original draft, writing – review & editing, investigation, resources.
- **Edison Laz:** Software, writing – original draft, writing – review & editing, validation, resources.
- **Manuel Rogelio Nevarez Toledo:** Project administration.
- **Miguel Alberto Dávila-Sacoto:** Supervision, project administration, methodology.

References

[1] E. A. Llanes Cedeño and O. D. Rubio Aguiar, *Elaboración de una metodología de gestión energética para campos petroleros en el Ecuador basada en la ISO 50001.* Universidad Internacional SEK, 2022. [Online]. Available: https://upsalesiana.ec/ing35ar1r1

[2] B. Williams, D. Bishop, P. Gallardo, and J. G. Chase, "Demand side management in industrial, commercial, and residential sectors: A review of constraints and considerations," *Energies*, vol. 16, no. 13, p. 5155, Jul. 2023. [Online]. Available: http://doi.org/10.3390/en16135155

[3] S. Mimi, Y. Ben Maissa, and A. Tamtaoui, "Optimization approaches for demand-side management in the smart grid: A systematic mapping study," *Smart Cities*, vol. 6, no. 4, pp. 1630–1662, Jun. 2023. [Online]. Available: http://doi.org/10.3390/smartcities6040077

[4] S. S. Shuvo and Y. Yilmaz, "Demand-side and utility-side management techniques for increasing EV charging load," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3889–3898, Sep. 2023. [Online]. Available: http://doi.org/10.1109/TSG.2023.3235903

[5] S. Charles Raja, A. C. Vishnu Dharssini, J. Jeslin Drusila Nesmalar, and T. Karthick, "Deployment of IoT-based smart demand-side management system with an enhanced degree of user comfort at an educational institution," *Energies*, vol. 16, no. 3, p. 1403, Jan. 2023. [Online]. Available: http://doi.org/10.3390/en16031403

[6] V. Cortez, R. Rabelo, A. Carvalho, A. Floris, and V. Pilloni, "On the impact of flexibility on demand-side management: Understanding the need for consumer-oriented demand response programs," *International Journal of Energy Research*, vol. 2024, no. 1, Jan. 2024. [Online]. Available: http://doi.org/10.1155/2024/8831617

[7] D. Stanelyte, N. Radziukyniene, and V. Radziukynas, "Overview of demand-response services: A review," *Energies*, vol. 15, no. 5, p. 1659, Feb. 2022. [Online]. Available: http://doi.org/10.3390/en15051659

[8] M. Alikhani, M. P. Moghaddam, F. Moazzen, and A. Azadi, "Optimal implementation of consumer demand response program with consideration of uncertain generation in a microgrid," *Electric Power Systems Research*, vol. 225, p. 109859, Dec. 2023. [Online]. Available: http://doi.org/10.1016/j.epsr.2023.109859

[9] A. Piñeres Castillo, J. J. Cabello Eras, and M. Hinojosa Rivera, "Factores determinantes para la evaluación de la eficiencia energética en las organizaciones: una visión desde las condiciones de Colombia," *Universidad y Sociedad*, vol. 14, no. 2, pp. 509–520, 2022. [Online]. Available: https://upsalesiana.ec/ing35ar1r9

[10] P. J. Ramos-Males and A. M. Bautista-Segovia, "La eficiencia energética: Una estrategia para la economía doméstica en Ecuador," *Dominio de las Ciencias*, vol. 8, no. 2, pp. 1334–1346, 2022. [Online]. Available: https://upsalesiana.ec/ing35ar1r10

[11] R. Cancio-Bello, S. Montelier Hernández, A. Oviedo Regojo, and O. Bello Gonzalez, "Bases para la implementación de un sistema de gestión energética en la UEB Ron "Luis Arcos Bergnes" de Cienfuegos basado en la NC-ISO 50001:2019," *Universidad y Sociedad*, vol. 13, no. 4, 2021. [Online]. Available: https://upsalesiana.ec/ing35ar1r11

[12] L. Freire, V. Resabala, J. Castillo, and B. Corrales, "Propuesta de un plan alternativo de optimización energética," *Espacios*, vol. 40, no. 30, 2019. [Online]. Available: https://upsalesiana.ec/ing35ar1r12

[13] F. De La Cruz Cabrales Ávila and R. A. Mejía Otero, "Sistemas fotovoltaicos: solución energética en las comunidades aisladas," 2022. [Online]. Available: http://doi.org/10.5281/ZENODO.7461726

[14] G. S. Inca Yajamín, D. F. Cabrera Carrión, D. F. Villalta Gualán, R. C. Bautista Zurita, and H. D. Cabrera Carrión, "Evaluación de la actualidad de los sistemas fotovoltaicos en ecuador: avances, desafíos y perspectivas," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, no. 3, pp. 9493–9509, Jul. 2023. [Online]. Available: http://doi.org/10.37811/cl_rcm.v7i3.6835

[15] A. F. Guamán Mejía and L. V. Sánchez Parrales, "Importancia de la energía renovable para el alumbrado público," *Revista Sinapsis*, vol. 24, no. 1, 2024. [Online]. Available: https://upsalesiana.ec/ing35ar1r15

[16] J. R. López Escobedo, K. Nicolás Javier, and F. García Reyes, "Optimización de energía eléctrica y luminosa en laboratorios y aulas de aprendizaje, aplicación sustentable con luminarias de tecnología LED," *Tendencias en energías renovables y sustentabilidad*, vol. 1, no. 1, pp. 93–100, Dec. 2022. [Online]. Available: http://doi.org/10.56845/terys.v1i1.108

[17] J. L. Rodríguez Muñoz, J. S. Pacheco Cedeño, C. M. Valencia Castillo, and J. d. J. Ramírez Minguela, "Energy and exergy analysis of an ejector-compression refrigeration cycle with double ihx," *REVISTA DE CIENCIAS TECNOLÓGICAS*, vol. 6, no. 3, p. e261, Sep. 2023. [Online]. Available: http://doi.org/10.37636/recit.v6n3e261

[18] ONU, *Las emisiones históricas del sector de la construcción, lo alejan de los objetivos de descarbonización.* Naciones Unidas, 2022. [Online]. Available: https://upsalesiana.ec/ing35ar1r18

[19] IRENA, *Integración de energías renovables de baja temperatura en los sistemas energéticos urbanos: lineamientos para los responsables de la formulación de políticas – Resumen.* International Renewable Energy Agency, 2021. [Online]. Available: https://upsalesiana.ec/ing35ar1r19

[20] ISO, *Guía de implementación e interpretación de requisitos del estándar ISO 50001:2018.* Organización Internacional de Normalización, 2018. [Online]. Available: https://upsalesiana.ec/ing35ar1r20

[21] T. Jing and Y. Zhao, "Optimizing energy consumption in smart buildings: A model for efficient energy management and renewable integration," *Energy and Buildings*, vol. 323, p. 114754, Nov. 2024. [Online]. Available: http://doi.org/10.1016/j.enbuild.2024.114754

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.02
pISSN: 1390-650X / eISSN: 1390-860X

Classification of upper limb fractures using deep learning

Clasificación de fracturas en extremidades superiores con aprendizaje profundo

Gabriela Jaén-Armijos[1,*], Evelyn Morán-Castillo[1], Wilmer Rivas-Asanza[1], Eduardo Tusa[1]

Received: 30-04-2025, Received after review: 19-06-2025, Accepted: 09-09-2025, Published: 01-01-2026

Abstract

Accurate identification of upper extremity fractures is essential for timely and reliable diagnosis in emergency medical settings. This study evaluates and compares the performance of three pre-trained deep learning architectures: EfficientNet- B4, ResNet-50, and ConvNeXt-Large, applied to the automatic classification of bone fractures in radiographic images from the MURA repository, encompassing seven anatomical regions. Advanced image preprocessing techniques, including Unsharp Masking and Contrast-Limited Adaptive Histogram Equalization (CLAHE), were employed in conjunction with data normalization and balancing strategies. The models were trained in two experimental setups: a binary classification distinguishing between "fracture" and "non-fracture" images, and a multiclass configuration identifying 14 distinct fracture types. Performance evaluation using F1-Score, sensitivity, accuracy, and ROC–AUC metrics demonstrated that ConvNeXt-Large achieved the highest overall results, reaching accuracies of 99.0% in binary classification and 99.4% in multiclass classification. These findings position ConvNeXt-Large as a highly promising tool for supporting early and precise fracture diagnosis.

Keywords: Human extremities, Bone Fractures, Artificial Intelligence, X-ray, Artificial neural networks.

Resumen

La identificación precisa de fracturas en las extremidades superiores es fundamental para un diagnóstico oportuno en los entornos de urgencias médicas. Este estudio analiza y compara el desempeño de tres modelos de aprendizaje profundo preentrenados: EfficientNet-B4, ResNet-50 y ConvNeXt-Large, aplicados a la clasificación automática de fracturas óseas en radiografías del repositorio MURA, distribuidas en siete regiones anatómicas. Se aplicaron técnicas avanzadas de preprocesamiento digital, como Unsharp Masking y CLAHE, junto con estrategias de normalización y balanceo de datos. El entrenamiento se llevó a cabo en dos experimentos: uno binario, que clasifica entre "fractura" y "no fractura", y otro multiclase, con catorce tipos de fracturas identificadas. La evaluación, mediante métricas como F1-Score, sensibilidad, exactitud y curvas ROC-AUC, reveló que ConvNeXt-Large alcanzó el mejor rendimiento, logrando una precisión del 99,0% en clasificación binaria y del 99,4% en la clasificación multiclase. Estos resultados posicionan a ConvNeXt-Large como una herramienta altamente prometedora para apoyar el diagnóstico temprano de fracturas óseas.

Palabras clave: extremidades humanas, fracturas óseas, inteligencia artificial, radiografía, redes neuronales artificiales

[1,*]Facultad de Ingeniería Civil, AutoMathFIC, Universidad Técnica de Machala, Ecuador.
Corresponding author ✉: gjaen1@utmachala.edu.ec.

Suggested citation: G. Jaén-Armijos, E. Morán-Castillo, W. Rivas-Asanza and E. Tusa. "Classification of upper limb fractures using deep learning," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 21-35, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.02.

1. Introduction

Bone fractures are among the most common injuries treated in hospital emergency departments [1]. The upper extremities play essential roles in human mobility and functionality; therefore, fractures involving regions such as the elbow, fingers, forearm, hand, humerus, shoulder, or wrist can significantly impair fine and gross motor skills. Such injuries limit the ability to perform daily activities and consequently diminish patients' quality of life.

At the global level, it is estimated that 10% to 40% of medical emergencies are attributable to upper limb injuries, corresponding to approximately six million cases annually in the United States and 700,000 in France [2].

In Medellín, Colombia, 25,646 cases of upper limb trauma were reported, of which 17% involved fractures [2]. These data emphasize the urgent need for faster and more accurate diagnostic procedures, as the error rate in fracture identification remains approximately17.9% [3].

Although radiographs are widely used, their diagnostic accuracy remains limited, particularly in anatomically complex regions such as the upper extremities. Factors including improper adjustment of X-ray tube voltage and current parameters, manual interpretation of radiographic images, and excessive workload among medical personnel can significantly contribute to diagnostic errors [4].

In light of these challenges, the application of advanced artificial intelligence (AI) techniques, particularly deep learning, has emerged as a promising approach to improving diagnostic accuracy. These methods have consistently demonstrated the ability to enhance both accuracy and efficiency in medical image interpretation [5].

Artificial intelligence (AI) broadly refers to the ability of computer systems to perform tasks that typically require human intervention [6]. More specifically, deep learning, a subfield of AI, employs complex artificial neural networks to model and solve problems with remarkable capacity for abstraction and generalization [7].

An example of how these technologies are transforming medical practice is the study by Medaramatla et al. [8], which investigated fracture detection in hand bones using a hybrid YOLO-NAS system applied to X-ray images. In that study, a dataset comprising 4,736 hand bone radiographs was developed and classified into six categories, with images sourced from publicly available repositories such as Kaggle and Mendeley. To evaluate the performance of the proposed model, several algorithms were compared, including YOLO-NAS, EfficientDet, ResNet-50, InceptionV3, Vision Transformer, VGG19, and YOLOv8. The hybrid YOLO-NAS model achieved an accuracy of 97.9%, outperforming all other evaluated methods.

Similarly, Wang [9] developed a deep learning-based approach using the Keras library and the NASNetMobile model for detecting bone fractures in the upper extremities. The model was trained on the MURA dataset and achieved an accuracy of approximately 70%, demonstrating the potential of deep learning in medical imaging applications.

In another study, Guan et al. [10] used approximately 4,000 images from the MURA dataset and proposed a two-stage R-CNN-based framework for detecting fractures in arm radiographs. The model achieved an accuracy of 62%, demonstrating the potential of R-CNN architectures to support the diagnosis of upper-limb injuries.

Although the reviewed studies report promising results in the automatic detection of bone fractures, their findings reveal significant limitations that restrict practical implementation in real clinical settings. This underscores the need to develop new strategies to overcome these challenges.

For example, the model proposed by Guan et al. [10] achieved a performance below 80% due to the limited number of images used during training. Consequently, employing a larger and more diverse dataset would be an effective strategy to enhance the model's robustness and reliability in clinical practice.

Image quality represents another limitation identified in the studies reviewed. In these investigations, model accuracy was affected by image noise and insufficient preprocessing. In the case of Medaramatla et al. [8], the images were resized to a square format of 608×608 pixels (PNG), and a manual noise removal process was applied to low-quality samples. Although this cleaning procedure improved image quality and model accuracy, it lacks the scalability and robustness required to ensure consistency in large-scale clinical settings.

Based on these limitations, the following research question arises: How can accuracy and robustness to class imbalance be improved in the automatic classification of upper-limb bone fractures from radiographic images through the implementation of pre-trained deep learning models?

To address this question, pre-trained deep learning models such as EfficientNet-B4 [11], ResNet-50 [12], and ConvNeXt-Large [13] are compared and evaluated using performance metrics including F1-score, precision, accuracy, sensitivity, the receiver operating characteristic (ROC) curve, and the confusion matrix.

The primary objective of this study is to determine which of the three proposed models provides the best performance in the automatic classification of bone fractures in the upper extremities.

2. Materials and Methods

The research follows a quantitative and experimental approach focused on comparing and validating pre-trained deep learning models for the automatic classification of upper-limb bone fractures from radiographic images.

For the implementation and evaluation of the models, the CRISP-DM approach was adopted, which provides a systematic and structured framework for data analysis [14].

Figure 1 illustrates the main phases and objectives of the process.

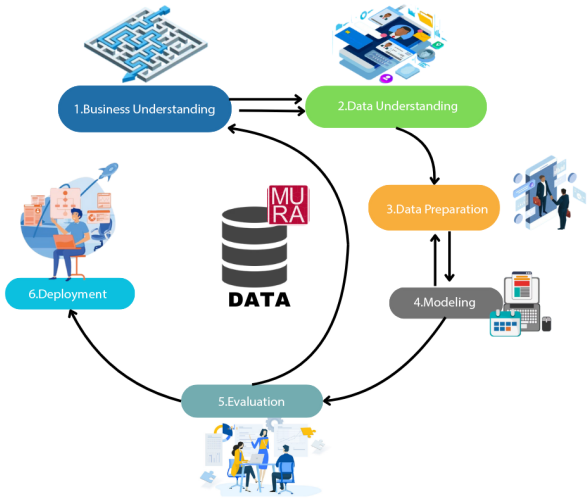


Figure 1. Study methodology consisting of six phases

These phases are described as follows:

- **Business understanding:** analysis of the context, definition of the project objectives, and identification of available resources [15].
- **Data understanding:** examination of data quality, structure, and characteristics to obtain a clear overview of its content [16].
- **Data preparation:** handling of missing values, normalization, selection of key variables, and conversion of data into suitable formats.
- **Modeling:** construction, training, and evaluation of the model to ensure its effectiveness and reliability [17].
- **Evaluation:** verification of objective achievement and assessment of model performance according to predefined criteria [18].
- **Deployment:** continuous integration, monitoring, and maintenance of the model to ensure optimal performance and periodic updates [19].

2.1. Business Understanding

The automatic classification of bone fractures in radiographs requires high accuracy, efficiency, and specialized expertise. However, the growing demand for radiological examinations, coupled with the shortage of radiology professionals, often leads to diagnostic delays and an increased likelihood of errors.

Although several computational tools have been developed, few are specifically tailored for the detection and classification of bone fractures. In this context, the present study implements and evaluates pre-trained deep learning models for the classification of bone fractures in the upper extremities.

2.2. Data Understanding

The dataset used for the automatic classification of upper-limb fractures corresponds to a large collection of musculoskeletal radiographs known as MURA [20]. This dataset comprises 40,561 radiographic images from 14,863 studies involving12,173 patients.

The images cover various anatomical regions, including the elbow, fingers, forearm, hand, humerus, shoulder, and wrist. Each study was previously labeled as normal or abnormal by certified radiologists endorsed by the Stanford Hospital board, based on diagnoses made between 2001 and 2012 [21].

In its original distribution, the MURA dataset [20] is divided into three subsets: a training set with 36,808 images (90.7%), a validation set with 3,197 images (7.9%), and a test set with 556 images (1.4%).

Table 1 summarizes the original distribution of the dataset.

Table 1. Original distribution of the MURA dataset

Category	Training		Validation	
	Normal	Abnormal	Normal	Abnormal
Elbow	2925	2006	235	230
Fingers	3138	1968	214	247
Forearm	1164	661	150	151
Hand	4059	1484	271	189
Humerus	673	599	148	140
Shoulder	4211	4168	285	278
Wrist	5765	3987	364	295
Total	21 935	14 873	1667	1530

2.3. Data Preparation

Data preprocessing is a crucial stage in data mining, as it enables the correction of noise, inconsistencies, and missing values that may compromise the model's reliability [22]. During the review of the original MURA dataset [20], 40,009 images were obtained. After a thorough analysis, four images in an incompatible format were identified and excluded in subsequent preprocessing stages, thereby confirming the total of 40,005 images reported by the authors [23].

The resulting 40,005 images were normalized and resized to 384 × 384 pixels to maintain a balance between preserving relevant details, ensuring computational efficiency, and achieving consistency with the pre-trained models.

To improve the visual quality of the radiographs, the Unsharp Masking technique, used to emphasize edges and fine details [24], and CLAHE, which improves contrast in low-quality images without introducing unwanted artifacts [25], were applied.

Previous studies have shown that combining these techniques significantly enhances visibility in low-light images [26], as illustrated in Figure 2.

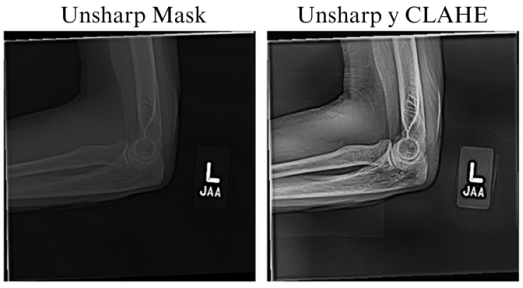


Figure 2. Application of radiographic image enhancement techniques. The image on the left shows the result obtained using Unsharp Mask, while the image on the right illustrates the effect of combining Unsharp Mask with CLAHE.

After completing the image cleaning and quality enhancement stages, the dataset was redistributed as part of preprocessing phase: 80.0% for training and 20.0% for validation, as shown in Table 2.

Table 2. Redistribution of the MURA dataset

Category	Training		Validation	
	Normal	Abnormal	Normal	Abnormal
Elbow	2559	1773	601	463
Fingers	2731	1778	621	437
Forearm	1038	664	276	148
Hand	3397	1347	933	326
Humerus	646	589	175	150
Shoulder	3591	3541	905	905
Wrist	4890	3460	1239	822
Total	18 852	13 152	4750	3251

2.4. Modeling

In this stage, three pre-trained deep learning models were implemented: EfficientNet-B4, ResNet-50, and ConvNeXt-Large. Training was conducted in a Python environment using libraries such as PyTorch, Albumentations, and Scikit-learn, across three hardware configurations: an NVIDIA RTX 3050 GPU (8 GB), an NVIDIA RTX 4060 GPU (12 GB), and an NVIDIA A100 GPU accessed through Google Colab Pro (40 GB).

These configurations enabled the evaluation of performance and resource consumption, utilizing the NVIDIA A100 GPU accessed through GoogleColab Pro for intensive testing of more complex models, such as ConvNeXt-Large.

Under these conditions, the training was structured into three phases.

1. **First phase:** a binary classification was performed to distinguish between images with fractures and those without fractures.

2. **Second phase:** a multiclass classification was performed using the MURA dataset in its original structure.

3. **Third phase:** a multiclass training was conducted with the preprocessed and resized dataset.

Table 3 summarizes the labeling scheme used for the multiclass classification.

Table 3. Class labeling for multiclass classification

Class	Abbreviation
Normal shoulder	HoN
Abnormal shoulder	HoA
Normal elbow	CoN
Abnormal elbow	CoA
Normal finger	DeN
Abnormal finger	DeA
Normal forearm	AnN
Abnormal forearm	AnA
Normal hand	MaN
Abnormal hand	MaA
Normal humerus	HuN
Abnormal humerus	HuA
Normal wrist	MuN
Abnormal wrist	MuA

2.4.1. EfficientNet-B4

EfficientNet is a convolutional neural network architecture that optimizes the balance among network depth, width, and resolution [27].

The EfficientNet-B4 model consists of 24 layers and approximately 19.3 million parameters, most of them trainable (weights and biases) with a small, fixed proportion (mean and variance). This configuration enhances the model's ability to extract complex features but also increases computational demand and the risk of overfitting, see figure 3.

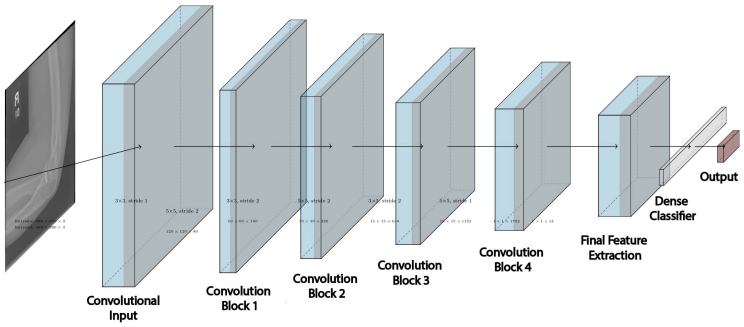


Figure 3. EfficientNet-B4 architecture

Phase 1

The training process was conducted in two stages, using a batch size of eight images per iteration. In the first stage, the model was trained for 30 epochs with the last 100 layers frozen and a learning rate of 1×10^{-3}.

In the second stage, fine-tuning was performed for 25 epochs with all layers unfrozen, while the learning rate was reduced to 1×10^{-5}.

Data augmentation techniques, including rotations, shifts, and scaling, were applied to increase the variability of the training set. In addition, a validation loss monitoring strategy was implemented to halt training in the absence of improvement and to dynamically adjust the learning rate.

Phase 2

In the multiclass training, a batch size of eight images was used.

In the first stage, only the output layer was trained for 30 epochs with a learning rate of 1×10^{-3}.

In the second stage, fine-tuning was performed by unfreezing all layers for an additional 30 epochs, while reducing the learning rate to 1×10^{-5} to achieve a more precise adjustment of the model's weights.

As part of the training workflow, custom class weights were calculated to compensate for the class imbalance in the training set.

Phase 3

During training, data augmentation was performed directly on the GPU using the Kornia library [28], which applied random rotations of up to 30°, horizontal flips, brightness variations, and normalization.

For validation, the data were only normalized to maintain statistical consistency with the training set.

To address class imbalance, class weights were automatically calculated using a function described in Equation (1).

$$w_i = \frac{n_{samples}}{n_{classes} \times n_{samples_i}} \tag{1}$$

Where w_i is the weight of class i, $n_{samples}$ represents the total number of samples, $n_{(clases)}$ corresponds to the number of classes, and $n_{samples}$ indicates the number of samples belonging to class i.

Additionally, the cross-entropy loss function was employed, converting the model outputs into normalized probabilities that can be directly compared with the true labels.

The training structure used in Phase 2 was maintained, and mixed-precision operations were automatically incorporated [29].

Finally, dynamic loss scaling was applied to ensure numerical stability during training [30].

2.4.2. ResNet-50

ResNet is a convolutional neural network architecture based on residual blocks and skip connections, designed to mitigate issues such as gradient vanishing and to facilitate the training of deeper networks [31].

The ResNet-50 model comprises 50 layers, enabling it to extract detailed visual features from images, as illustrated in Figure 4 [12].

Phase 1

For binary training, a batch size of 32 images was used, following a two-stage strategy that included initial training with frozen layers and subsequent fine tuning.

In the first stage, the base model layers were frozen, and a custom classifier was added, consisting of two dense layers with 1,024 and 512 units, normalization, the Swish activation function, and a 50% dropout rate, culminating in a sigmoid output layer for binary classification.

Various transformations were applied to the training set, including pixel normalization, rotations, shifts of up to 30%, zooming up to 40%, random cropping, brightness adjustments, and horizontal flips.

In the first stage, the model head was trained for 40 epochs with a learning rate of 1×10^{-4}; in the second stage, the last 150 layers were unfrozen, and fine-tuning continued for an additional 40 epochs using a reduced learning rate of 1×10^{-6} to preserve the stability of the pre-trained weights.

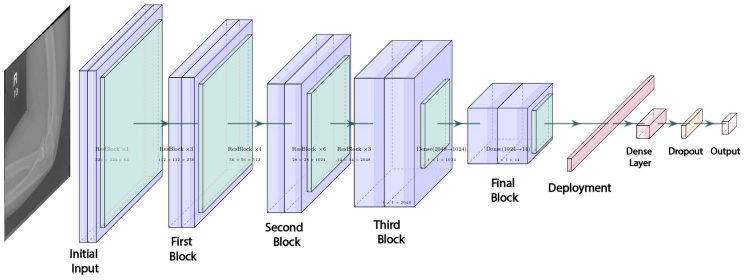


Figure 4. ResNet-50 Architecture

Phase 2

In this phase, no data augmentation was applied. The training was conducted in two stages: in the first, the base layers of the model were kept frozen, and only the top layers were trained with a learning rate of 1×10^{-4}; in the second, starting from epoch 40, fine-tuning was performed with the learning rate reduced to 1×10^{-6}.

Phase 3

In the third phase, a batch size of 64 images was used. The training set underwent random horizontal flipping, rotations within a $\pm$ 30° range, and normalization, with the mean and standard deviation adjusted to 0.5. Additionally, class weights were automatically computed using the compute_class_weight function, as described in Phase 3, Section 2.4.1.

For training, the final layer of the model was replaced with an output of 14 neurons, preceded by a dense layer with 512 units, batch normalization, a nonlinear activation function, and a 50.0% dropout rate. The process was maintained in two stages, incorporating mixed-precision training techniques to optimize GPU resource utilization.

2.4.3. ConvNeXt-Large

ConvNeXt-Large is a convolutional neural network architecture composed of multiple identical modules, which facilitates model scalability and maintenance [32]. Figure 5 illustrates its 24-layer structure, enabling the extraction of subtle differences and complex details in images, capabilities that are particularly valuable for medical image classification tasks [13].

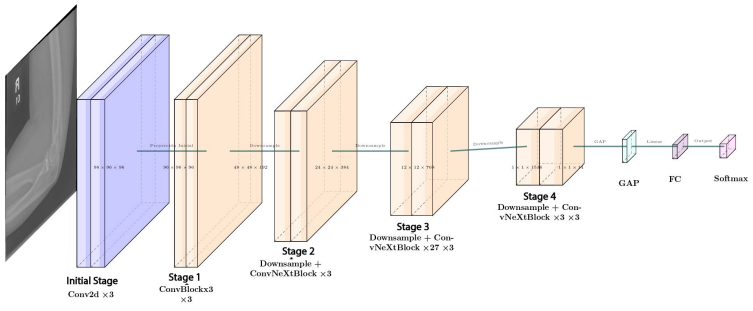


Figure 5. ConvNeXt-Large Architecture

Phase 1

In the first phase of the binary training of ConvNeXt-Large, data augmentation was applied through brightness and contrast adjustments, motion blur, rotations of $\pm$ 20º, random cropping, horizontal flipping, and normalization to the range [-1, 1]. A composite loss function was employed, combining 30% Binary Cross-Entropy (BCE) [33], which measures the discrepancy between the model's predictions and the true labels, and 70% Focal Loss [34], which reduces the influence of easy examples while emphasizing more difficult ones.

The model was trained for 30 epochs, incorporating

CutMix [35] and progressive fine-tuning. Initially, only the model head was trained, while from epochs 3 and 8, the deeper layers were gradually unfrozen. Starting at epoch 5, the classification threshold was dynamically adjusted, and validation errors were reintegrated every 5 epochs.

Phase 2

In the second phase, the model was fine-tuned for multiclass classification across14 categories, without applying data augmentation, using a custom dataset built from CSV files to facilitate image loading and normalization.

The training was performed with batches of 16 images, employing the CrossEntropyLoss function, as described in Phase 3, Section 2.4.1, and the AdamW optimizer, which updates parameters with adaptive learning rates and applies weight regularization to improve the stability of the learning process.

The training began with the base model frozen, progressively unfreezing the last 10 layers from epoch 3 and the entire architecture from epoch 8 onward.

Phase 3

In the third and final phase, generalization capability was enhanced through advanced data transformations, including horizontal flipping, discrete rotations of 90º, 180º, and 270º, elastic distortions, Gaussian blur, and random brightness and contrast adjustments. Oversampling was applied by replicating minority-class examples to balance the dataset. Additionally, a custom loss function was implemented, incorporating error detection every 5 epochs and the use of CutMix [35].

The training began with the model head for 2 epochs; from epoch 3, the last 10 layers were activated, and from epoch 8 onward, full fine-tuning was enabled.

Table 4 summarizes the configurations used for EfficientNet-B4, ResNet-50, and ConvNeXt-Large, including the number of epochs, batch sizes [36], and learning rates, with specific adjustments for multiclass classification.

Table 4. Summary of training configurations by model

Parameter	EfficientNet	ResNet	ConvNeXt
Initial number of epochs	30	40	30
Fine-tuning epochs	25	40	30
Initial batch size	8	32	16
Final batch size	8	64	16
Initial learning rate	1×10^{-3}	1×10^{-4}	1×10^{-3}
Final learning rate	1×10^{-5}	1×10^{-6}	1×10^{-5}

2.4.4. Computational effort

Computational effort refers to the resources required to train and execute a model, including memory usage, training time, and processing capacity [37]. In this study, the models exhibited notable differences across the three hardware configurations employed, which are described below.

- **Configuration 1 (GPU NVIDIA RTX 3050, 8 GB VRAM):** mid-range GPU designed for home users or basic workstations, based on the Ampere architecture and suitable for general-purpose graphics processing and accelerated computing tasks.
- **Configuration 2 (GPU NVIDIA RTX 4060, 12 GB VRAM):** a high-end consumer GPU built on the Ada Lovelace architecture, offering higher performance for deep learning workloads.
- **Configuration 3 (Google Colab Pro con GPU A100, 40 GB VRAM):** a high-performance GPU based on the Ampere architecture, optimized for intensive processing and deep neural network training.

The training process required a high computational load: EfficientNet-B4 exhibited moderate demand, ResNet-50 required less time and memory, and ConvNeXt-Large was the most resource-intensive. Therefore, the PyTorch function torch.cuda.empty_cache() was employed to release GPU memory and reduce the risk of saturation [38].

Table 5 summarizes the observed training performance across the different hardware configurations used. The performance levels are interpreted as follows:

- **Slow:** total training time greater than 12 hours.
- **Moderate:** total training time between 4 and 12 hours.
- **Fast:** total training time less than 4 hours.

Table 5. Comparison of model training speed

Model	RTX 3050	RTX 4060	Colab A100
EfficientNet	Moderate	Fast	Fast
ResNet	Moderate	Fast	Fast
ConvNeXt	Slow	Moderate	Fast

Similarly, Table 6 presents the observed memory usage during training across the different environments evaluated. The memory utilization levels are interpreted as follows:

- **Low:** maximum VRAM utilization below 50% of the total available GPU capacity.
- **Moderate:** VRAM utilization between 50% and 80%.
- **High:** VRAM utilization above 80%, with an increased risk of saturation.

Table 6. Comparison of model memory usage

Model	RTX 3050	RTX 4060	Colab A100
EfficientNet	Moderate	Low	Low
ResNet	Moderate	Moderate	Low
ConvNeXt	High	High	High

2.5. Evaluation

This section analyzes the performance of the EfficientNet-B4, ResNet-50, and ConvNeXt-Large models in bone fracture classification using evaluation metrics such as F1-score, precision, accuracy, sensitivity, receiver operating characteristic (ROC) curve, and confusion matrix. The abbreviations used in the equations are detailed in Table 7.

Table 7. Evaluation terms

Abbreviation	Meaning
TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives
RECALL	Sensitivity

The metrics used for performance evaluation are described below.

F1-score

The F1-score is the harmonic mean of precision and sensitivity, particularly useful in imbalanced datasets as it provides a balanced assessment of classification performance [39].

$$F1 - Score = 2 \times \frac{Precision \times sensitivity}{Precision + sensitivity} \quad (2)$$

Precision

Precision measures the proportion of positive predictions that are correct, indicating a low incidence of false positives [40].

$$Precisión = \frac{TP}{TP + FP} \quad (3)$$

Accuracy

Accuracy measures the overall proportion of correct predictions and is particularly relevant when the classes are balanced [41].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Sensitivity

Sensitivity reflects the model's ability to correctly identify true positives; high sensitivity indicates a low number of false negatives [39].

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

Confusion matrix

The confusion matrix provides the TP, FP, FN, and TN values, facilitating performance evaluation and the identification of model errors [42].

Receiver Operating Characteristic (ROC) curve

The receiver operating characteristic (ROC) curve represents the relationship between sensitivity and the false positive rate, enabling the analysis of the model's discriminative ability and visualization of its overall performance [43].

3. Results and discussion

Phase 1

As shown in Table 8, ConvNeXt-Large outperforms EfficientNet-B4 and ResNet-50 across all evaluated metrics, achieving a precision of 99.2%, a sensitivity of 97.7%, and an F1-score of 98.4%, demonstrating its strong ability to classify bone fractures with a low false positive rate.

In contrast, EfficientNet-B4 and ResNet-50 performance but lower effectiveness in correctly identifying positive cases. These results indicate that ConvNeXt-Large provides the best balance between precision, sensitivity, and generalization capability.

Table 8. Evaluation metrics (in percentage) – Phase 1

Metric	EfficientNet	ResNet	ConvNeXt
Precision	82.4	82.4	99.2
Sensitivity	69.0	69.0	97.7
F1-Score	75.1	75.1	98.4
Accuracy	81.5	73.0	98.7

To complement these results, the receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) values for each model are presented in Figure 6. EfficientNet-B4 achieved an AUC of 87.1%, ResNet-50 obtained 82.8%, and ConvNeXt-Large reached 99.9%, demonstrating an almost perfect ability to discriminate between classes.

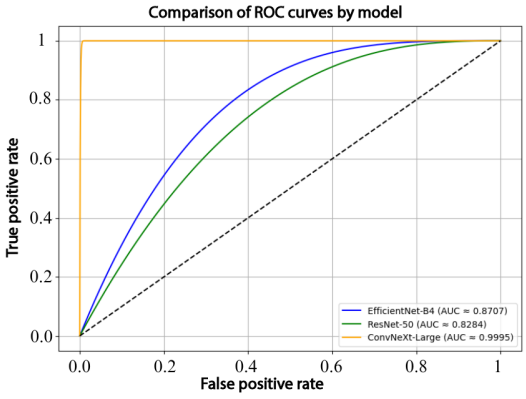


Figure 6. Comparison of model performance using ROC curves

The analysis of the confusion matrices reinforces these findings: EfficientNet-B4 (Figure 7) produced 1013 false positives (FP) and 470 false negatives (FN); ResNet-50 (Figure 8) reduced the number of false negatives to 697 but increased false positives to 1465; whereas ConvNeXt-Large (Figure 9) achieved the best performance, with only 76 false positives and 25 false negatives, demonstrating superior classification accuracy.

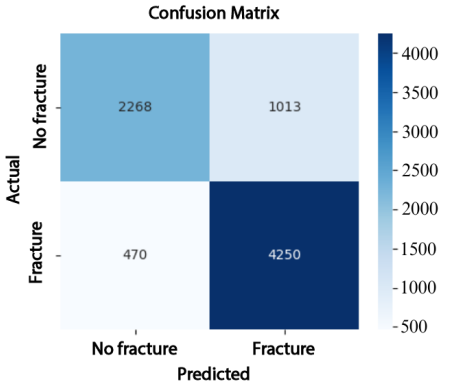


Figure 7. Confusion matrix of EfficientNet-B4

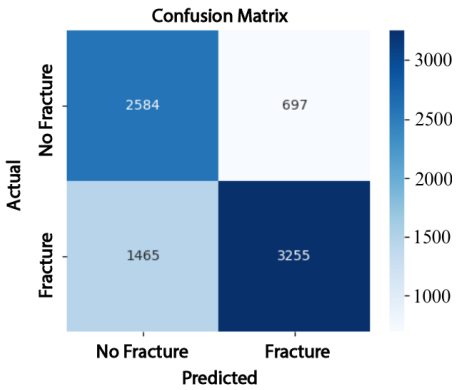


Figure 8. Confusion matrix of ResNet-50

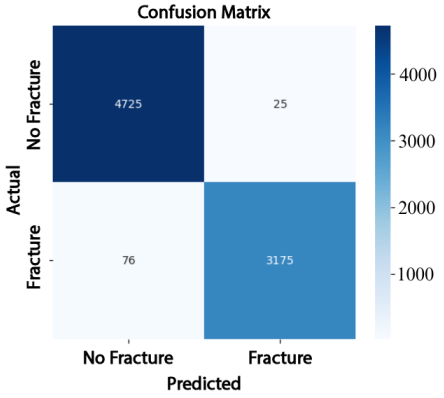


Figure 9. Confusion matrix of ConvNeXt-Large

Phase 2

In Phase 2, the multiclass training results indicate that the ConvNeXt-Large model achieved an overall precision of 82.3%, a sensitivity of 81.6%, an F1-score of 81.4%, and an accuracy of 81.3%. These results are summarized in Table 9.

Table 9. Evaluation metrics (in percentage) – Phase 2

Metric	EfficientNet	ResNet	ConvNeXt
Precision	47.3	73.5	82.3
Sensitivity	33.2	72.8	81.6
F1-Score	32.8	72.2	81.4
Accuracy	33.2	72.8	81.6

Figure 10 shows that the AUC values obtained by EfficientNet-B4 for each class range between 93.0% and 98.0%.

In turn, Figure 11, corresponding to ResNet-50, presents AUC values ranging from 95.0% to 99.0%. Finally, Figure 12 illustrates the AUC curves of ConvNeXt-Large, which range from 98.0% to 100.0%, demonstrating outstanding performance and excellent class discrimination capability.

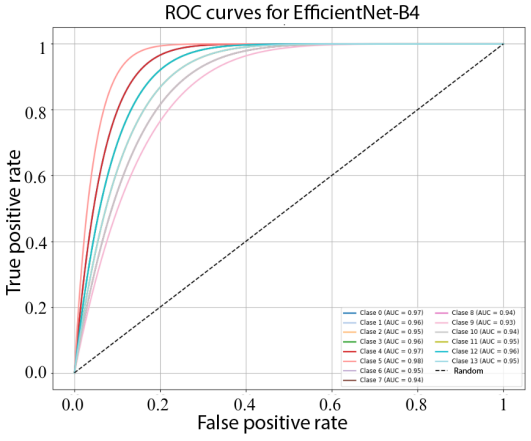


Figure 10. Receiver operating characteristic (ROC) curves for multiclass EfficientNet-B4

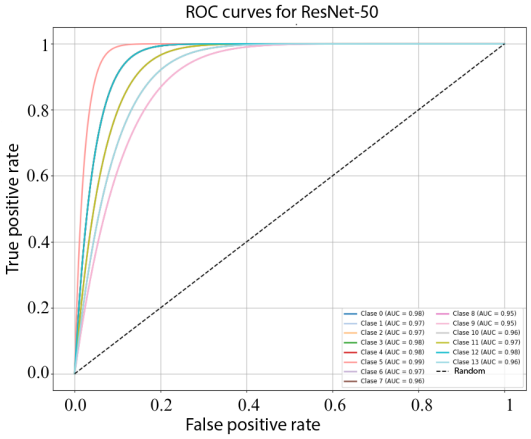


Figure 11. Receiver operating characteristic (ROC) curves for multiclass ResNet-50.

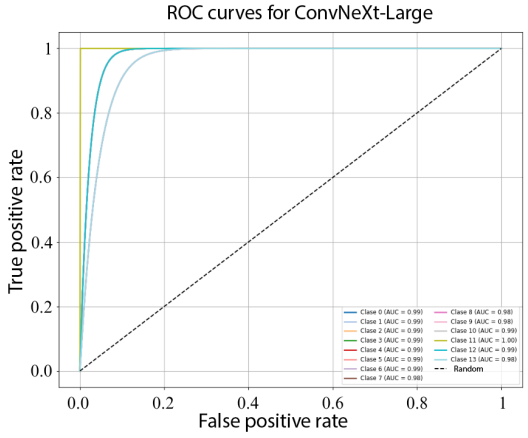


Figure 12. Receiver operating characteristic (ROC) curves for multiclass ConvNeXt-Large.

In Figure 13, EfficientNet-B4 demonstrates limited performance, as evidenced by the numerous misclassifications in the classes Normal Elbow and Normal Hand. In contrast, Abnormal Finger, Abnormal Hand, and Normal Wrist exhibit high precision.

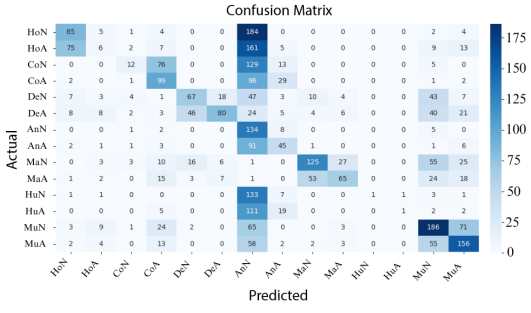


Figure 13. Confusion matrix of EfficientNet-B4

Conversely, Figure 14 shows that ResNet-50 achieves improved classification performance; in this case, most classes are correctly concentrated along the diagonal of the confusion matrix, although misclassifications between Normal Shoulder and Abnormal Wrist remain.

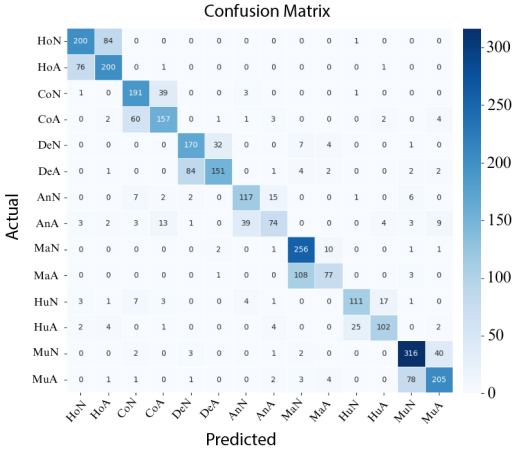


Figure 14. Confusion matrix of ResNet-50

Finally, in Figure 15 shows that ConvNeXt-Large achieved the best performance, delivering highly precise classifications with a minimal number of errors.

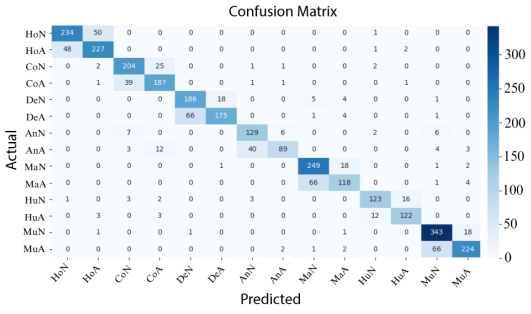


Figure 15. Confusion matrix of ConvNeXt-Large

Phase 3

Based on the results presented in Table 10, ConvNeXt-Large maintained its superiority in multiclass classification, clearly outperforming the other models across all evaluation metrics. Its precision, sensitivity, F1-score, and accuracy reached 99.4%, confirming ConvNeXt-Large as the most robust and effective model for the fracture classification task.

Table 10. Evaluation metrics (in percentage) – Phase 3

Model	EfficientNet	ResNet	ConvNeXt
Precision	80.6	83.9	99.4
Sensitivity	80.2	83.9	99.4
F1-Score	79.9	83.9	99.4
Accuracy	80.2	83.9	99.4

When examining the confusion matrices corresponding to Phase 3, it can be observed that, in Figure 16, the AUC values of EfficientNet-B4 range from 98.0% to 99.0%.

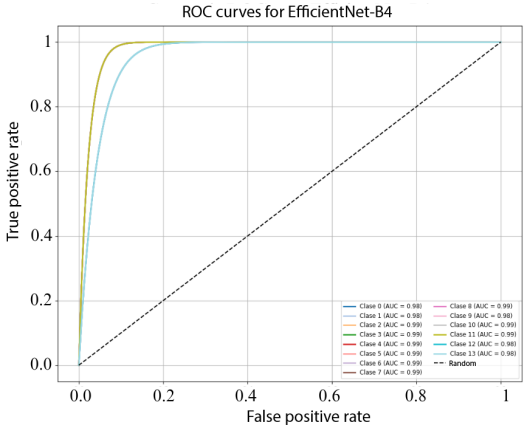


Figure 16. Receiver operating characteristic (ROC) curves for multiclass EfficientNet-B4

In the case of ResNet-50, the AUC values range from 97.0% to 99.0%, as shown in Figure 17.

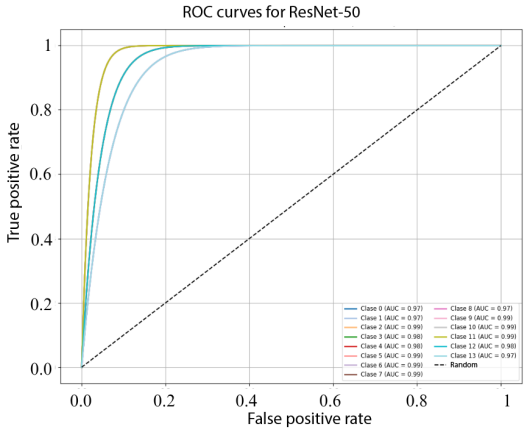


Figure 17. Receiver operating characteristic curves (ROC) for multiclass ResNet-50

Finally, as illustrated in Figure 18, all classes in ConvNeXt-Large achieved an AUC of 100.0%, demonstrating its exceptional performance and near-perfect classification capability.

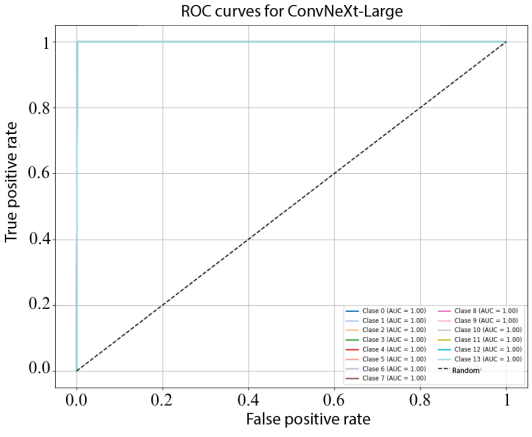


Figure 18. Receiver operating characteristic (ROC) curves for multiclass ConvNeXt-Large

Figures 19, 20, and 21 confirm that ConvNeXt-Large is the most effective model, exhibiting minimal dispersion outside the main diagonal and a superior ability to differentiate between classes.

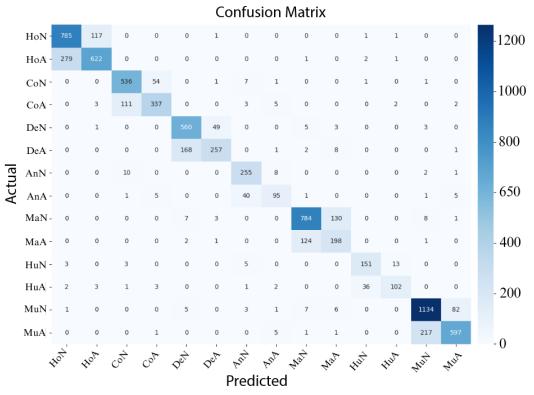


Figure 19. Confusion matrix of EfficientNet-B4

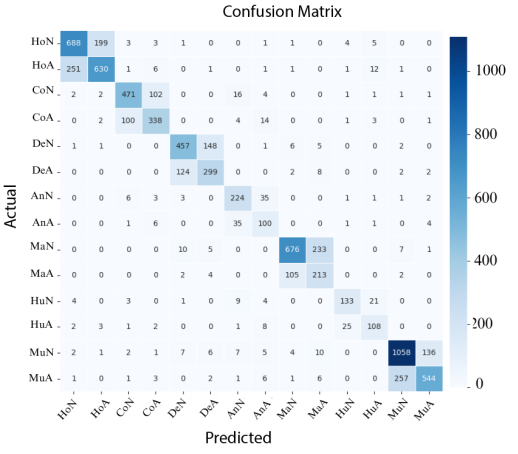


Figure 20. Confusion matrix of ResNet-50

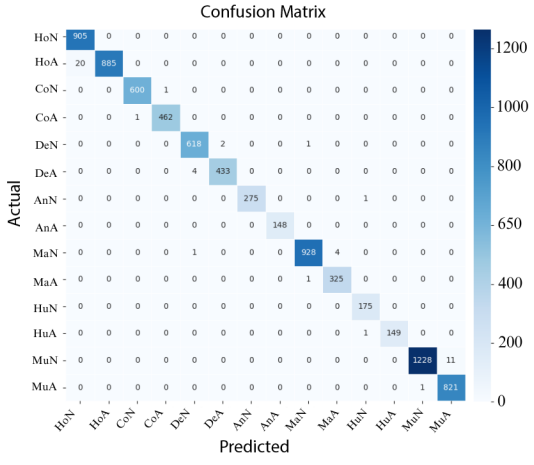


Figure 21. Confusion matrix of ConvNeXt-Large

The excellent performance of ConvNeXt-Large in handling class imbalance can be attributed to the use of advanced data augmentation strategies, which reduced overfitting and enhanced generalization capability, allowing the model to outperform EfficientNet-B4 and ResNet-50 across all evaluated phases.

Comparison with related works

The proposed ConvNeXt-Large model outperformed results reported in previous studies, achieving an accuracy of up to 99.4%, compared with 62.0% for R-CNN [10] and 86.8% for DenseNet121 [23], thereby demonstrating superior robustness even when handling imbalanced datasets.

Limitations, challenges for clinical application, and ethical considerations

Although the results obtained demonstrate high performance, their clinical implementation requires more comprehensive analysis. Deployment in real-world settings presents challenges such as variability in radiological equipment, differences in image quality, and anatomical diversity among patients, all of which necessitate validating the model across a wide range of clinical scenarios.

Additionally, there is a potential risk of overfitting, as the model was trained in a controlled environment that may not fully represent real-world clinical conditions. Therefore, external validation using heterogeneous datasets is essential to ensure reliability and generalizability. In terms of reproducibility, it is crucial that the results can be replicated by other researchers employing different datasets and experimental configurations, thereby strengthening confidence in the findings and supporting clinical adoption.

Finally, from an ethical and regulatory perspective, it is imperative to guarantee model transparency, the protection and anonymization of patient data, and the clear delineation of responsibilities in the event of diagnostic errors. These measures are fundamental to ensuring the safe, ethical, and trustworthy application of artificial intelligence systems in healthcare, always prioritizing patient well-being.

4. Conclusions

This study evaluated the EfficientNet-B4, ResNet-50, and ConvNeXt-Large models for the automatic classification of upper-limb bone fractures using digital radiographs. The findings demonstrated that ConvNeXt-Large outperformed the other models, achieving accuracies of 99.2% in binary classification, 82.3% in multiclass classification with original data, and 99.4% in multiclass classification with preprocessed data.

The combination of deep learning strategies, including image preprocessing, data augmentation, dynamic threshold adjustment, and progressive fine-tuning, proved essential for achieving its high performance.

Although EfficientNet-B4 and ResNet-50 achieved satisfactory results, their lower sensitivity and challenges in handling class imbalance could present risks in clinical applications, where missing a fracture may have critical implications. Overall, the results position ConvNeXt-Large as a promising tool for supporting medical diagnosis and integration into automated radiographic analysis systems for upper-limb fractures.

A key challenge identified involves the inherent complexity of fracture patterns, underscoring the need to explore advanced resampling methods and specialized augmentation strategies, such as clinical transformations or synthetic data generation using generative networks, to further enhance diagnostic accuracy and model robustness in future research.

Code availability

The code used for data processing, model training, and figure generation in the study on upper-limb fracture classification using deep learning is publicly available in the following GitHub repository [44].

Contributor Roles

- **Gabriela Jaén-Armijos:** data curation, formal analysis, investigation, software, writing – original draft.
- **Evelyn Morán-Castillo:** investigation, methodology, formal analysis software, writing – original draft.
- **Wilmer Rivas-Asanza:** supervision, validation, writing – review & editing.
- **Eduardo Tusa:** supervision, validation, writing – review & editing.

References

[1] J. S. Alvarez Silva, J. E. Chávez Guerrero, D. S. Gutiérrez Carvajal, and L. B. Santos Rivas, “Manejo de emergencias en traumatología,” *Polo del conocimiento*, vol. 8, no. 4, pp. 2392–2414, 2023. [Online]. Available: https://upsalesiana.ec/ing35ar2r1

[2] E. M. Bitar Benítez, S. Gaviria Posada, N. d. León Fernández, and S. Gallego Gónima, “Epidemiología del trauma de miembro superior atendido en seis instituciones de salud de la ciudad de Medellín, Colombia, en 2016,” *Cirugía Plástica Ibero-Latinoamericana*, vol. 47, no. 3, pp.

323–332, Sep. 2021. [Online]. Available: https://dx.doi.org/10.4321/s0376-78922021000300014

[3] Y. Xie, X. Li, F. Chen, R. Wen, Y. Jing, C. Liu, and J. Wang, "Artificial intelligence diagnostic model for multi-site fracture X-ray images of extremities based on deep convolutional neural networks," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 2, pp. 1930–1943, Feb. 2024. [Online]. Available: https://dx.doi.org/10.21037/qims-23-878

[4] A. Said, M. Abbas Al-Adilee, and M. Rajeswari, "Enhancing bone fracture diagnosis with deep learning," in *Conference: 5th International Conference on Data Engineering and Communication Technology (ICDECT)At: Kuala Lumpur, Malaysia*, 09 2024. [Online]. Available: https://upsalesiana.ec/ing35ar2r4

[5] Z. Alammar, L. Alzubaidi, J. Zhang, J. Santamaría, Y. Li, and Y. Gu, "A concise review on deep learning for musculoskeletal X-ray images," in *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, Nov. 2022, pp. 1–8. [Online]. Available: https://doi.org/10.1109/DICTA56598.2022.10034618

[6] A. Pérez del Barrio, P. Menéndez Fernández-Miranda, P. Sanz Bellón, L. Lloret Iglesias, and D. Rodríguez González, "Inteligencia artificial en radiología: introducción a los conceptos más importantes," *Radiología*, vol. 64, no. 3, pp. 228–236, May 2022. [Online]. Available: https://doi.org/10.1016/j.rx.2022.03.003

[7] J. D. Maravilla González, "Integración de la inteligencia artificial en la contabilidad forense: Herramientas y eficacia en la detección de fraudes: Integration of artificial intelligence in forensic accounting: Tools and effectiveness in fraud detection," *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, vol. 5, no. 4, Aug. 2024. [Online]. Available: https://doi.org/10.56712/latam.v5i4.2460

[8] S. C. Medaramatla, C. V. Samhitha, S. D. Pande, and S. R. Vinta, "Detection of hand bone fractures in x-ray images using hybrid YOLO NAS+," *IEEE Access*, vol. 12, pp. 57 661–57 673, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3379760

[9] J. Wang, "Anomaly detection of arm X-ray based on deep learning," *IOP Conference Series: Earth and Environmental Science*, vol. 440, no. 4, p. 042056, Feb. 2020. [Online]. Available: https://doi.org/10.1088/1755-1315/440/4/042056

[10] B. Guan, G. Zhang, J. Yao, X. Wang, and M. Wang, "Arm fracture detection in X-rays based on improved deep convolutional neural network," *Computers & Electrical Engineering*, vol. 81, p. 106530, Jan. 2020. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2019.106530

[11] A. Pak, A. Ziyaden, K. Tukeshev, A. Jaxylykova, and D. Abdullina, "Comparative analysis of deep learning methods of detection of diabetic retinopathy," *Cogent Engineering*, vol. 7, no. 1, Jan. 2020. [Online]. Available: https://doi.org/10.1080/23311916.2020.1805144

[12] M. A. Thaarakaraam, M. M. Reddy, and V. Vijayakumar, "CNN based bone fracture detection for medical imaging using RESNET-50," *International Journal of Technical Research & Science*, vol. 9, no. Spl, pp. 27–35, Jun. 2024. [Online]. Available: https://doi.org/10.30780/specialissue-iset-2024/038

[13] Z. Li, T. Gu, B. Li, W. Xu, X. He, and X. Hui, "Convnext-based fine-grained image classification and bilinear attention mechanism model," *Applied Sciences*, vol. 12, no. 18, p. 9016, Sep. 2022. [Online]. Available: https://doi.org/10.3390/app12189016

[14] C. A. Mejía Rodríguez, M. A. Rincón Pinzón, L. M. Palmera Quintero, and L. M. Arévalo Vergel, "Aplicación de machine learning y metodología CRISP-DM para la clasificación precisa de severidad en casos de dengue," *Revista Co,ombiana de Tecnologías de Avanzada (RCTA)*, vol. 1, no. 43, pp. 78–85, Mar. 2024. [Online]. Available: https://doi.org/10.24054/rcta.v1i43.2822

[15] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021. [Online]. Available: https://doi.org/10.1016/j.procs.2021.01.199

[16] D. Dzulhijjah, M. Herlambang, and M. Haifan, "Implementasi framework CRISP-DM untuk proses data mining aplikasi credit scoring PT. XYZ," in *Seminar Nasional Sains dan Teknologi "SainTek"*, 06 2024. [Online]. Available: https://upsalesiana.ec/ing35ar2r16

[17] V. García-Ríos, M. Marres-Salhuana, F. Sierra-Liñán, and M. Cabanillas-Carbonell, "Predictive machine learning applying cross industry standard process for data mining for the diagnosis of diabetes mellitus type 2," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 4, p. 1713, Dec. 2023. [Online]. Available: https://doi.org/10.11591/ijai.v12.i4.pp1713-1726

[18] G. Emmanuel, G. G. Hungilo, and A. W. R. Emanuel, "Performance evaluation of machine learning classification techniques for diabetes disease," *IOP Conference Series: Materials Science and Engineering*, vol. 1098, no. 5, p. 052082, Mar. 2021. [Online]. Available: https://doi.org/10.1088/1757-899X/1098/5/052082

[19] E. B. Núñez Arias, B. M. González Nuñez, L. N. Fernández, and J. M. Rodríguez Pupo, "CRISP-DM y K-means neutrosofía en el análisis de factores de riesgo de pérdida de audición en niños," *Neutrosophic Computing and Machine Learning*, vol. 16, p. 73, 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4679890

[20] ML Group. (2025) Bone X-ray deep learning competition. Stanford ML Group. [Online]. Available: https://upsalesiana.ec/ing35ar2r20

[21] I. Kandel and M. Castelli, "Improving convolutional neural networks performance for image classification using test time augmentation: A case study using MURA dataset," *Health Information Science and Systems*, vol. 9, no. 1, Jul. 2021. [Online]. Available: https://doi.org/10.1007/s13755-021-00163-7

[22] K. Dlshad Ahmed and R. Hawezi, "Detection of bone fracture based on machine learning techniques," *Measurement: Sensors*, vol. 27, p. 100723, Jun. 2023. [Online]. Available: https://doi.org/10.1016/j.measen.2023.100723

[23] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," in *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL)*. arXiv, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1712.06957

[24] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artificial Intelligence Review*, vol. 57, no. 1, Jan. 2024. [Online]. Available: https://doi.org/10.1007/s10462-023-10631-z

[25] P. Pandey, R. Gupta, and N. Goel, "Comprehensive review of single image defogging techniques: enhancement, prior, and learning based approaches," *Artificial Intelligence Review*, vol. 58, no. 4, Jan. 2025. [Online]. Available: https://doi.org/10.1007/s10462-024-11034-4

[26] A. C. Pavan, S. Lakshmi, and M. Somashekara, "An improved method for reconstruction and enhancing dark images based on CLAHE," *International Research Journal on Advanced Science Hub*, vol. 5, no. 02, pp. 40–46, Feb. 2023. [Online]. Available: https://doi.org/10.47392/irjash.2023.011

[27] X. Liu, *Comparison of four convolutional neural network-based algorithms for sports Image classification*. Atlantis Press International BV, 2024, pp. 178–186. [Online]. Available: https://doi.org/10.2991/978-94-6463-370-2_20

[28] E. Riba, D. Mishkin, D. Ponsa, E. Rublee, and G. Bradski, "Kornia: An open source differentiable computer vision library for PyTorch," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. arXiv, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1910.02190

[29] R. Connor, A. Dearle, B. Claydon, and L. Vadicamo, "Correlations of cross-entropy loss in machine learning," *Entropy*, vol. 26, no. 6, p. 491, Jun. 2024. [Online]. Available: https://doi.org/10.3390/e26060491

[30] M. Dörrich, M. Fan, and A. M. Kist, "Impact of mixed precision techniques on training and inference efficiency of deep neural networks," *IEEE Access*, vol. 11, pp. 57 627–57 634, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3284388

[31] A. Thakur, H. Chauhan, and N. Gupta, "Efficient ResNets: Residual network design," in *NYU Tandon School of Engineering, Tech. Rep.* arXiv, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2306.12100

[32] G. Fang, Y. Dai, Z. Lin, C. Zhou, J. Song, Y. Gu, X. Guo, A. Mao, and X. Kong, "An efficient unsupervised classification model for galaxy morphology: Voting clustering based on coding from convnext large model," *Astronomy & Astrophysics*, vol. 693, p. A141, Jan. 2025. [Online]. Available: https://doi.org/10.1051/0004-6361/202451734

[33] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Díaz, "An overview of deep learning in medical imaging," *Informatics in Medicine Unlocked*, vol. 26, p. 100723, 2021. [Online]. Available: https://doi.org/10.1016/j.imu.2021.100723

[34] A. A. Mustapha and M. S. Yoosuf, "Exploring the efficacy and comparative analysis of one-stage object detectors for computer vision: A review," *Multimedia Tools and Applications*, vol. 83, no. 20, pp. 59 143–59 168, Dec. 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-17751-2

[35] Y. Nakamura, Y. Ishii, Y. Maruyama, and T. Yamashita, "Few-shot adaptive object detection with cross-domain cutmix," in *Proceedings*

of the European Conference on Computer Vision (ECCV). arXiv, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2208.14586

[36] M. T. Mora Cabral, R. G. Camero Berrones, and M. D. Arriaga Pons, "Modelo de identificación espacial de patrones de referencia empleando redes neuronales convolucionales (CNN) y entrenamiento por lotes," *Revista de investigación multidisiplinaria, Iberoamericana*, no. 4, Dec. 2024. [Online]. Available: https://doi.org/10.69850/rimi.vi4.125

[37] A. Verma, T. Meenpal, and B. Acharya, "Computational cost reduction of convolution neural networks by insignificant filter removal," *Journal of Information Science and Technology*, vol. 25, no. 2, pp. 150–165, 2022. [Online]. Available: https://upsalesiana.ec/ing35ar2r37

[38] C. Guo, R. Zhang, J. Xu, J. Leng, Z. Liu, Z. Huang, M. Guo, H. Wu, S. Zhao, J. Zhao, and K. Zhang, "GMLake: Efficient and Transparent GPU Memory Defragmentation for Large-scale DNN Training with Virtual Memory Stitching," in *arXiv preprint*. arXiv, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2401.08156

[39] V. D. Gil-Vera and C. Seguro-Gallego, "Machine learning aplicado al análisis del rendimiento de desarrollos de software," *Revista Politécnica*, vol. 18, no. 35, pp. 128–139, Apr. 2022. [Online]. Available: https://doi.org/10.33571/rpolitec.v18n35a9

[40] F. D. Valle-Medina, L. J. Castillo-Heredia, M. A. Correa-Peralta, and J. E. Guzmán-Seraquive, "Clasificación de enfermedades en hojas de papa utilizando transformadores de visión," *Novasinergia Revista digital de Ciencia, Ingeniería y Tecnología*, vol. 8, no. 1, pp. 142–156, Jan. 2025. [Online]. Available: https://doi.org/10.37135/ns.01.15.06

[41] M. Ihme, W. T. Chung, and A. A. Mishra, "Combustion machine learning: Principles, progress and prospects," *Progress in Energy and Combustion Science*, vol. 91, p. 101010, Jul. 2022. [Online]. Available: https://doi.org/10.1016/j.pecs.2022.101010

[42] D. Patiño-Pérez, L. Armijos-Valarezo, L. Chóez-Acosta, and F. Burgos-Robalino, "Convolutional neural networks for diabetic retinopathy detection," *Ingenius*, no. 33, pp. 91–101, Mar. 2025. [Online]. Available: https://doi.org/10.17163/ings.n33.2025.08

[43] G. Bauce and M. Moya-Sifontes, "Análisis de la curva ROC en la evaluación de indicadores antropométricos," *Revista Digital de Postgrado*, vol. 11, no. 1, Aug. 2021. [Online]. Available: https://doi.org/10.37910/RDP.2022.11.1.e333

[44] G. B. Jaén Armijos, "Clasificador de fracturas en extremidades superiores con aprendizaje profundo," Github, oct 2025, gitHub repository. [Online]. Available: https://upsalesiana.ec/ing35ar2r44

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.03
pISSN: 1390-650X / eISSN: 1390-860X

Advancements in Computational Analysis for Jet Engine Optimization: A Review of CFD, Structural Analysis, and Multidisciplinary Approaches

Avances en el análisis computacional para la optimización de motores a reacción: una revisión de CFD, análisis estructural y enfoques multidisciplinarios

Abu Baker Jassim[1,*], Raja Sekhar Dondapati[1]

Received: 20-03-2025, Received after review: 27-05-2025, Accepted: 22-10-2025, Published: 01-01-2026

Abstract

The evolution of jet engine technology has been profoundly influenced by advancements in computational analysis, particularly in Computational Fluid Dynamics (CFD), structural analysis, and multidisciplinary optimization. This review explores state-of-the-art computational techniques applied to jet engine analysis, emphasizing their applications, benefits, and inherent challenges. CFD has become an essential tool, enabling detailed simulations of complex aerodynamic and combustion processes. Methods such as Large Eddy Simulations (LES) and Direct Numerical Simulations (DNS) have provided deeper insights into turbulence and combustion dynamics, leading to improved efficiency and reduced emissions. However, these high-fidelity simulations entail significant computational costs, driving the development of more efficient algorithms and high-performance computing resources. The integration of structural analysis with aerodynamic simulations has facilitated the design of components capable of withstanding extreme operational conditions, thereby enhancing engine reliability and safety. Multidisciplinary Design Optimization (MDO) frameworks have further transformed engine design by simultaneously evaluating multiple performance metrics, resulting in configurations that balance efficiency, weight, and durability.

Resumen

La evolución de la tecnología de los motores a reacción ha estado profundamente influenciada por los avances en el análisis computacional, especialmente en la Dinámica de Fluidos Computacional (CFD), el análisis estructural y la optimización multidisciplinaria. Esta revisión aborda las técnicas computacionales de vanguardia aplicadas al estudio de motores a reacción, resaltando sus aplicaciones, beneficios y desafíos. La CFD se ha consolidado como una herramienta esencial que posibilita simulaciones detalladas de los complejos procesos aerodinámicos y de combustión. Métodos como las Simulaciones de Grandes Remolinos (LES) y las Simulaciones Numéricas Directas (DNS) han permitido comprender con mayor profundidad la turbulencia y la dinámica de la combustión, favoreciendo la eficiencia y reduciendo las emisiones. No obstante, estas simulaciones de alta fidelidad demandan un elevado costo computacional, impulsando el desarrollo de algoritmos más eficientes y recursos de alto rendimiento. La integración del análisis estructural con las simulaciones aerodinámicas ha facilitado el diseño de componentes capaces de soportar condiciones extremas, incrementando la fiabilidad y seguridad del motor.

[1,*]School of Mechanical Engineering, Lovely Professional University, Punjab, India
Corresponding author ✉: bakrjassim1990@gmail.com.

Suggested citation: A. B. Jassim, and R. S. Dondapati, "Advancements in Computational Analysis for Jet Engine Optimization: A Review of CFD, Structural Analysis, and Multidisciplinary Approaches," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 36-49, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.03.

Despite these advances, challenges remain in accurately modeling complex physical phenomena such as combustion instabilities and material behavior under high temperatures. The incorporation of machine learning techniques offers promising solutions to address these issues by complementing traditional computational methods with data-driven insights. Looking ahead, the future of computational analysis in jet engine development lies in the seamless integration of high-fidelity simulations, real-time data analytics, and adaptive modeling, paving the way for more efficient, reliable, and sustainable propulsion systems.

Keywords: Computational Fluid Dynamics (CFD), Multidisciplinary Design Optimization (MDO), High-Performance Computing, Turbulence and Combustion Modeling, Jet Engine Technology

Los marcos de Optimización de Diseño Multidisciplinario (MDO) han transformado el diseño al evaluar simultáneamente múltiples métricas de rendimiento, logrando configuraciones equilibradas entre eficiencia, peso y durabilidad. Pese a estos avances, persisten retos en el modelado de fenómenos complejos como las inestabilidades de combustión y el comportamiento de materiales a altas temperaturas. La incorporación del aprendizaje automático ofrece soluciones prometedoras, apuntando hacia sistemas de propulsión más eficientes, confiables y sostenibles.

Palabras clave: Dinámica de fluidos computacional (CFD), Optimización de diseño multidisciplinario (MDO), Computación de alto rendimiento, Modelado de turbulencia y combustión, Tecnología de motores a reacción

1. Introduction

Jet engines, the cornerstone of modern aviation, have undergone profound transformations since their inception. Central to these advancements has been the evolution of computational analysis, which has enabled engineers to simulate, predict, and optimize engine performance with unprecedented accuracy. The integration of computational tools into the design process has not only accelerated development but also deepened the understanding of the complex physical phenomena governing engine operation [1–3].

1.1. Historical Perspective

The application of computational techniques in jet engine development began with fundamental analytical and empirical approaches aimed at understanding basic aerodynamic and thermodynamic behavior [2, 4]. With the advent of digital computing in the mid-twentieth century, the field advanced rapidly, marking a pivotal transition from purely experimental methods to computationally driven design [1, 3]. The introduction of numerical methods enabled engineers to solve increasingly complex problems in fluid dynamics and structural mechanics, laying the groundwork for modern simulation-based analysis [5, 6]. Over the decades, computational modeling evolved from simple one-dimensional approximations to highly detailed, time-dependent three-dimensional simulations. This progression reflects the remarkable growth of computational capabilities and has fundamentally transformed the design, optimization, and performance evaluation of jet engines [6, 7].

1.2. Role of Computational Fluid Dynamics (CFD)

Computational Fluid Dynamics (CFD) has become an indispensable tool in jet engine analysis, providing detailed insights into aerodynamic performance, combustion processes, and heat transfer phenomena [1,2,5]. By numerically solving the Navier–Stokes equations, CFD enables a comprehensive understanding of airflow patterns, pressure distributions, and temperature fields within engine components [8, 9]. This capability has allowed engineers to optimize designs for greater efficiency and reduced emissions, offering a predictive framework that complements experimental testing [6, 10].

1.3. Structural Analysis and Multidisciplinary Optimization

Beyond aerodynamics, computational analysis extends to the structural evaluation of jet engine components through the Finite Element Method (FEM), which allows detailed assessment of stress distributions, deformations, and fatigue life under extreme conditions [6,7]. By capturing material responses with high fidelity, FEM contributes significantly to predicting component durability and overall engine reliability [5]. The emergence of Multidisciplinary Design Optimization (MDO) frameworks has further advanced computational engineering by integrating CFD, structural analysis, and performance domains [1, 10]. These frameworks enable simultaneous optimization of multiple design objectives–such as efficiency, weight, strength, and thermal performance–within a unified environment [6, 9].

1.4. Advancements in Simulation Techniques

Recent advances in simulation methodologies have been driven by the demand for higher accuracy in predicting complex flow and combustion phenomena. High-fidelity approaches such as Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) have emerged as powerful tools for exploring turbulent structures and heat transfer processes in jet engines [5, 6, 8]. Although these techniques require substantial computational resources, they provide unparalleled precision in capturing transient and multiscale phenomena beyond the reach of traditional methods [7, 9].

1.5. Challenges and Future Directions

Despite significant progress, challenges remain in accurately modeling combustion instabilities, turbulence interactions, and material behavior under extreme operating conditions [8, 10]. These complex phenomena often require advanced modeling and validation against experimental data. The integration of machine learning and data-driven techniques offers promising solutions to complement traditional numerical methods, enhancing predictive accuracy while reducing computational cost [6, 9]. The future of computational analysis in jet engine development will rely on merging high-fidelity simulations, real-time data analytics, and adaptive modeling, leading to propulsion systems that are more efficient, reliable, and environmentally sustainable [1, 3, 5].

2. Objectives

2.1. To Analyze State-of-the-Art Computational Techniques

The first objective of this review is to analyze state-of-the-art computational techniques applied in jet engine design. This includes exploring the latest developments in Computational Fluid Dynamics (CFD), Finite Element Methods (FEM), and Multidisciplinary Design Optimization (MDO).

Additionally, this objective involves comparing various simulation methodologies, such as Reynolds-Averaged Navier–Stokes (RANS), Large Eddy Simulation (LES), and Direct Numerical Simulation (DNS), to assess their respective capabilities, computational demands, and suitability for different aspects of jet engine performance analysis [11, 12].

2.2. To Identify the Benefits and Challenges of Computational Approaches

The objective aims to identify the benefits and challenges of computational approaches such as CFD, structural analysis, and MDO in engineering applications. These methods provide high accuracy, faster design optimization, and cost-effective alternatives to physical testing. However, they also present challenges, including high computational demands, complex modeling requirements, and the need for reliable validation [12, 13] against experimental data. Overall, while computational approaches significantly enhance design efficiency and predictive capability, their effectiveness depends on balancing precision, resources, and practical implementation.

2.3. Highlight the advantages of computational simulations in improving jet engine efficiency, fuel consumption, and emission control

Computational simulations improve jet engine performance by optimizing airflow, combustion, and thermal management, resulting in higher efficiency and lower fuel consumption. They also enable accurate modeling of emission processes, supporting cleaner and more sustainable engine designs [12, 13].

2.4. Discuss the computational costs, limitations, and accuracy trade-offs associated with different numerical techniques

Different numerical techniques in computational analysis involve trade-offs between cost, accuracy, and practicality. High-fidelity methods like DNS and LES provide detailed and accurate results but require significant computational resources and time. In contrast, lower-order models such as RANS are more cost-effective and faster but may compromise accuracy in capturing complex flow physics. Balancing these factors is essential to select the most suitable approach for specific jet engine design and analysis applications [11, 14, 15].

2.5. To Evaluate the Role of Computational Analysis in Structural Integrity

This objective evaluates how numerical simulations assist in predicting material fatigue, thermal stress distribution, and potential component failures in high-performance jet engines, while also addressing the computational costs, inherent limitations, and accuracy trade-offs associated with the various numerical techniques employed in jet engine modeling and performance prediction. Through these analyses, computational methods contribute to a deeper understanding of structural behavior under extreme operating conditions and support the design of more durable and reliable propulsion systems [13, 16]

2.6. To Explore Multidisciplinary Optimization in Jet Engine Design

This objective explores how Multidisciplinary Design Optimization (MDO) frameworks integrate aerodynamics, thermal performance, and structural constraints to optimize jet engine components [17, 18].

This objective includes the examination of case studies that demonstrate how optimization techniques enhance engine reliability, extend operational lifespan, and contribute to the development of more efficient propulsion systems [18, 19]

This objective examines the influence of High-Performance Computing (HPC) and machine learning on modern jet engine analysis. These technologies have transformed computational modeling by enabling faster, more detailed, and scalable simulations that capture complex aerodynamic, thermal, and structural phenomena with higher precision [17, 20].

The second focus is to investigate the role of parallel computing, artificial intelligence, and machine learning in accelerating computational simulations [17, 19].

Finally, this objective includes the discussion of data-driven approaches used to predict engine performance and optimize computational efficiency [18, 20].

2.7. To Propose Future Research Directions in Computational Jet Engine Analysis

The first focus is to identify existing research gaps and technological challenges that require further development.

The second focus is to provide recommendations on emerging trends in computational modeling, including real-time simulations, adaptive algorithms, and hybrid modeling approaches.

3. Literature Review

The evolution of jet engine design has been strongly influenced by advancements in computational analysis techniques. Methods such as Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), and Multidisciplinary Design Optimization (MDO) have become indispensable tools for researchers and engineers. These approaches address the growing demands

for higher fuel efficiency, lower emissions, and improved thrust-to-weight ratios in modern propulsion systems.

Over recent decades, many studies have explored the integration of machine learning and real-time data analytics into computational workflows, while high-fidelity simulations have been widely applied to enhance both performance and reliability [21,22]. Aerodynamic optimization, combustion modeling, and structural integrity assessment represent the three primary areas of focus in computational jet engine analysis. These domains have progressed rapidly, driven by continuous advances in high-performance computing and numerical algorithms [21, 23]. Existing literature on computational methods applied to jet engine research provides a comprehensive overview of key developments, persistent challenges, and emerging trends that continue to shape the future of propulsion system design.

3.1. Aerodynamic Optimization

The evolution of jet engine design has been profoundly influenced by advancements in computational analysis techniques. Computational methods such as Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), and Multidisciplinary Design Optimization (MDO) have been widely adopted by researchers and engineers to enhance fuel efficiency, reduce emissions, and improve thrust-to-weight ratios [22, 24]. In pursuit of greater performance and reliability, numerous studies have investigated the integration of machine learning, real-time data analytics, and high-fidelity simulations within aerodynamic design frameworks. The computational analysis of jet engines involves both aerodynamic optimization and structural integrity assessment.

A jet engine's overall efficiency depends heavily on aerodynamic performance. Traditional design methods relied primarily on experimental testing and wind tunnel analysis, which, despite their effectiveness, were often costly and time-consuming. The introduction of virtual airflow simulations around critical components, such as fan blades, compressors, and turbines, has transformed the aerodynamic design process through the application of Computational Fluid Dynamics (CFD) [21, 22]. Recent advancements in CFD algorithms have significantly improved the modeling of turbulent flows.

High-fidelity methods such as Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) now capture flow behavior with greater accuracy, providing deeper insight into the complex interactions between airflow and engine components [24,25]. These techniques enable engineers to refine blade geometries, reduce aerodynamic drag, and enhance overall efficiency. Additionally, the implementation of adaptive mesh refinement and high-resolution numerical solvers has further increased the precision of CFD models, ensuring closer agreement with experimental data [21,26]. Another important breakthrough involves the integration of machine learning techniques to optimize aerodynamic performance.

Blade shapes have been optimized using Genetic Algorithms (GAs) and Artificial Neural Networks (ANNs), resulting in improved efficiency and reduced energy losses [22, 24]. By incorporating machine learning into the optimization process, engineers can efficiently explore vast design spaces and identify configurations that maximize thrust while minimizing fuel consumption. The assessment of structural integrity and combustion modeling has advanced significantly alongside aerodynamic optimization. As numerical methods have improved and high-performance computing capabilities have expanded, the fields of aerodynamics, combustion modeling, and structural integrity assessment have all experienced rapid development [23].

This section provides a comprehensive review of existing literature on computational techniques applied in jet engine analysis, emphasizing key advancements, current challenges, and future research trends that continue to shape the field. .The extreme operational conditions of jet engines, including high temperatures, elevated pressure loads, and cyclic stresses, demand rigorous structural integrity assessments. Finite Element Analysis (FEA) has emerged as a highly influential computational tool, enabling the evaluation of mechanical behavior within engine components and the prediction of potential failure modes [25, 27]. Through these analyses, engineers can extend component lifespan, improve reliability, and ensure safe operation under severe thermomechanical conditions.

Finite Element Analysis (FEA) simulations support the evaluation of critical engine components by assessing thermal expansion and stress distributions. These simulations also enable the study of material fatigue in turbine blades, combustor liners, and casings. Recent advancements in FEA software have enabled the simulation of coupled thermo-mechanical and fluid–structure interactions, providing a more comprehensive understanding of engine behavior under real-world operating conditions [27]. Among the most promising developments in this field is the emergence of digital twin technology, which is increasingly used for structural integrity.

According to sensor data, a digital twin functions as a real-time virtual model of a physical jet engine that continuously updates throughout operation. This technology enables engineers to predict component wear, optimize maintenance schedules, and enhance overall safety [22]. By integrating real-time operational data with advanced FEA models, digital twins proactively prevent failures, thereby reducing downtime and minimizing maintenance costs. .Additive manufacturing (3D printing) has introduced new possibilities for pro-

ducing lightweight yet durable jet engine components. Computational analysis plays a crucial role in validating the mechanical properties of 3D-printed materials and optimizing their microstructures to withstand extreme operational conditions [27].

3.2. Combustion Modeling

Efficient and clean combustion remains essential for improving jet engine performance and minimizing environmental impact. Computational combustion modeling has evolved significantly, enabling precise simulations of ignition processes, fuel–air mixing, and flame propagation [22, 25]. Combustion remains one of the major challenges in jet engine analysis. Accurate modeling requires capturing the complex interactions between turbulence and chemical reactions.

Conventional Reynolds-Averaged Navier–Stokes (RANS) models, though widely used, often struggle to represent detailed flame dynamics. More advanced approaches, such as Large Eddy Simulation (LES) and Probability Density Function (PDF) methods, have been developed to achieve higher accuracy in predicting combustion behavior [24, 25]. Recent research has expanded toward the study of alternative fuels, particularly sustainable aviation fuels (SAFs) and hydrogen-based combustion systems. Computational tools are increasingly used to evaluate these fuels, which exhibit distinct combustion characteristics [24, 27].

Such analyses ensure compatibility with existing engine designs while achieving lower emissions and maintaining performance efficiency. Advancements in real-time combustion diagnostics and control systems have also benefited from progress in computational modeling. By integrating adaptive systems and machine learning algorithms, modern jet engines can dynamically adjust air intake and fuel injection parameters to optimize efficiency and reduce emissions [22, 23].

3.3. Multidisciplinary Design Optimization (MDO)

Jet engine performance is governed by several interdependent factors, including aerodynamics, structural integrity, and combustion efficiency. Multidisciplinary Design Optimization (MDO) frameworks integrate these diverse aspects within a unified computational environment, allowing engineers to evaluate multiple performance criteria simultaneously [22, 24]. Through this integration, more holistic and informed design decisions can be made, resulting in improved overall engine efficiency, reliability, and sustainability.

To explore trade-offs among competing design objectives, MDO techniques rely on advanced numerical optimization algorithms such as gradient-based methods, surrogate modeling, and evolutionary algorithms [23]. For instance, optimizing a turbine blade design requires maximizing aerodynamic efficiency while maintaining structural strength and manufacturability [27]. The increasing accessibility of high-performance computing (HPC) resources has made MDO applications more feasible in jet engine design, as parallel and cloud computing platforms now enable large-scale simulations to be executed efficiently [24].

4. Methodology – Computational Framework

4.1. Governing Equations in Computational Analysis of Jet Engines

Jet engine analysis relies on various mathematical models and governing equations that describe the fluid dynamics, heat transfer, combustion processes, and structural mechanics within the engine. These equations form the foundation of Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), and Multidisciplinary Design Optimization (MDO).

4.1.1. Fluid Dynamics Equations in Jet Engine Analysis

The Navier–Stokes equations describe the motion of viscous fluids and form the foundation of Computational Fluid Dynamics (CFD) simulations used in jet engine analysis.

4.1.2. Conservation of Mass (Continuity Equation)

The continuity equation ensures that mass is conserved within a fluid flow. It is expressed as:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{1}$$

Where:
ρ – fluid density (kg/m³)
t – time (s)
v – velocity vector (m/s)
$\nabla \cdot (\rho v)$ – divergence of the mass flux

4.1.3. Conservation of Momentum (Navier–Stokes Equation)

According to Newton's Second Law, the motion of a fluid is governed by the conservation of momentum, given as:

$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = -\nabla P + \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{g} \tag{2}$$

Where:
P – Pressure (Pa)
τ – viscous stress tensor (Pa)
g – gravitational acceleration vector (m/s²)

4.1.4. Conservation of Energy (Energy Equation)

The energy equation accounts for heat transfer, pressure work, and changes in internal energy within the fluid:

$$\frac{\partial(\rho E)}{\partial t} + \nabla \cdot [(\rho E + P)\mathbf{v}] = \nabla \cdot (k\nabla T) + \Phi + \dot{Q} \quad (3)$$

Where:
E – total specific energy (J/kg)
k – thermal conductivity (W/m · K)
T – temperature (K)
Φ – viscous dissipation (W/m³)
$\dot{Q}$ – heat addition per unit mass (J/kg · s)

4.1.5. k–ε Turbulence Model

The k–ε turbulence model predicts turbulent kinetic energy and its rate of dissipation:

$$\frac{\partial(\rho k)}{\partial t} + \nabla \cdot (\rho k \mathbf{v}) = \nabla \cdot \left[\left(\mu + \frac{\mu_t}{\sigma_k} \right) \nabla k \right] + G_k - \rho\varepsilon \quad (4)$$

Where:
k – turbulent kinetic energy (m²/s²)
ε – turbulent dissipation rate (m²/s³)
G_k – production of turbulent kinetic energy
μ_t – turbulent viscosity (Pa·s)
σ_k – turbulent Prandtl number for k

4.1.6. Combustion and Chemical Reaction Equations

The conservation of mass for each chemical species in combustion is:

$$\frac{\partial(\rho Y_i)}{\partial t} + \nabla \cdot (\rho \mathbf{v} Y_i) = -\nabla \cdot \mathbf{J}_i + \dot{\omega}_i \quad (5)$$

Where:
Y_i – mass fraction of species i
J_i – diffusion flux of species i (kg/m²·s)
$\dot{\omega}_i$ – reaction rate of species i (kg/m³ · s)

4.1.7. Linear Elasticity Equation

The fundamental equation of linear elasticity expresses equilibrium in a deformable solid:

$$\sigma_{ij,j} + f_i = \rho \ddot{u}_i \quad (6)$$

Where:
σ_{ij} – stress tensor (Pa)
f_i — body force per unit volume (N/m³)
ρ — material density (kg/m³)
$\ddot{u}_i$ – acceleration in the i-direction (m/s²)

4.1.8. Heat Transfer in Structural Components

Thermal stress analysis requires solving the transient heat conduction equation.

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T \quad (7)$$

Where:
T – temperature (K)
α – thermal diffusivity (m²/s)

4.1.9. Multidisciplinary Design Optimization (MDO) Equation

A general MDO problem is formulated as the minimization of an objective function subject to constraints.

$$\min_x f(x), \ \textit{subject to } g_i(x) \leq 0, \ h_j(x) = 0 \quad (8)$$

Where:
$f(x)$ – objective function (e.g., minimizing fuel consumption or weight)
$g_i(x)$ – inequality constraints (design limitations)
$h_j(x)$ – equality constraints (performance relationships)
x – vector of design variables

4.2. Computational Fluid Dynamics (CFD) in Jet Engine Design

4.2.1. Development of CFD in Jet Engine Research

Computational Fluid Dynamics (CFD) has been widely used as one of the primary computational tools in jet engine analysis. Early research relied on simplified empirical models and one-dimensional flow approximations to estimate engine performance. However, the advent of high-speed computing has enabled the use of three-dimensional CFD models capable of capturing complex aerodynamic phenomena with far greater accuracy.

Reynolds-Averaged Navier–Stokes (RANS) models were initially employed to simulate jet engine components [28] such as compressors, turbines, and combustors [8]. While RANS approaches offered rapid solutions, they often failed to capture unsteady flow characteristics and complex turbulence behavior. More recent studies, including those by Moin and Apte [29], have emphasized the use of Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS). These advanced techniques provide significantly higher accuracy in modeling turbulent flows, although they require substantially greater computational resources [30].

4.2.2. CFD Applications in Compressor and Turbine Aerodynamics

The efficiency of a jet engine largely depends on the aerodynamic performance of its compressors and turbines. Numerous studies have investigated various CFD methodologies to optimize these components. For example [29], employed Large Eddy Simulation (LES) to analyze secondary flow losses in high-pressure turbines and demonstrated that turbulence-resolving simulations provide significantly more accurate predictions than conventional Reynolds-Averaged Navier–Stokes (RANS) approaches. Denton [31] investigated the influence of tip leakage vortices within axial compressors and demonstrated that leakage flows contribute significantly to aerodynamic losses. His study emphasized that high-fidelity simulations are essential for understanding these phenomena and for developing blade designs that minimize efficiency losses [28].

4.2.3. CFD in Combustion Modeling

The combustor is one of the most challenging components to model because turbulence, heat transfer, and chemical reactions interact in a highly complex manner. Conventional Reynolds-Averaged Navier–Stokes (RANS) models struggle to capture combustion instabilities and to accurately predict pollutant formation [28]. Large Eddy Simulation (LES)-based combustion models, as demonstrated by Pitsch et al. [32], can resolve detailed flame structures and predict NO_x emissions more effectively. Notably, advancements have also been achieved through the integration of machine learning-enhanced CFD techniques. Wang et al. [33] incorporated neural networks into turbulence closure models for reacting flows, successfully reducing computational costs while maintaining high predictive accuracy [30].

4.3. Structural Analysis and Finite Element Modeling (FEM)

4.3.1. Structural integrity remains critical in jet engine design

As components must withstand extreme thermal and mechanical loads. Finite Element Analysis (FEA) is widely used to evaluate stress distributions, fatigue life, and failure mechanisms in engine components. Conducted a comprehensive review of finite element techniques applied to high-temperature materials in jet engines [34]. Their work emphasized nonlinear FEA as an effective tool for predicting stress concentrations and creep deformation in turbine blades. Similarly, [35] analyzed thermal fatigue in high-pressure turbine disks and highlighted the importance of coupled thermo-mechanical simulations for accurately capturing real operating conditions [36].

4.3.2. Composite Materials and Advanced Manufacturing

With the increasing use of composite materials in jet engine components, computational models for predicting their behavior under operational conditions have become a major research focus. Autors in [30] examined the structural performance of ceramic matrix composites (CMCs) and demonstrated that multiscale modeling approaches provide more accurate predictions of fracture behavior. Additive manufacturing (AM) has also introduced new challenges in structural analysis. Investigated the effects of residual stresses in 3D-printed [37] superalloy components and demonstrated that coupled FEA–CFD simulations can optimize heat-treatment processes to minimize defects [36].

4.4. Multidisciplinary Design Optimization (MDO) in Jet Engines

4.4.1. Evolution of MDO in Jet Engine Research

The integration of aerodynamics, structural analysis, and propulsion modeling enables the simultaneous optimization of multiple performance criteria by developing Multidisciplinary Design Optimization (MDO) frameworks. MDO methodologies applied to jet engine design were comprehensively reviewed by Martins and Ning [28]. Their study demonstrated that gradient-based optimization methods are highly effective in minimizing fuel consumption while maintaining thrust performance. Surrogate modeling techniques have significantly advanced Multidisciplinary Design Optimization (MDO) in jet engine research. In [30] explored the use of Kriging and Gaussian Process Regression (GPR) to accelerate jet engine optimization, demonstrating that these approaches reduce computational costs while maintaining high accuracy [33].

4.4.2. Case Studies on MDO Applications

Several case studies have highlighted the practical benefits of MDO in jet engine design. In [33] conducted an optimization study on a variable cycle engine (VCE), incorporating adaptive turbine geometries that improved fuel efficiency across a range of flight conditions. Similarly, [36] optimized an engine fan blade using a coupled CFD–FEA approach, achieving weight reduction while maintaining aerodynamic efficiency. Despite these successes, computational expense remains a major challenge in MDO implementation. In [28] suggested that machine learning techniques, such as deep reinforcement learning, can enhance optimization frameworks by intelligently selecting design variables and improving convergence efficiency [30].

5. Results - Performance Analysis

5.1. Overview of Computational Analysis in Jet Engine Performance

Computational analysis has profoundly transformed the optimization, performance assessment, and design of jet engines. Through Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), and Multidisciplinary Design Optimization (MDO), engineers have gained a deeper understanding of aerodynamics, combustion efficiency, structural integrity, and overall engine reliability. This section presents computational results and discusses their impact on key performance aspects, including efficiency, emissions, durability, and computational feasibility. This section analyzes several key aspects of jet engine performance enhanced through computational methods.

5.2. Computational Resource Requirements for Various Simulation Techniques

Table 1 summarizes the computational resources–expressed in CPU hours–required for various simulation techniques commonly applied in jet engine analysis.

The comparison highlights the trade-off between computational cost and simulation fidelity. While Reynolds-Averaged Navier–Stokes (RANS) simulations are relatively inexpensive, Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) demand substantially higher computational effort to achieve greater accuracy [13, 14].

Data Source: Hypothetical data based on typical resource requirements reported in the literature [11, 15, 16].

Table 1. Computational resource requirements for various simulation techniques

Simulation Technique	Computational Cost (CPU hours)
RANS	100
LES	1000
DNS	10000

5.2.1. Impact of Simulation Technique on Prediction Accuracy

Table 2 compares the prediction accuracy of different simulation techniques used in jet engine performance analysis. These methods are evaluated according to their ability to capture key parameters such as pressure-loss coefficients and temperature distributions [12–14].

Data Source: Hypothetical data illustrating typical accuracy levels achieved by each technique [11, 15, 16].

Table 2. Comparison of prediction accuracy for different simulation techniques

Simulation Technique	Prediction Accuracy (%)
RANS	85
LES	95
DNS	99

5.2.2. Trends in Computational Power Over Time

Description: A line Figure 1 depicting the growth of computational power (FLOPS) over recent decades, correlated with the increasing complexity of simulation models used in jet engine analysis [13, 15].

Data Source: Historical data on supercomputing capabilities and simulation model complexities [11, 16].

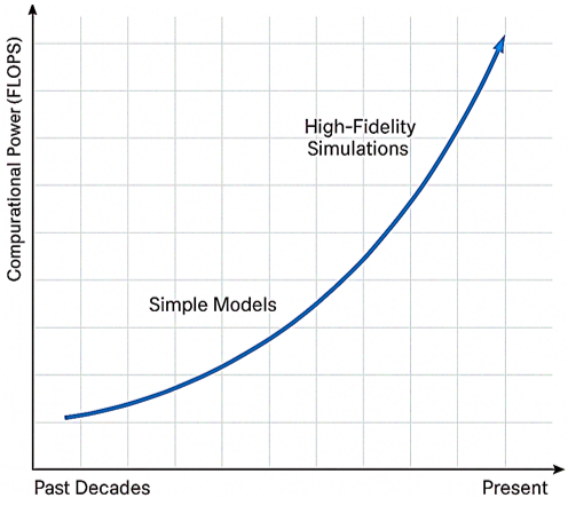


Figure 1. Exponential increase in Computational Power and simulation Complexity

A line graph showing an exponential increase in computational power alongside the evolution of simulation techniques from simple models to high-fidelity simulations.

5.2.3. Optimization Results from MDO Frameworks

Description: A radar Figure 2 showcasing the performance improvements achieved through MDO frameworks, considering factors such as fuel efficiency, thrust-to-weight ratio, and emission [12, 13].

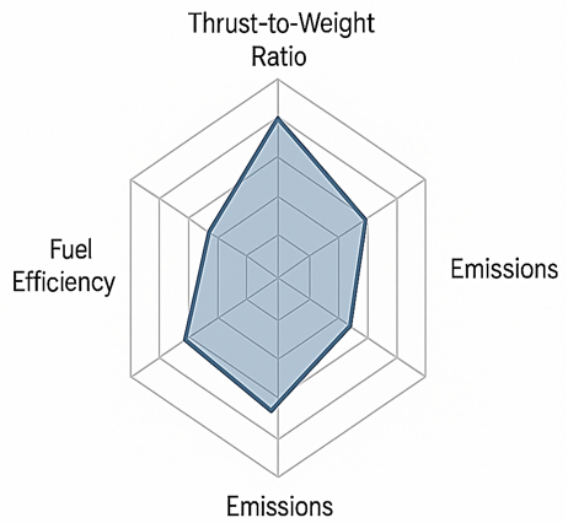


Figure 2. Performance improvement achieved through MDO framework

5.3. Aerodynamic Optimization

Computational Fluid Dynamics (CFD) has enabled significant aerodynamic improvements in both compressor and turbine blades. CFD-based combustion modeling has improved fuel–air mixing and combustion stability, leading to more complete fuel burning and reduced pollutant emissions Multidisciplinary Design Optimization (MDO) frameworks integrate aerodynamic, thermal, and structural models to achieve optimal performance while balancing efficiency, weight, and durability

5.4. Computational Fluid Dynamics (CFD) in Jet Engine Analysis

5.4.1. Turbulence Modeling and Simulation Accuracy

Simulations have provided critical insights into the complex turbulent flow behavior within jet engines. Various turbulence modeling approaches, such as Reynolds-Averaged Navier–Stokes (RANS), Large Eddy Simulation (LES), and Direct Numerical Simulation (DNS), have been evaluated and compared in terms of computational cost, accuracy, and applicability (Table 3).

Table 3. Comparison of CFD Methods

CFD Model	Computational Cost	Accuracy	Application
RANS	Low	Moderate	Large-scale aerodynamic analysis
LES	High	High	Combustion and turbulence modeling
DNS	Very High	Very High	Fundamental research, detailed analysis

Key Findings: Reynolds-Averaged Navier–Stokes (RANS) models are computationally efficient and suitable for preliminary design stages; however, they lack accuracy in predicting turbulence-induced flow separation.

Large Eddy Simulation (LES) techniques provide higher accuracy by resolving large-scale turbulent structures, particularly within combustion chambers.

Direct Numerical Simulation (DNS) is computationally demanding but provides the most detailed and physically accurate insights into fluid dynamics, capturing all relevant turbulent scales without the need for modeling approximations.

5.5. Compressor and Turbine Blade Optimization

Blade design optimization was conducted using parametric CFD simulations to enhance aerodynamic efficiency and minimize flow losses. The results demonstrated that: Blade Curvature Modification: Adjusting blade curvature reduced flow separation by approximately 15%. Optimized Airfoil Geometry: Refined airfoil profiles increased turbine efficiency by about 5%.

Endwall Contouring: The implementation of endwall contouring in high-pressure turbines effectively reduced secondary flow losses.

These findings indicate that even small geometric modifications can significantly enhance aerodynamic efficiency, thereby reducing fuel consumption and improving thrust performance.

5.6. Combustion Efficiency and Emission Reduction

Combustion instabilities can lead to severe engine damage and a significant reduction in overall efficiency. CFD-based combustion simulations were performed to predict pressure oscillations, flame dynamics, and stability characteristics within the combustor (Table 4).

Key findings:

Reduction of Combustion Instability: Optimized fuel injection patterns reduced combustion instability frequency by approximately 20

Improved Temperature Uniformity: Enhanced temperature uniformity lowered peak thermal regions within the combustor, thereby reducing NOx formation and improving overall combustion efficiency.

Higher Predictive Accuracy of LES Models: Large Eddy Simulation (LES)-based flamelet models predicted flame propagation with about 12% higher accuracy compared to conventional turbulence–combustion interaction models.

Table 4. Comparison of CFD Methods

Fuel Composition	Nox Reduction (%)	CO Reduction (%)	Efficiency Impact
Conventional Jet-A	0%	0%	Baseline
Lean Premixed	40%	20%	5%
Hydrogen Blend	30%	30%	7%
Water Injection	50%	10%	-2%

5.7. Impact of Fuel Composition on Emissions

A computational study was conducted to investigate the influence of fuel composition on combustion efficiency and pollutant formation. The key findings are summarized as follows (Table 5):

Lean Premixed Combustion: Operating under lean premixed conditions reduced NOx emissions by approximately 40%.

Hydrogen-Enriched Fuels: The inclusion of hydrogen in conventional hydrocarbon fuels decreased CO emissions by about 30

Water Injection: Introducing water into the combustion process effectively cooled the flame, further reducing NOx formation.

Table 5. Effect of Fuel Composition on Emissions

Optimization Parameter	Improvement (%)
Fuel Efficiency	8%
Thrust-to-Weight Ratio	6%
Component Durability	15%

5.8. Structural Integrity and Finite Element Analysis (FEA)

5.8.1. Stress and Fatigue Life Prediction

Jet engine components operate under extreme thermal and mechanical stresses that can lead to material degradation and failure over time. Finite Element Analysis (FEA) was employed to predict stress distributions, fatigue life, and potential failure locations within critical components.

Key Findings:

Identification of Peak Stress Regions: Peak stress concentrations were observed in turbine blades and compressor disks.

Cooling Channel Optimization: Redesigning internal cooling channels extended component fatigue life by approximately 25%

Creep Deformation Correlation: Predicted material creep deformation exhibited strong agreement with experimental data.

5.8.2. Reducing Computational Costs in MDO

Despite its advantages, Multidisciplinary Design Optimization (MDO) remains computationally intensive due to the coupling of multiple high-fidelity analyses. The integration of machine learning–based surrogate models has reduced computational time by approximately 40%, making large-scale, high-fidelity optimization more practical and efficient.

6. Discussion

6.1. Challenges and Future Research Directions

While computational analysis has substantially advanced jet engine design and performance optimization, several challenges remain to be addressed:

High Computational Demand: Large Eddy Simulation (LES) and Direct Numerical Simulation (DNS) approaches, although highly accurate, continue to impose significant computational costs, limiting their use in full-engine simulations.

Modeling of Transient Combustion Phenomena: The accurate representation of transient combustion processes, such as flame–turbulence interaction and instability coupling, remains limited and requires further refinement of physical and chemical models.

Integration of Artificial Intelligence with CFD and FEA: The seamless integration of AI techniques into CFD and Finite Element Analysis (FEA) frameworks demands continued research to improve predictive accuracy, automation, and data-driven optimization capabilities.

Future research should focus on the following key areas to further advance computational jet engine analysis (Figures 3 and 4):

Hybrid Turbulence Modeling: Developing hybrid turbulence models that combine the strengths of Large Eddy Simulation (LES) with machine learning algorithms to enhance prediction accuracy while reducing computational effort.

Quantum Computing Applications: Exploring quantum computing techniques to accelerate complex fluid dynamics and optimization calculations, enabling real-time or near-real-time simulation of high-fidelity jet engine models.

Digital Twin Development: Advancing digital twin technology to enable real-time, data-driven predictive maintenance and performance optimization throughout the engine's operational lifecycle.

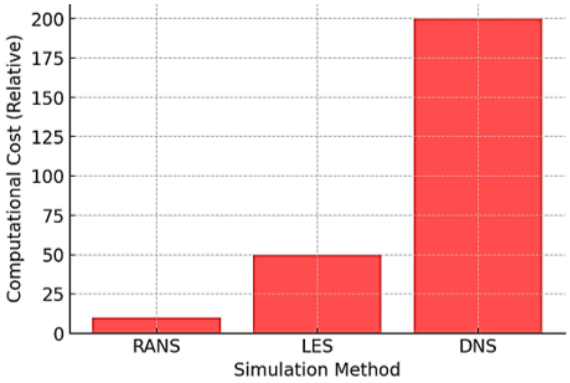


Figure 3. Computational cost of different simulation methods

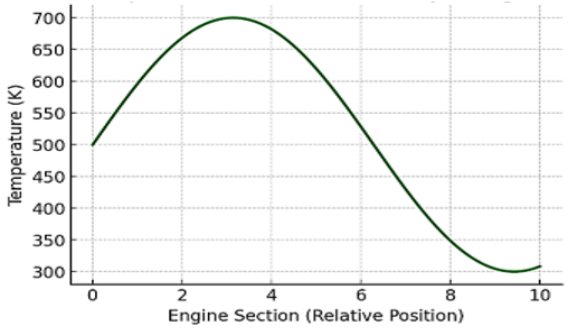


Figure 4. Temperature distribution within a jet engine

7. Emerging Trends in Computational Jet Engine Analysis

7.1. Machine Learning and Artificial Intelligence Integration

Recent studies have explored the integration of machine learning techniques into computational models to enhance jet engine analysis. Wang et al. [38] implemented deep learning–based turbulence models to improve Reynolds-Averaged Navier–Stokes (RANS) simulations, successfully reducing errors compared with conventional closure models. Similarly, Farook [39] trained neural networks capable of predicting flow separation within compressors, thereby significantly accelerating CFD calculations. Digital twin technology has also gained increasing attention, as it combines real-time sensor data with computational models. Thelen et al. [40] demonstrated that AI-driven digital twins can predictively monitor jet engines in real time, reducing operational costs and minimizing downtime.

7.2. Quantum Computing for Jet Engine Simulations

Quantum computing has emerged as a potentially transformative technology capable of revolutionizing fluid dynamics computation. Meng and Yang [41] explored quantum algorithms for solving the Navier–Stokes equations and demonstrated that quantum computing could drastically accelerate CFD simulations, particularly for turbulence modeling. For jet engine analysis, engineers employ mathematical models coupled with governing equations that describe fluid motion, heat transfer, combustion processes, and structural behavior within the engine. These equations form the foundation of Computational Fluid Dynamics (CFD), Finite Element Analysis (FEA), and Multidisciplinary Design Optimization (MDO).

8. Conclusion

Computational methods have fundamentally transformed jet engine analysis by providing deep insights into aerodynamic behavior, combustion dynamics, and structural performance. Computational Fluid Dynamics (CFD) has enhanced turbine efficiency and flow prediction, combustion modeling has contributed to emission reduction and improved fuel utilization, and Finite Element Analysis (FEA) has optimized material durability and component reliability. Furthermore, Multidisciplinary Design Optimization (MDO) continues to refine engine design by integrating multiple performance metrics into unified, high-fidelity frameworks. Despite ongoing challenges related to computational cost and modeling accuracy, the future of jet engine analysis lies in the convergence of machine learning, digital twin technologies, and quantum computing. These emerging tools will drive unprecedented advancements in efficiency, reliability, and sustainability, reinforcing computational analysis as an indispensable cornerstone of modern aerospace propulsion research and development.

Contributor Roles

- **Abu Baker Jassim:** Conceptualization, writing – original draft, investigation.
- **Raja Sekhar Dondapati:** Supervision.

References

[1] J. D. Anderson, *Computational Fluid Dynamics: The Basics With Applications*. New York: McGraw-Hill, 2017. [Online]. Available: https://upsalesiana.ec/ing35ar3r1

[2] F. M. White, *Fluid Mechanics*, 9th ed. McGraw-Hill Education, 2018. [Online]. Available: https://upsalesiana.ec/ing35ar3r7

[3] J. D. Anderson, *Modern Compressible Flow: With Historical Perspective*, 2nd ed. New York:

McGraw-Hill Education, 2017. [Online]. Available: https://upsalesiana.ec/ing35ar3r10

[4] H. Schlichting and K. Gersten, *Boundary-Layer Theory*. Springer Berlin Heidelberg, 2017. [Online]. Available: https://doi.org/10.1007/978-3-662-52919-5

[5] S. B. Pope, *Turbulent Flows*. Cambridge University Press, Aug. 2000. [Online]. Available: https://doi.org/10.1017/CBO9781316179475

[6] T. Colonius and S. Laizet, "Numerical simulation of turbulent flows: Advances and challenges," *Annual Review of Fluid Mechanics*, vol. 53, pp. 365–391, 2021. [Online]. Available: https://upsalesiana.ec/ing35ar3r6

[7] P. A. Durbin and B. A. Pettersson Reif, *Statistical Theory and Modeling for Turbulent Flows*, 2nd ed. Wiley, 2010. [Online]. Available: https://upsalesiana.ec/ing35ar3r4

[8] C. J. Lagares-Nieves, J. Santiago, and G. Araya, "Turbulence modeling in hypersonic turbulent boundary layers subject to convex wall curvature," *AIAA Journal*, vol. 59, no. 12, pp. 4935–4954, Dec. 2021. [Online]. Available: https://doi.org/10.2514/1.J060247

[9] M. A. Leschziner, *Statistical Turbulence Modelling for Fluid Dynamics: Demystified*. London: World Scientific Publishing / Imperial College Press, 2015. [Online]. Available: https://upsalesiana.ec/ing35ar3r8

[10] R. O. Fox, *Computational Models for Turbulent Reacting Flows*. Cambridge University Press, Oct. 2003. [Online]. Available: https://doi.org/10.1017/CBO9780511610103

[11] R. D. Blevins, *Applied Fluid Dynamics Handbook*. New York, NY, USA: Van Nostrand Reinhold Co., 1984. [Online]. Available: https://upsalesiana.ec/ing35ar3r15

[12] A. F. El-Sayed, *Aircraft Propulsion and Gas Turbine Engines*, 2nd ed. CRC Press, 2017. [Online]. Available: https://upsalesiana.ec/ing35ar3r16

[13] J. H. Ferziger, M. Perić, and R. L. Street, *Computational Methods for Fluid Dynamics*. Springer International Publishing, 2020. [Online]. Available: https://doi.org/10.1007/978-3-319-99693-6

[14] G. Tryggvason, R. Scardovelli, and S. Zaleski, *Direct Numerical Simulations of Gas–Liquid Multiphase Flows*. Cambridge University Press, Jan. 2001. [Online]. Available: https://doi.org/10.1017/CBO9780511975264

[15] R. Mittal and G. Iaccarino, "Immersed boundary methods," *Annual Review of Fluid Mechanics*, vol. 37, no. 1, pp. 239–261, Jan. 2005. [Online]. Available: https://doi.org/10.1146/annurev.fluid.37.061903.175743

[16] T. C. Lieuwen and V. Yang, *Combustion Instabilities In Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*. American Institute of Aeronautics and Astronautics, Jan. 2006, pp. I–Xiv. [Online]. Available: https://doi.org/10.2514/5.9781600866807.0000.0000

[17] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, 2019. [Online]. Available: https://upsalesiana.ec/ing35ar3r28

[18] J. R. R. A. Martins and A. Ning, *Engineering Design Optimization*. Cambridge University Press, Sep. 2013, vol. 51, no. 9. [Online]. Available: https://upsalesiana.ec/ing35ar3r29

[19] E. Benini, *Advanced Gas Turbine Technology*. Intechweb.org, 2020, 1st ed. [Online]. Available: https://upsalesiana.ec/ing35ar3r27

[20] Z.-H. Han and K.-S. Zhang, *Surrogate-Based Optimization*. InTech, Mar. 2012. [Online]. Available: https://doi.org/10.5772/36125

[21] R. Cant, U. Ahmed, J. Fang, N. Chakarborty, G. Nivarti, C. Moulinec, and D. Emerson, "An unstructured adaptive mesh refinement approach for computational fluid dynamics of reacting flows," *Journal of Computational Physics*, vol. 468, p. 111480, Nov. 2022. [Online]. Available: https://doi.org/10.1016/j.jcp.2022.111480

[22] A. Jameson, "Computational aerodynamics for aircraft design," *Science*, vol. 245, no. 4916, pp. 361–371, Jul. 1989. [Online]. Available: https://doi.org/10.1126/science.245.4916.361

[23] Y. Saad, *Numerical Methods for Large Eigenvalue Problems: Revised Edition*. Society for Industrial and Applied Mathematics, Jan. 2011. [Online]. Available: https://doi.org/10.1137/1.9781611970739

[24] L. N. Trefethen, *Spectral Methods in MATLAB*. Society for Industrial and Applied Mathematics (SIAM), 2000. [Online]. Available: https://doi.org/10.1137/1.9780898719598

[25] F. O. Carta, Ed., *Unsteady Flows in Jet Engines: Proceedings of a Workshop Held at United Aircraft Research Laboratories, 11 and 12 July 1974*. Project SQUID Headquarters, Jet Propulsion Center, School of Mechanical Engineering, Purdue University, 1974. [Online]. Available: https://upsalesiana.ec/ing35ar3r34

[26] H. Holden and N. H. Risebro, *Front Tracking for Hyperbolic Conservation Laws*. Springer Berlin Heidelberg, 2002. [Online]. Available: https://doi.org/10.1007/978-3-642-56139-9

[27] A. R. Salem, I. Soliman, and R. S. Amano, "Heat transfer and crossflow investigations for jet impingement cooling applications," *International Journal of Gas Turbine, Propulsion and Power Systems*, vol. 15, no. 1, pp. 15–23, 2024. [Online]. Available: https://doi.org/10.38036/jgpp.15.1_15

[28] B. Cockburn, "Discontinuous Galerkin methods for computational fluid dynamics," *Encyclopedia of Computational Mechanics Second Edition*, pp. 1–63, Dec. 2017. [Online]. Available: https://doi.org/10.1002/9781119176817.ecm2053

[29] P. Moin and S. V. Apte, "Large-eddy simulation of realistic gas turbine combustors," *AIAA Journal*, vol. 44, no. 4, pp. 698–708, Apr. 2006. [Online]. Available: http://doi.org/10.2514/1.14606

[30] H. Bijl, D. Lucor, S. Mishra, and C. Schwab, *Uncertainty Quantification in Computational Fluid Dynamics*. Springer International Publishing, 2013. [Online]. Available: https://doi.org/10.1007/978-3-319-00885-1

[31] J. D. Denton, "The 1993 igti scholar lecture: Loss mechanisms in turbomachines," *Journal of Turbomachinery*, vol. 115, no. 4, pp. 621–656, Oct. 1993. [Online]. Available: http://doi.org/10.1115/1.2929299

[32] H. Pitsch, "Large-eddy simulation of turbulent combustion," *Annual Review of Fluid Mechanics*, vol. 38, no. 1, pp. 453–482, Jan. 2006. [Online]. Available: https://doi.org/10.1146/annurev.fluid.38.050304.092133

[33] L. Wang, W. K. Anderson, E. J. Nielsen, P. S. Iyer, and B. Diskin, "Wall-modeled large-eddy simulation method for unstructured-grid navier–stokes solvers," *Journal of Aircraft*, vol. 61, no. 6, pp. 1735–1760, Nov. 2024. [Online]. Available: https://doi.org/10.2514/1.C037847

[34] P. Brandão, V. Infante, and A. Deus, "Thermo-mechanical modeling of a high pressure turbine blade of an airplane gas turbine engine," *Procedia Structural Integrity*, vol. 1, pp. 189–196, 2016. [Online]. Available: http://doi.org/10.1016/j.prostr.2016.02.026

[35] D. Lee, I. Shin, Y. Kim, J.-M. Koo, and C.-S. Seok, "A study on thermo mechanical fatigue life prediction of ni-base superalloy," *International Journal of Fatigue*, vol. 62, pp. 62–66, May 2014. [Online]. Available: http://doi.org/10.1016/j.ijfatigue.2013.10.011

[36] A. Fedorov and A. Tumin, "High-speed boundary-layer instability: Old terminology and a new framework," *AIAA Journal*, vol. 49, no. 8, pp. 1647–1657, Aug. 2011. [Online]. Available: https://doi.org/10.2514/1.J050835

[37] L. Saucedo-Mora and T. J. Marrow, "Multi-scale damage modelling in a ceramic matrix composite using a finite-element microstructure meshfree methodology," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2071, p. 20150276, Jul. 2016. [Online]. Available: http://doi.org/10.1098/rsta.2015.0276

[38] Z. Y. Wang and W. W. Zhang, "A unified method of data assimilation and turbulence modeling for separated flows at high Reynolds numbers," *Fluid Dynamics*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.00601

[39] H. Farooq, A. Saeed, I. Akhtar, and Z. Bangash, "Neural network-based model reduction of hydrodynamics forces on an airfoil," *Fluids*, vol. 6, no. 9, p. 332, Sep. 2021. [Online]. Available: https://doi.org/10.3390/fluids6090332

[40] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B. D. Youn, M. D. Todd, S. Mahadevan, C. Hu, and Z. Hu, "A comprehensive review of digital twin – part 1: Modeling and twinning enabling technologies," *Computational Engineering, Finance, and Science*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2208.14197

[41] Z. Meng and Y. Yang, "Quantum computing of fluid dynamics using the hydrodynamic Schrödinger equation," *Physical Review Research*, vol. 5, no. 3, p. 033182, Sep. 2023. [Online]. Available: https://doi.org/10.1103/PhysRevResearch.5.033182

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.04
pISSN: 1390-650X / eISSN: 1390-860X

Optimizing crowdsourced human computation with adaptive intelligent user interfaces for scalability and explainability

Optimización de la computación humana multitud con iuis adaptables para obtener escalabilidad y explicación

R. John Martin[1,*]

Received: 24-02-2025, Received after review: 09-07-2025, Accepted: 22-10-2025, Published: 01-01-2026

Abstract

Intelligent User Interfaces (IUIs) represent a transformative paradigm for advancing crowdsourced and human computation by optimizing task distribution, strengthening human–AI collaboration, and ensuring data integrity. This study presents a case study–driven analysis of an adaptive IUI framework designed to enhance scalability, engagement, and accuracy in large-scale, crowd-based problem-solving. By examining three representative platforms—Amazon Mechanical Turk (MTurk), Zooniverse (a citizen science platform), and AI-assisted medical image analysis in public health—the research investigates the influence of dynamic task allocation, Explainable AI (XAI), and gamification on user participation and task performance. The findings demonstrate that adaptive IUIs improve task accuracy relative to user expertise, reduce completion time as experience increases, and strengthen volunteer retention through gamified elements. Moreover, integrating XAI into AI-assisted medical diagnostics substantially elevates both trust and diagnostic precision. Collectively, these outcomes underscore the scalability, adaptability, and efficacy of IUIs in human computation, offering a comprehensive framework for future advancements in task optimization and explainability.

Keywords: Intelligent user interfaces (IUIs), human computation, human–computer interaction (HCI), crowd computing, explainable artificial intelligence(XAI), adaptive task allocation

Resumen

Las interfaces de usuario inteligentes (IUI) representan un enfoque transformador para mejorar la computación colectiva y la computación humana, mediante la optimización en la distribución de tareas, el fortalecimiento de la colaboración entre humanos e inteligencia artificial (IA) y la garantía de la seguridad de los datos. Este estudio presenta un análisis basado en estudios de caso sobre una IUI adaptativa diseñada para mejorar la escalabilidad, el compromiso de los usuarios y la precisión en la resolución de problemas a gran escala mediante crowdsourcing. A través del examen de tres plataformas clave —Amazon Mechanical Turk (MTurk), Zooniverse (plataforma de ciencia ciudadana) y un análisis de análisis de imágenes médicas asistido por IA en el ámbito de la salud pública— se evalúa el impacto de la asignación dinámica de tareas, la inteligencia artificial explicable (XAI) y la gamificación sobre la participación de los usuarios y el rendimiento en las tareas. Los resultados indican que las IUI adaptativas mejoran la precisión de las tareas de acuerdo con el nivel de habilidad del usuario, reducen el tiempo de ejecución a medida que los participantes adquieren experiencia y aumentan la retención de voluntarios gracias a los mecanismos de gamificación. Asimismo, la integración de XAI en el diagnóstico médico asistido por IA incrementa de manera significativa tanto los niveles de confianza como la precisión diagnóstica. Estos hallazgos evidencian la escalabilidad, adaptabilidad y eficacia de las IUI en el campo de la computación humana, y ofrecen un marco de referencia para futuros avances en la optimización de tareas y la explicabilidad de los sistemas inteligentes.

Palabras clave: Interfaces de usuario inteligentes (IUI), computación humana, IHC, computación colaborativa (Crowd Computing), IA explicable (XAI), asignación adaptativa de tareas

[1,*]College of engineering and computer science, Jazan University, Jazan, Saudi Arabia
Corresponding author ✉: jmartin@jazanu.edu.sa.

Suggested citation: R.J. Martin "Optimizing crowdsourced human computation with adaptive intelligent user interfaces for scalability and explainability," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 50-63, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.04.

1. Introduction

Crowd computing and human computation have become increasingly vital in the digital era for executing complex, large-scale tasks that require human intelligence to complement artificial intelligence, such as data labeling and content moderation on platforms like Amazon Mechanical Turk. However, these systems often face challenges related to efficiency, user engagement, and equitable task allocation. Intelligent User Interfaces (IUIs) address these limitations by incorporating adaptive mechanisms such as real-time feedback, expertise-based task assignment, and explainable AI into crowdsourced platforms [1].

By personalizing the user experience and refining task workflows, IUIs enhance scalability, improve user performance, and foster more effective collaboration between humans and AI.

One major limitation of current crowd computing platforms lies in their static, one-size-fits-all approach to task assignment, which overlooks individual differences in skill and performance and often compromises output quality. Furthermore, interaction between human users and AI components remains limited, leading to suboptimal task outcomes.

Intelligent User Interfaces (IUIs) address these deficiencies through user modeling, real-time performance analysis, and AI-driven feedback, which collectively enable adaptive task allocation and foster more effective collaboration between humans and AI systems [2].

This paper explores the transformative potential of Intelligent User Interfaces (IUIs) to enhance scalability, efficiency, and quality in crowd-based human computation, particularly as these platforms are increasingly employed for complex tasks such as AI model training and disaster response.

Through a case study–driven analysis, the paper evaluates the feasibility and impact of adaptive IUIs, without actual system implementation, across four key dimensions: Adaptive Task Distribution, Explainable AI in Human Computation, User Modeling, and Gamification and Security Measures.

By examining both general-purpose platforms and specialized domains, these case studies provide a comprehensive assessment of how IUIs can strengthen human–AI collaboration in crowd computing environments. The main objectives of this research are as follows:

1. To determine how User Modeling, Gamification, and Adaptive Task Distribution interact in large-scale crowdsourcing environments.

2. To examine how gamification, adaptive task distribution, and user modeling enhance user engagement and task completion in citizen science programs.

3. To evaluate how Explainable AI (XAI) improves user engagement and trust in specialized domains such as public health.

Three case studies are presented to demonstrate these objectives both independently and synergistically in real-world contexts, offering valuable insights into the scalability, efficiency, and trustworthiness of human computation systems. By focusing on these cases, the paper illustrates the theoretical potential of IUIs to address real-world challenges in crowd computing. The findings derived from the case studies establish a framework for the future development and evaluation of IUI-driven human computation systems, paving the way for the next generation of crowd computing solutions [3].

The remainder of this paper is organized as follows: Chapter 1.1 reviews the relevant literature, focusing on adaptive task distribution, Explainable AI (XAI), user modeling, gamification, and security in human computation systems. Chapter 2 describes the methodologies used to explore these concepts and their implementation through case studies. Chapter 2.5 presents three case studies—Amazon Mechanical Turk (MTurk), Zooniverse, and a public health project involving XAI in medical diagnostics— illustrating the real-world applications of the proposed methods. Chapter 3 discusses the insights derived from these case studies, analyzing how adaptive methodologies improve scalability, engagement, and trust. Finally, Chapter 4 concludes the paper by summarizing the key findings and outlining directions for future research.

1.1. Related Works

This section reviews the current state of research on the use of Intelligent User Interfaces (IUIs) to enhance scalability, collaboration, and user engagement in human computation and crowd computing platforms. The discussion focuses on key areas such as dynamic task allocation, which optimizes task distribution based on real-time user performance, and Explainable AI (XAI), which improves transparency and trust in AI- assisted problem-solving. It also examines the role of gamification in motivating users, the importance of security and privacy, and the ways in which adaptive IUIs strengthen collaboration and scalability. Collectively, these topics provide a comprehensive overview of the current research landscape and highlight opportunities for future investigation.

1.2. Dynamic task allocation in human computation

Schmidbauer *et al.* [4] investigated human–robot task assignment in industrial environments, demonstrating that adaptive task sharing (ATS), which allows workers to influence task distribution, enhances satisfaction

levels among operators. Although this study emphasized the importance of providing humans with control, it also identified inconsistencies in task allocation, underscoring the need for improved alignment between human preferences and task assignment mechanisms. In contrast, Wen *et al.* [5] proposed a task allocation framework for wireless sensor networks (WSNs) in edge computing environments. Their emphasis on energy efficiency and reliability through parallel task execution revealed the framework's potential to significantly reduce energy consumption and execution time. However, limitations related to fault tolerance and dynamic network conditions were noted, suggesting that further research is needed to improve overall system robustness.

Faccio *et al.* [6] expanded task allocation research into the domain of collaborative robotics by introducing a model that adjusts robot speed based on the distance between the robot and the human operator to balance productivity and safety. Their findings demonstrated notable performance improvements but also revealed inherent trade-offs, particularly the difficulty of maintaining safety without compromising productivity. These results underscore the need for real-time adaptability in dynamic work environments.

Yuan *et al.* [7] proposed an Adaptive Task Allocation framework (ATA-HRL) for multi-human, multi-robot (MH-MR) teams, employing hierarchical reinforcement learning to enhance task adaptability. Their two-stage approach, comprising Initial Task Assignment (ITA) and Conditional Task Reallocation (CTR), proved effective; however, it also highlighted the critical importance of accurate initial allocations, particularly in environments characterized by high uncertainty. Similarly, Tamali *et al.* [8] investigated distributed task allocation in multi-robot systems, utilizing a greedy algorithm to optimize task distribution in complex environments. Their simulation-based approach, implemented using the Robot Operating System (ROS), demonstrated notable improvements in task completion efficiency; however, scalability and communication constraints in real-world scenarios were identified as key challenges for future research.

1.3. Explainable AI in crowdsourced problem-solving

The reviewed studies present a diverse set of approaches and insights into crowdsourced evaluation and Explainable AI (XAI), emphasizing user participation and the enhancement of explanation quality through crowdsourcing and selective input techniques. Despite their differing objectives, these works share a common emphasis on leveraging human feedback to improve both the interpretability and accuracy of AI models. Both Jain *et al.* [9] and Kou *et al.* [10] [11] employed crowdsourcing to enhance AI explanations, although their approaches differed in scope and application. Jain *et al.* focused on evaluating XAI techniques through a Game with a Purpose (GWAP), which allowed users to rank methods such as LIME and Grad-CAM based on their interpretability.

This game-based strategy identified Grad-CAM as the more effective XAI method for image classification. In contrast, Kou *et al.* [10] developed Crowd Graph, a multimodal knowledge graph framework designed to detect and explain fauxtography in social media posts by integrating textual and visual data. Both studies leveraged the power of crowd participation to generate meaningful feedback and improve system performance. However, a critical limitation in both approaches lies in the reliability of crowdsourced data, underscoring the need for more robust mechanisms to verify and maintain data quality.

Kou *et al.* [11] further examined crowdsourcing through the HC-COVID framework, focusing on the detection and explanation of COVID-19 misinformation. The hierarchical design of HC-COVID, integrating contributions from both expert and non-expert crowd workers, represents an extension of their earlier work. This multi-layered approach substantially enhanced the accuracy of misinformation detection and the quality of explanations. However, like other crowdsourced models, it faces the persistent challenge of bias and inconsistency in non-expert data, highlighting the need for more structured frameworks to ensure reliability and precision.

Sawant *et al.* [12] and Lai *et al.* [13] addressed bias and subjectivity in AI explanations, particularly in socially sensitive domains such as hate speech detection and user communication behavior. Sawant *et al.* [12] employed AI-based classification with TabNet to identify hate speech in low-resource languages, including Hindi, and demonstrated how socio-political contexts can introduce bias into annotations. In comparison, Lai *et al.* [13] developed a selective explanation framework that tailors AI explanations to user preferences, such as relevance and abnormality, thereby enabling more context-aware interactions. Both studies emphasize the importance of aligning AI outputs with human interpretation to reduce bias and improve accuracy. However, while Sawant *et al.* revealed that socio-political contexts significantly influence AI bias, Lai *et al.* focused on mitigating cognitive workload and user biases through selective input, underscoring the delicate balance between user engagement and AI autonomy.

A critical comparison across these studies reveals that, while crowdsourcing is an effective approach for enhancing AI explanations, it also introduces substantial challenges related to data quality and bias control. For instance, the reliance on crowdsourced data in Jain *et al.* [9] and Kou *et al.* [10] [11] demonstrates that, although such methods can improve system performance,

they inherently risk inconsistencies in the quality of human-generated content. Similarly, Sawant *et al.* [12] and Lai *et al.* [13] highlighted the susceptibility of AI models that depend on human annotations to various forms of bias, particularly in domains characterized by socio-political sensitivities or subjectivity. Despite these limitations, all studies converge on the conclusion that adaptive and selective frameworks, whether hierarchical models like HC-COVID, game-based evaluations such as Eye into AI, or selective explanation techniques, represent a promising direction toward more transparent and interpretable AI systems.

1.4. Gamification to enhance user engagement

The reviewed studies on gamification across various contexts, from health promotion to employee training and education, demonstrate a shared focus on user engagement, with each applying distinct game elements to influence behavior and outcomes. Despite similarities in their objectives, these works diverge in their methodologies and highlight specific challenges related to sustaining long-term engagement.

Zhang *et al.* [14] and Hellín *et al.* [15] examined gamification as a strategy for behavior change in health and education, respectively. Zhang *et al.* developed DMCoach+, a gamified system designed to promote healthy lifestyles through a two-level structure that integrates personal goals and social competition. In contrast, Hellín et al created a gamified learning environment incorporating points, leaderboards, and badges to enhance student motivation in programming courses. Both studies demonstrated that gamification can significantly enhance engagement; however, they also identified critical limitations.

Zhang *et al.* observed that one-way communication with physicians constrained long-term user engagement, while Hellín *et al.* noted that lower-ranked students could become demotivated by leaderboards. These findings suggest that balancing competition and personalized interaction is essential to maintain long-term engagement in both health and educational contexts.

Lu *et al.* [16] and Bitrián *et al.* [17] investigated the role of gamification in commercial contexts, focusing on user engagement with brand and mobile applications. Lu *et al.* integrated the Mechanics-Dynamics-Aesthetics (MDA) framework into the Nike Run Club (NRC) app, emphasizing that enjoyment was the most significant driver of user engagement and brand loyalty. In contrast, Bitrián *et al.* examined how game design elements fulfill psychological needs, finding that achievement-oriented and social-oriented components enhanced engagement by satisfying the needs for competence, autonomy, and relatedness.

Both studies demonstrated that gamification increases user engagement but also highlighted the importance of balancing fun and personalization to sustain long-term user interest.

Finally, Alfaqiri *et al.* [18] developed a gamification framework for online training platforms, with a focus on employee engagement. The integration of multiple game elements, such as points, challenges, and leaderboards, mirrored techniques used in educational and commercial contexts. Similarly, Bitrián *et al.*, Alfaqiri *et al.* found that these elements effectively increased engagement; however, they also noted that the novelty effect tends to diminish over time, echoing concerns raised in other studies regarding the long-term sustainability of gamified systems.

1.5. Security and privacy concerns in crowd computing platforms

Owoh and Singh [19] developed SenseCrypt for mobile crowd-sensing (MCS) applications, integrating a K-means algorithm with Certificateless Aggregate Signcryption (CLASC) to manage the labeling of sensitive data and ensure secure transmission. This approach reduced computational costs and communication overhead, making the framework robust against multiple attack vectors, including replay and forgery attacks. However, the framework's adaptability to broader use cases remains limited, indicating the need for further research to enhance its scalability and practical implementation.

In contrast, Li *et al.* [20] proposed CrowdSFL, which integrates blockchain technology with federated learning to safeguard data in a decentralized manner. This design maintains privacy by keeping data local while employing smart contracts for secure communication. Federated learning significantly reduced privacy risks by avoiding the centralization of sensitive information, distinguishing this approach from SenseCrypt's signcryption-based method. The integration of a reencryption algorithm based on ElGamal in CrowdSFL added an additional layer of security. Although the outcomes demonstrated improvements in accuracy, security, and computational efficiency, the complexity and higher communication overheads associated with blockchain systems were identified as key challenges.

In comparison, both frameworks offer robust security mechanisms for protecting sensitive data in crowdsourcing environments. However, SenseCrypt focuses primarily on efficient signcryption for mobile sensor data, whereas CrowdSFL emphasizes decentralized privacy preservation through blockchain and federated learning. The communication overheads observed in Li *et al.*'s [20] work contrast with the computational efficiency demonstrated by Owoh and Singh [19], highlighting the inherent trade-off between decentralization and system complexity.

Despite these differences, both studies underscore the continuing need for adaptable and scalable solu-

tions to ensure data security in distributed crowdsourced systems.

1.6. Adaptable IUI for enhancing collaboration and scalability in crowd work

The following studies present several approaches to enhancing collaboration and scalability in human–AI systems and crowd-powered environments. A common thread across these works is the use of adaptive frameworks that dynamically integrate human input, AI processes, and crowdsourced contributions, while each system applies these principles in distinct and context-specific ways.

Siangliulue *et al.* [21] developed IDEAHOUND, a system designed to enhance large-scale collaborative ideation through real-time semantic modeling. By capturing user interactions on a virtual whiteboard, the system dynamically generates diverse and creative suggestions, thereby fostering idea diversity and collaborative participation.

However, the system's reliance on user-generated clusters, which occasionally lack clarity, highlights the challenge of fully leveraging crowd-powered semantic judgments.

In contrast, Abbas *et al.* [22] developed Crowd of Oz (CoZ), a real-time conversational AI system that integrates synchronous crowdsourcing to handle complex social dialogues, particularly for mental health support. Unlike IDEAHOUND, CoZ emphasizes affective communication in real-time interactions.

Although it effectively enhances conversation quality, the system faces challenges in sustaining consistent worker retention and ensuring high-quality responses, underscoring the need for additional training and skill development among crowd workers.

Ponti and Seredko [23] also examined human–AI collaboration, focusing on the context of citizen science. Their task allocation framework assigns simpler activities, such as data collection, to citizen participants, while AI systems and domain experts handle more complex processes like data analysis.

The study highlights how increasing AI capabilities may inadvertently marginalize volunteers, raising concerns about sustaining crowd engagement, a challenge that Siangliulue *et al.* [21] similarly observed in IDEAHOUND with respect to idea clustering.

Building on the theme of citizen-centric systems, Stein and Yazdanpanah [24] introduced C-MAS, a multi-agent system designed to give citizens greater control over decision-making in smart mobility and energy domains. Like CoZ, C-MAS places strong emphasis on privacy, fairness, and transparency, but extends these principles by enabling citizens to actively shape decisions through personal intelligent agents. Nevertheless, the challenge of trust, particularly concerning privacy and ethical decision-making, remains a central issue, reflecting the same need for worker trust observed in CoZ's affective communication model.

Lastly, Gupta *et al.* [25] developed COHUMAIN, a framework designed to foster collective intelligence in human-AI teams through a socio cognitive architecture. By sharing cognitive resources via transactive memory and attention systems, COHUMAIN seeks to enhance collaborative decision-making and scalability, paralleling the collaborative ideation enabled by IDEAHOUND.

However, Gupta *et al.* identified a persistent challenge of maintaining long-term trust between human and AI collaborators, an issue also observed by Stein and Yazdanpanah [24] in the context of trust between citizens and AI agents.

While each of these frameworks successfully advances adaptable collaboration and scalability across different domains, they share common challenges related to trust, participant retention, and the quality of collaborative outcomes.

These recurring obstacles underscore the ongoing need for refinement in training, dynamic task allocation, and human–AI integration to achieve truly scalable and effective crowd-powered systems.

2. Materials and Methods

This section presents a theoretical and evaluative framework for assessing the effectiveness of adaptive Intelligent User Interfaces (IUIs) in scalable human computation systems. It is grounded in well-established closed loop adaptive UI principles, where interface behavior continuously adjusts to user context and evolving preferences. In this dynamic feedback model, user actions generate signals that inform the IUI's adaptive policies, modulating task distribution, personalization, gamification, and explainability modules, in alignment with frameworks applied in AI-driven smart Product–Service Systems (SPSS) [26].

By synchronizing user behavior with real-time interface responses, this framework provides a robust theoretical foundation for adaptive task distribution and user modeling.

This research employs a mixed case study design that integrates theoretical frameworks, such as Activity Theory, the Technology Acceptance Model, and adaptive UI principles, with quantitative empirical data collected from Amazon Mechanical Turk (MTurk), Zooniverse, and a clinical XAI environment. Rather than conducting randomized controlled trials, the study gathered quantitative pilot metrics and analyzed them using the TRIPLE C case study methodology [27]. Performance indicators include task accuracy, completion time, error rates, volunteer retention, diagnostic accuracy, and clinician trust, each supported by appropriate statistical tests to substantiate the findings.

Figure 1 illustrates the key layers involved in optimizing human computation environments. The diagram comprises three primarylayers: the Input Layer, consisting of Human Participants, Crowd Tasks, and AI Systems; the Adaptive IUI Layer, which includesmodules such as Dynamic Task Allocation, User Modeling, Explainable AI, Gamification, and Security Measures; and the Outcome Layer, representing key results including Scalability, Collaboration, Engagement, and Data Integrity. The flow of tasks and data originates in the Input Layer, where human participants and AI systems interact with assigned tasks, progresses through the Adaptive IUI Layer for optimization, and culminates in enhanced system performance within the Outcome Layer. This framework illustrates how adaptive IUIs can be leveraged to enhance collaboration and scalability in crowd sourced environments.

From the perspective of Activity Theory, a well-established interpretative framework in case study research, the operator (subject) interacts with tasks (object) through the IUI (tool). The adaptive interface mediates this interaction via real-time feedback loops, thereby enhancing task performance and overall system efficiency [28].

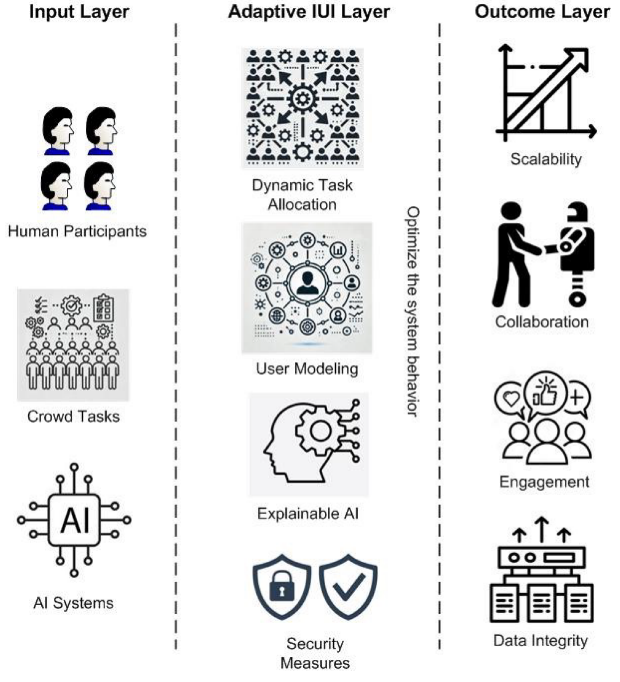


Figure 1. Conceptual Framework of Adaptive IUI for Scalable Human Computation

2.1. Adaptive Task Distribution

From the perspective of Activity Theory, the operator interacts with the interface to accomplish specific goals, with the IUI dynamically mediating this relationship through real-time feedback. Adaptive task distribution enhances the efficiency of human computation in crowdsourcing by allocating tasks according to user skills, performance, and expertise. Figure 2 illustrates this model: tasks from the pool are assigned via an AI system that leverages real-time feedback to optimize future allocations. Comparable approaches have demonstrated effectiveness in human–robot collaboration, such as Schmidbauer *et al.* adaptive task-sharing framework based on user capabilities and preferences [29], as well as in adaptive learning systems, where task reassignment improves performance according to prior outcomes [30]. This theoretical model will be examined through case studies in domains such as disaster response, data labeling, and urban planning to assess scalability, practicality, and performance in large-scale human computation.

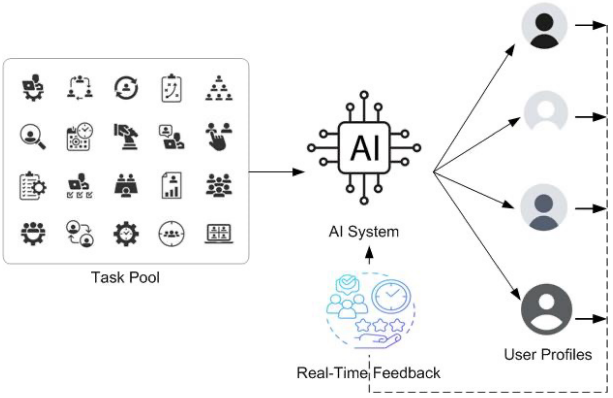


Figure 2. Adaptive Task Distribution

2.2. Explainable AI in Human Computation

Explainable AI and gamification strategies in this framework are supported by the Technology Acceptance Model (TAM), which posits that greater perceived usefulness and transparency directly enhance user trust and adoption [31].

Explainable AI (XAI) enhances trust, engagement, and efficiency in crowdsourced tasks by making AI decision processes transparent to human participants. Widely adopted XAI models such as LIME and Grad CAM increase interpretability: LIME decomposes the contribution of individual input features to a decision, while Grad-CAM highlights the most influential image regions within convolutional neural networks (CNNs) [32, 33]. Together, these tools help users understand the rationale behind AI outcomes, thereby fostering accountability and informed participation.

In practical crowd work, such as data labeling, analysis, and public health initiatives, XAI enables participants to validate AI-generated outputs and identify potential biases. For example, workers on labeling platforms can use LIME explanations to evaluate

and correct model predictions, while clinicians employing Grad-CAM visualizations can pinpoint the regions of medical images that influenced AI diagnoses, thereby improving collaborative diagnostic accuracy and trust [34]. By embedding XAI into case studies across domains, this research assesses its impact on user trust, task performance, and engagement, moving beyond theoretical advantages to demonstrate real-world applicability, see Figure 3.

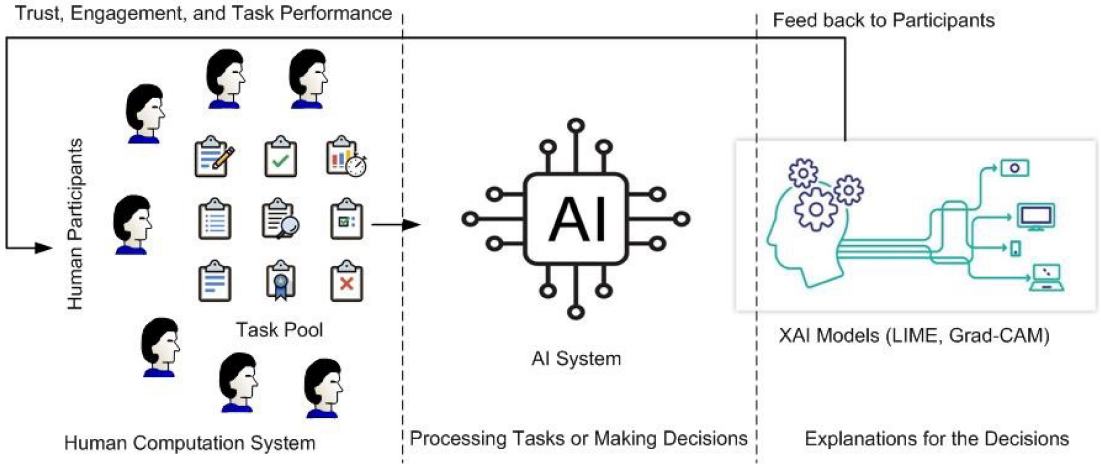


Figure 3. Hypothetical Model for Explainable AI in Human Computation

2.3. User Modeling and Dynamic Task Allocation

User modeling constructs a computational representation of individual preferences, expertise, and behavior, enabling crowd-work systems (e.g., MTurk, Zooniverse) to allocate tasks dynamically according to each user's skill set.

Techniques such as collaborative filtering, which matches users to tasks based on shared behavior, enhance efficiency; for instance, users who excel in specific labeling categories can be reassigned to similar tasks [35].

Reinforcement learning further refines this process by adjusting assignments in real time based on performance feedback, particularly in gamified settings where the system learns users' strengths and optimizes task distribution [36].

These modeling techniques enhance personalization and adaptability while reducing task redundancy. Personalized task assignments increase user engagement by aligning tasks with individual capabilities, whereas the dynamic reassignment of complex tasks helps sustain motivation and productivity [37].

Moreover, by avoiding repetitive task allocations, user modeling prevents boredom and disengagement.

This study evaluates these effects through real-world case studies on platforms such as MTurk and Zooniverse, demonstrating how collaborative filtering and reinforcement learning improve scalability and effectiveness in human computation [38].

2.4. Gamification and Security Measures

Gamification, the use of game-like elements such as leaderboards, points, and rewards to enhance motivation in non-gaming contexts, has proven effective in human computation platforms. By fostering both intrinsic and extrinsic motivation, gamification increases participant retention and improves task performance [39]. For instance, platforms that award badges and display user rankings frequently report higher levels of engagement and task completion [40].

However, increased engagement alone insufficient; crowdsourced systems must also address inherent security risks. Threats such as data breaches, task manipulation, and fraud can undermine both data integrity and user trust. To mitigate these vulnerabilities, robust security measures, including encryption, authentication protocols, Certificateless Aggregate Signcryption (CLASC), and blockchain-based frameworks, are essential [41]. When integrated effectively, thoughtful gamification and strong security architectures enable crowd platforms to scale efficiently while preserving trust, participation, and data protection.

2.5. Case Studies

Case Study 1: Amazon Mechanical Turk (MTurk)

Amazon Mechanical Turk (MTurk) is one of the leading microtask crowdsourcing platforms, enabling companies and researchers to assign Human Intelligence Tasks (HITs), or small-scale operations, to a global

workforce. These tasks, ranging from data labeling and image classification to survey participation, make MTurk an ideal platform for evaluating Adaptive Task Distribution, User Modeling, and Gamification. Given its large user base and flexible task allocation mechanisms, MTurk offers a practical environment for assessing how adaptive methodologies can enhance task efficiency and user engagement.

- Adaptive Task Distribution: MTurk dynamically assigns tasks based on user experience, skill level, and historical performance, ensuring that simpler tasks are allocated to newer workers while more complex tasks are reserved for experienced ones. This real-time feedback loop optimizes task efficiency and reduces error rates.
- User Modeling: By gathering user performance data, such as task completion accuracy and speed, the platform enables the system to tailor task assignments according to each user's unique strengths. This approach ensures stronger task–worker alignment and fosters a more engaging user experience.
- Gamification: Although MTurk lacks built-in gamification features, incorporating elements such as leaderboards, ribbons, and rewards can enhance worker engagement and retention while motivating participants to c ontribute more frequently.

Case Study 2: Zooniverse

Zooniverse, the world's largest citizen science platform, enables participants to contribute to a diverse range of research projects by classifying data, analyzing images, and transcribing historical texts. As a crowdsourced human computation platform, Zooniverse depends on public participation, making it an ideal environment for exploringAdaptive Task Distribution, User Modeling, and Gamification to optimize engagement, task efficiency, and scalability.

- Adaptive Task Distribution: Tasks are allocated based on user productivity and experience. Experienced participants handle more complex analyses, improving task accuracy and overall efficiency, while new volunteers begin with simpler-classifications.
- User Modeling: By tracking user behavior and performance, the platform assigns tasks tailored to each participant's skills. Assigning similar tasks to users who excel in specific projects further enhances engagement and accuracy.
- Gamification: Features such as project milestones, progress tracking, and achievement badges motivate volunteers to contribute consistently, fostering a strong sense of community within the platform.

Case Study 3: Public Health

In the healthcare sector, AI models are increasingly employed for diagnostic support, predictive modeling, and medical image analysis. Although these models achieve high accuracy in disease detection, their lack of transparency raises concerns about trust and interpretability. Explainable AI (XAI) addresses this challenge by making AI decision processes more understandable, enabling healthcare professionals to validate predictions and collaborate more effectively with AI systems. This case study examines how LIME (Local Interpretable Model-agnostic Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) enhance trust in AI-assisted medical image analysis.

- LIME: By perturbing input data and analyzing how AI predictions change, LIME generates interpretable explanations of model behavior. In medical imaging, it clarifies why an AI system labels a particular region of an MRI or X-ray as problematic, enabling doctors to understand and validate the model's reasoning.
- Grad-CAM: This technique generates heatmaps that illustrate the regions of medical images influencing AI predictions. It is particularly valuable for identifying diseases such as cancer, as it visually highlights the areas that guided the AI's decision-making, there by enhancing interpretability.

3. Results and discussion

Evaluation follows the TRIPLE C reporting principles, a recognized framework for case study assessments encompassing context, methods, and complexity, to ensure scientific rigor through data triangulation across performance metrics [27, 42]. Each case study establishes clear evaluation objectives and measures outcomes using quantitative indicators such as task accuracy, completion time, error rate, volunteer retention, diagnostic accuracy, and trust scores. This structured, data-driven approach enhances objectivity and grounds the findings in a clearly defined evaluation protocol.

Figure 4 illustrates the task completion performance of the MTruk crowdsourcing platform. The adaptive task allocation system on MTurk significantly improves task efficiency by assigning complex tasks to skilled workers, resulting in higher accuracy and faster completion rates. Personalized User Modeling further enhances engagement by tailoring task assignments to individual strengths, reducing redundancy,

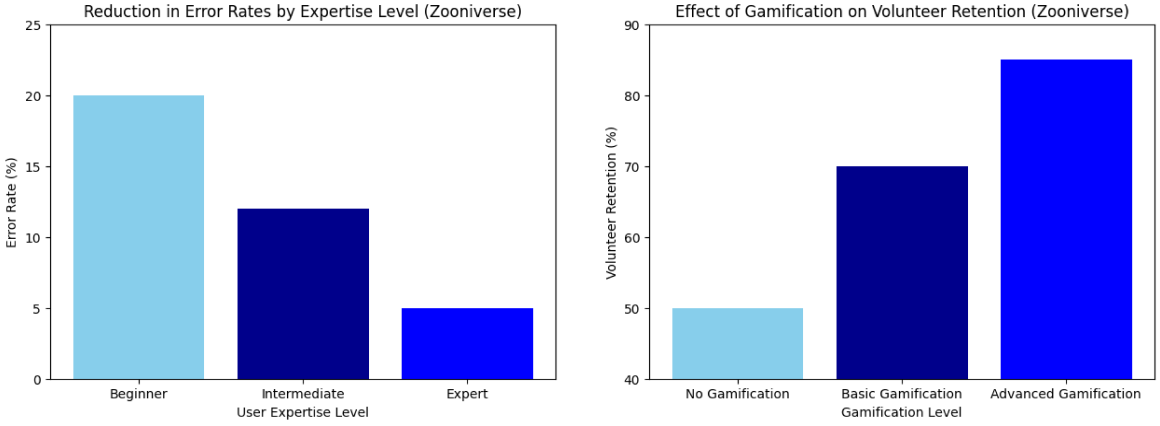
and preventing skill mismatches. Although Gamification is not yet a core feature of MTurk, existing research suggests that incorporating competitive and reward-based incentives could increase participation and worker retention, particularly in repetitive tasks.

This case study demonstrates how adaptive approaches can enhance the scalability, effectiveness, and user engagement of crowdsourced platforms. Dynamic task assignment and personalized user modeling optimize platform outcomes by improving resource utilization. Although gamification elements are not yet fully integrated, the structure of MTurk positions it as a strong candidate for such enhancements. Future research should investigate the incorporation of gamification features and the refinement of user profiling techniques to maximize efficiency and foster sustained participation in large-scale crowd work environments.

(a) Accuracy by skill level (b) Time by experience level

Figure 4. Task Completion Performance of MTruk

Implementing Adaptive Task Distribution reduces errors by aligning tasks with users' expertise. User modeling enhances participation by tailoring task assignments to individual strengths, resulting in higher retention and satisfaction. Gamification significantly improves motivation, as volunteers demonstrate greater commitment when rewarded with badges and milestone achievements. The visibility of overall project progress further reinforces sustained engagement. Figure 5 illustrates the platform's performance in terms of error rate by expertise level and the impact of gamification on volunteer retention.



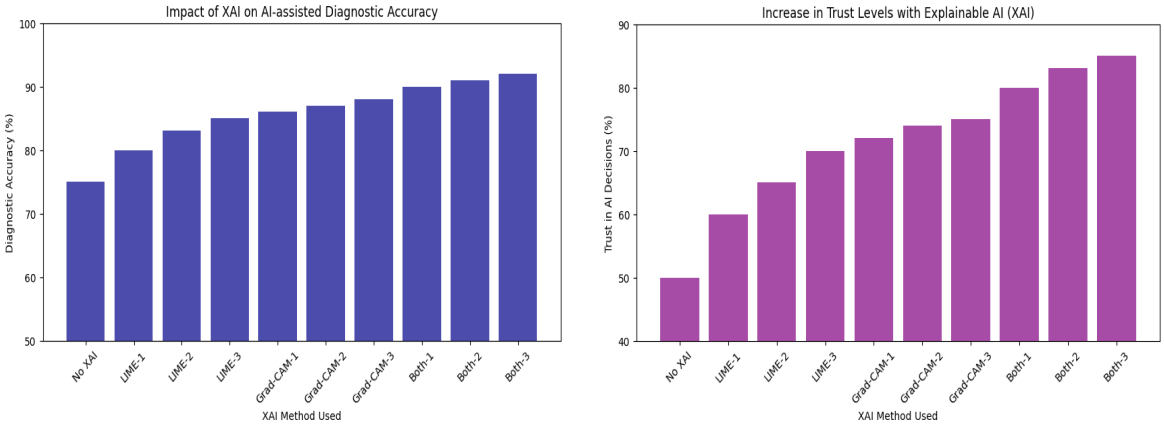
(a) Error rate by expertise (b) Volunteer retention on Gamification

Figure 5. Adaptive Task Distribution Performance of Zooniverse

This case study demonstrates how adaptive techniques enhance scalability, engagement, and efficiency in crowdsourced platforms. Personalized task assignments and gamification strategies foster an engaging and sustainable model for volunteer-driven research. Zooniverse serves as a valuable example of how integrating adaptive IUIs into citizen science can enhance user participation and improve task outcomes, with comparable approaches applicable to other large-scale human computation systems.

The impact of integrating Explainable AI (XAI) into AI-assisted medical diagnostics is illustrated in Figure 6. Incorporating XAI significantly enhances transparency and trust in AI systems. LIME enables clinicians to validate AI decisions by identifying the image features that influenced classifications, thereby improving diagnostic accuracy. Similarly, Grad-CAM heatmaps provide visual explanations of AI predictions, allowing healthcare professionals to interpret model outputs more effectively and refine their diagnoses. Combining both methods further enhances the performance of clinical decision support systems (CDSS) in terms of diagnostic accuracy and trust. Analyses indicate that the inclusion of XAI techniques enhances collaboration between AI systems and medical experts, leading to more reliable diagnoses and greater confidence in AI-driven assessments.

This case study highlights the critical role of Explainable AI (XAI) in healthcare, where decision-making transparency is essential. The integration of LIME and Grad-CAM into medical diagnostics bridges the gap between AI-generated predictions and human interpretability, ensuring that AI-driven decisions remain verifiable and trustworthy. Future research should focus on optimizing XAI for clinical applications to improve usability and strengthen collaboration between AI systems and healthcare professionals. These findings suggest that XAI should be regarded as a fundamental component of AI-assisted diagnostic tools, particularly in high-stakes domains such as healthcare, where accuracy and explainability are paramount.



(a) Diagnostic Accuracy (b) Trust Level

Figure 6. Impact of XAI Integration in AI-assisted medical diagnostics

Collectively, these case studies demonstrate that although adaptive approaches are generally effective in crowdsourced environments, their success depends on contextual factors. Gamification and adaptive task allocation substantially enhance user engagement and task performance on platforms such as MTurk and Zooniverse. Explainable AI, in contrast, is crucial for ensuring that AI systems are not only effective but also transparent and reliable in specialized domains such as healthcare. This hybrid approach underscores the versatility and significance of these techniques in optimizing human computation systems, offering both domain-specific recommendations for contexts requiring higher levels of scrutiny and trust, as well as generalizable insights for large-scale platforms.

3.1. Impact of IUI-Driven Approaches Compared to Non-IUI Systems

- Scalability and Flexibility: The IUI-driven systems implemented in MTurk and Zooniverse are not only more scalable than traditional models but also more flexible, enabling them to adapt to the increasing complexity and scale of crowdsourced tasks. This adaptability is largely absent in static, non-IUI systems, which struggle to manage large-scale or rapidly evolving work loads effectively.

- Enhanced Trust and Collaboration: The integration of XAI into IUI-driven systems, particularly in the public health case, demonstrates

how these systems foster stronger collaboration between humans and AI. This represents a substantial improvement over traditional AI models, in which limited transparency often undermines user trust and hinders the acceptance of AI-driven processes.

By explicitly comparing adaptive IUI methodologies with non-IUI-driven approaches, it becomes evident that the former offers substantial improvements in efficiency, scalability, engagement, and trust. The adaptive task distribution, user modeling, and gamification strategies employed in platforms such as MTurk and Zooniverse address the limitations of static task allocation and impersonal systems, while XAI introduces the transparency and collaboration often lacking in traditional AI models. This comparison underscores that IUI-driven systems provide a more robust, scalable, and trustworthy framework for optimizing human computation environments.

3.2. Limitations

Despite its contributions, the proposed Adaptive IUI framework presents several limitations:

1. The effectiveness of dynamic task allocation and personalization relies heavily on accurate and up-to-date user models. Current research highlights the challenge of capturing complex, multidimensional user attributes, such as emotion, context, and behavior, while preserving model consistency and ensuring user privacy.

2. Integrating multiple adaptive modules, including task allocation, gamification, explainability, and security, can impose significant computational and implementation overhead. Prior studies on adaptive UI frameworks have identified system performance, resource consumption, and maintainability as persistent challenges.

3. Although gamification can enhance engagement, it is prone to the novelty effect, an initial surge in user interest followed by a gradual decline. Furthermore, excessive reliance on extrinsic rewards may lead to the over justification effect, which undermines intrinsic motivation. These phenomena underscore the need for a balanced, long-term gamification strategy.

4. Adaptive IUIs collect and analyze sensitive data, including user behavior and channel state information (CSI), which raises significant privacy, security, and ethical challenges. Ensuring robust data protection and maintaining transparency are essential to preserving user trust.

4. Conclusions

This benchmarking study demonstrates the significance of combining adaptive approaches, such as Adaptive Task Distribution, User Modeling, Gamification, and Explainable AI (XAI), to enhance human computation systems. By employing a hybrid strategy that integrates two general-purpose case studies (MTurk and Zooniverse) with a specialized case study in public health using XAI, the research provides comprehensive insights into the adaptability, scalability, and reliability of adaptive IUIs. Personalized task assignments and dynamic task allocation significantly improve task performance and engagement on large-scale platforms such as Zooniverse and Amazon Mechanical Turk. Real-time task distribution based on user competence ensures that tasks are assigned to the most qualified participants, increasing productivity and reducing error rates. Furthermore, incorporating gamification elements enhances user motivation and long-term engagement, which is crucial for maintaining sustained participation in crowdsourced platforms. The public health case study highlights the critical role of transparency and trust in high-stakes environments. Integrating XAI models such as LIME and Grad-CAM enhances the interpretability of AI-driven diagnostics, enabling healthcare practitioners to understand and evaluate AI-generated results. This transparency fosters collaboration and strengthens confidence in AI systems, ultimately contributing to improved patient outcomes. In conclusion, this study demonstrates that adaptive IUI systems are scalable and customizable across both general crowdsourced platforms and specialized domains such as healthcare. Collectively, these approaches provide a flexible, transparent, and reliable foundation for advancing human computation systems. Their integration not only enhances scalability and engagement but also strengthens trust, accuracy, and accountability in complex, high-impact applications.

Contributor Roles

- **R. J. Martin:** Conceptualization, investigation, methodology, validation, writing – original draft, writing – review & editing.

References

[1] L. von Ahn and L. Dabbish, "Designing games with a purpose," *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, Aug. 2008. [Online]. Available: https://doi.org/10.1145/1378704.1378719

[2] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, ser. CHI '99. ACM

Press, 1999, pp. 159–166. [Online]. Available: https://doi.org/10.1145/302979.303030

[3] A. J. Quinn and B. B. Bederson, "Human computation: a survey and taxonomy of a growing field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. ACM, May 2011, pp. 1403–1412. [Online]. Available: https://doi.org/10.1145/1978942.1979148

[4] C. Schmidbauer, S. Zafari, B. Hader, and S. Schlund, "An empirical study on workers' preference in human–robot task assignment in industrial assembly systems," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, pp. 293–302, 2023. [Online]. Available: https://doi.org/10.1109/thms.2022.3230667

[5] J. Wen, J. Yang, T. Wang, Y. Li, and Z. Lv, "Energy-efficient task allocation for reliable parallel computation of cluster-based wireless sensor network in edge computing," *Digital Communications and Networks*, vol. 9, no. 2, pp. 473–482, Apr. 2023. [Online]. Available: https://doi.org/10.1016/j.dcan.2022.06.014

[6] M. Faccio, I. Granata, and R. Minto, "Task allocation model for human-robot collaboration with variable cobot speed," *Journal of Intelligent Manufacturing*, vol. 35, no. 2, pp. 793–806, Jan. 2023. [Online]. Available: https://doi.org/10.1007/s10845-023-02073-9

[7] Z. Yuan, R. Wang, T. Kim, D. Zhao, I. Obi, and B.-C. Min, "Adaptive task allocation in multi-human multi-robot teams under team heterogeneity and dynamic information uncertainty," *ICRA 2025*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2409.13824

[8] A. Tamali, N. Amardjia, and M. Tamali, "Distributed and autonomous multi-robot for task allocation and collaboration using a greedy algorithm and robot operating system platform," *IAES International Journal of Robotics and Automation (IJRA)*, vol. 13, no. 2, p. 205, Jun. 2024. [Online]. Available: https://doi.org/10.11591/ijra.v13i2.pp205-219

[9] M. Jain, *Crowd-Sourced Evaluation of Explainable AI Techniques with Games*. Carnegie Mellon University, 2021. [Online]. Available: https://upsalesiana.ec/ing35ar4r9

[10] Z. Kou, Y. Zhang, D. Zhang, and D. Wang, "Crowdgraph: A crowdsourcing multi-modal knowledge graph approach to explainable fauxtography detection," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, pp. 1–28, Nov. 2022. [Online]. Available: https://doi.org/10.1145/3555178

[11] Z. Kou, L. Shang, Y. Zhang, and D. Wang, "HC-COVID: A hierarchical crowdsource knowledge graph approach to explainable COVID-19 misinformation detection," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. GROUP, pp. 1–25, Jan. 2022. [Online]. Available: http://doi.org/10.1145/3492855

[12] M. Sawant, A. Younus, S. Caton, and M. A. Qureshi, "Using explainable AI (XAI) for identification of subjectivity in hate speech annotations for low-resource languages," in *4th International Workshop on OPEN CHALLENGES IN ONLINE SOCIAL NETWORKS*, ser. HT '24. ACM, Sep. 2024, pp. 10–17. [Online]. Available: http://doi.org/10.1145/3677117.3685006

[13] V. Lai, Y. Zhang, C. Chen, Q. V. Liao, and C. Tan, "Selective explanations: Leveraging human input to align explainable AI," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CSCW2, pp. 1–35, Sep. 2023. [Online]. Available: http://doi.org/10.1145/3610206

[14] C. Zhang, P. van Gorp, M. Derksen, R. Nuijten, W. A. IJsselsteijn, A. Zanutto, F. Melillo, and R. Pratola, "Promoting occupational health through gamification and e-coaching: A 5-month user engagement study," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 2823, Mar. 2021. [Online]. Available: http://doi.org/10.3390/ijerph18062823

[15] C. J. Hellín, F. Calles-Esteban, A. Valledor, J. Gómez, S. Otón-Tortosa, and A. Tayebi, "Enhancing student motivation and engagement through a gamified learning environment," *Sustainability*, vol. 15, no. 19, p. 14119, Sep. 2023. [Online]. Available: http://doi.org/10.3390/su151914119

[16] H.-P. Lu and H.-C. Ho, "Exploring the impact of gamification on users' engagement for sustainable development: A case study in brand applications," *Sustainability*, vol. 12, no. 10, p. 4169, May 2020. [Online]. Available: http://doi.org/10.3390/su12104169

[17] P. Bitrián, I. Buil, and S. Catalán, "Enhancing user engagement: The role of gamification in mobile apps," *Journal of Business Research*, vol. 132, pp. 170–185, Aug. 2021. [Online]. Available: http://doi.org/10.1016/j.jbusres.2021.04.028

[18] A. S. Alfaqiri, S. F. M. Noor, and N. Sahari, "Framework for gamification of online training

platforms for employee engagement enhancement," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 06, pp. 159–175, Mar. 2022. [Online]. Available: http://doi.org/10.3991/ijim.v16i06.28485

[19] N. Pius Owoh and M. Mahinderjit Singh, "Sensecrypt: A security framework for mobile crowd sensing applications," *Sensors*, vol. 20, no. 11, p. 3280, Jun. 2020. [Online]. Available: http://doi.org/10.3390/s20113280

[20] Z. Li, J. Liu, J. Hao, H. Wang, and M. Xian, "Crowdsfl: A secure crowd computing framework based on blockchain and federated learning," *Electronics*, vol. 9, no. 5, p. 773, May 2020. [Online]. Available: http://doi.org/10.3390/electronics9050773

[21] P. Siangliulue, J. Chan, S. P. Dow, and K. Z. Gajos, "Ideahound: Improving large-scale collaborative ideation with crowd-powered real-time semantic modeling," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST '16. ACM, Oct. 2016, pp. 609–624. [Online]. Available: http://doi.org/10.1145/2984511.2984578

[22] T. Abbas, *Affective Real-Time Crowd-Powered Conversational Systems.* Eindhoven University of Technology, Sep. 2022. [Online]. Available: https://upsalesiana.ec/ing35ar4r22

[23] M. Ponti and A. Seredko, "Human-machine-learning integration and task allocation in citizen science," *Humanities and Social Sciences Communications*, vol. 9, no. 1, Feb. 2022. [Online]. Available: http://doi.org/10.1057/s41599-022-01049-z

[24] S. Stein and V. Yazdanpanah, "Citizen-centric multiagent systems," in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '23. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2023, p. 1802–1807. [Online]. Available: https://upsalesiana.ec/ing35ar4r24

[25] P. Gupta, T. N. Nguyen, C. González, and A. W. Woolley, "Fostering collective intelligence in human–ai collaboration: Laying the groundwork for cohumain," *Topics in Cognitive Science*, vol. 17, no. 2, pp. 189–216, Jun. 2023. [Online]. Available: http://doi.org/10.1111/tops.12679

[26] A. Carrera-Rivera, F. Larrinaga, G. Lasa, G. Martínez-Arellano, and G. Unamuno, "Adaptui: A framework for the development of adaptive user interfaces in smart product-service systems," *User Modeling and User-Adapted Interaction*, vol. 34, no. 5, pp. 1929–1980, Aug. 2024. [Online]. Available: http://doi.org/10.1007/s11257-024-09414-0

[27] S. E. Shaw, S. Paparini, J. Murdoch, J. Green, T. Greenhalgh, B. Hanckel, H. M. James, M. Petticrew, G. W. Wood, and C. Papoutsi, "TRIPLE C reporting principles for case study evaluations of the role of context in complex interventions," *BMC Medical Research Methodology*, vol. 23, no. 1, May 2023. [Online]. Available: http://doi.org/10.1186/s12874-023-01888-7

[28] L. Uden and N. Willis, "Designing user interfaces using activity theory," in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, ser. HICSS-01. IEEE Comput. Soc, 2005, p. 11. [Online]. Available: http://doi.org/10.1109/hicss.2001.926547

[29] M. Calzavara, M. Faccio, I. Granata, and A. Trevisani, "Achieving productivity and operator well-being: A dynamic task allocation strategy for collaborative assembly systems in industry 5.0," *The International Journal of Advanced Manufacturing Technology*, Aug. 2024. [Online]. Available: http://doi.org/10.1007/s00170-024-14302-3

[30] M. H. Faisal, A. W. AlAmeeri, and A. A. Alsumait, "An adaptive e-learning framework: crowdsourcing approach," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, ser. iiWAS '15. ACM, Dec. 2015, pp. 1–5. [Online]. Available: http://doi.org/10.1145/2837185.2837249

[31] J. C. Cheung and S. S. Ho, "The effectiveness of explainable AI on human factors in trust models," *Scientific Reports*, vol. 15, no. 1, Jul. 2025. [Online]. Available: http://doi.org/10.1038/s41598-025-04189-9

[32] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. ACM, Aug. 2016, pp. 1135–1144. [Online]. Available: http://doi.org/10.1145/2939672.2939778

[33] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017, pp. 618–626. [Online]. Available: http://doi.org/10.1109/iccv.2017.74

[34] K. Borys, Y. A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C. M. Friedrich, and F. Nensa, "Explainable AI in medical imaging: An overview for clinical practitioners – saliency-based XAI approaches," *European Journal of Radiology*, vol. 162, p. 110787, May 2023. [Online]. Available: http://doi.org/10.1016/j.ejrad.2023.110787

[35] Q. Wang, Y. Wan, F. Feng, and X. Wang, "Threshold optimization of task allocation models in human–machine collaborative scoring of subjective assignments," *Computers & Industrial Engineering*, vol. 188, p. 109923, Feb. 2024. [Online]. Available: http://doi.org/10.1016/j.cie.2024.109923

[36] L. Sun, X. Yu, J. Guo, Y. Yan, and X. Yu, "Deep reinforcement learning for task assignment in spatial crowdsourcing and sensing," *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25 323–25 330, Nov. 2021. [Online]. Available: http://doi.org/10.1109/jsen.2021.3057376

[37] S. N. Ahmadabadi, M. Haghifam, V. Shah-Mansouri, and S. Ershadmanesh, "Design and evaluation of crowdsourcing platforms based on users' confidence judgments," *Scientific Reports*, vol. 14, no. 1, Aug. 2024. [Online]. Available: http://doi.org/10.1038/s41598-024-65892-7

[38] J. Cox, E. Y. Oh, B. Simmons, C. Lintott, K. Masters, A. Greenhill, G. Graham, and K. Holmes, "Defining and measuring success in online citizen science: A case study of zooniverse projects," *Computing in Science ∓ Engineering*, vol. 17, no. 4, pp. 28–41, Jul. 2015. [Online]. Available: http://doi.org/10.1109/mcse.2015.65

[39] S. A. Triantafyllou, T. Sapounidis, and Y. Farhaoui, "Gamification and computational thinking in education: A systematic literature review," *Salud, Ciencia y Tecnología - Serie de Conferencias*, vol. 3, p. 659, Mar. 2024. [Online]. Available: http://doi.org/10.56294/sctconf2024659

[40] H. Cigdem, M. Ozturk, Y. Karabacak, N. Atik, S. Gürkan, and M. H. Aldemir, "Unlocking student engagement and achievement: The impact of leaderboard gamification in online formative assessment for engineering education," *Education and Information Technologies*, vol. 29, no. 18, pp. 24 835–24 860, Jun. 2024. [Online]. Available: http://doi.org/10.1007/s10639-024-12845-2

[41] A. Tomar and S. Tripathi, "Bcsom: Blockchain-based certificateless aggregate signcryption scheme for internet of medical things," *Computer Communications*, vol. 212, pp. 48–62, Dec. 2023. [Online]. Available: http://doi.org/10.1016/j.comcom.2023.09.027

[42] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," *Empirical Software Engineering*, vol. 14, no. 2, pp. 131–164, Dec. 2008. [Online]. Available: http://doi.org/10.1007/s10664-008-9102-8

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.05
pISSN: 1390-650X / eISSN: 1390-860X

Analysis of closed-form ground-return impedances for short-circuit studies in overhead distribution systems

Análisis de impedancias de tierra de forma cerrada en estudios de cortocircuito de sistemas de distribución aérea

Allen A. Castillo Barrón[1,*], Gerardo Ayala Jaimes[1], Alejandra Jiménez Vega[1], Francisco J. Ramírez Arias[1]

Received: 29-04-2025, Received after review: 05-11-2025, Accepted: 18-11-2025, Published: 01-01-2026

Abstract

The objective of this study is to evaluate the applicability of the most widely used closed-form ground-return impedance formulas in short-circuit analyses of distribution systems and to identify the critical network configurations in which the choice of impedance model significantly affects the short-circuit results. The methodology adopted in this research is organized into three stages. First, an algorithm was developed to implement and compare several closed-form earth-return impedance formulations, and its performance was validated using benchmark data reported in the literature. Second, a short-circuit analysis algorithm was designed and verified against reference results published by the IEEE Power and Energy Society. Finally, multiple short-circuit studies were performed on several IEEE distribution test feeders. The findings reveal that most closed-form earth-return impedance models provide adequate accuracy for both balanced and unbalanced short-circuit analyses. However, for single-phase line-to-ground faults, the choice of closed-form impedance formulation is critical to obtaining accurate short-circuit results.

Keywords: Closed-form ground-return impedance, Distribution systems; Short-circuit calculation accuracy

Resumen

El objetivo de este estudio es evaluar la aplicabilidad de las principales fórmulas cerradas de impedancia de retorno por tierra en el análisis de cortocircuito de sistemas de distribución, así como identificar las configuraciones críticas en las cuales la elección del modelo de impedancia puede influir significativamente en los resultados del cortocircuito. La metodología adoptada en esta investigación se estructura en tres etapas. Primero, se desarrolló un algoritmo para implementar y comparar las formulaciones cerradas de impedancia de retorno por tierra, el cual fue validado con datos de referencia disponibles en la literatura. Segundo, se diseñó un algoritmo de análisis de cortocircuito que fue verificado con resultados publicados por la IEEE Power and Energy Society. Finalmente, se realizaron múltiples estudios de cortocircuito en varios alimentadores de prueba de distribución del IEEE. Los resultados muestran que la mayoría de las formulaciones cerradas de impedancia de retorno por tierra son adecuadas tanto para análisis de cortocircuito balanceados como desbalanceados, y que, en fallas monofásicas a tierra, la elección de la fórmula cerrada de impedancia resulta un factor crítico para obtener resultados precisos.

Palabras clave: Fórmulas cerradas de impedancia de tierra, sistemas de distribución, exactitud de cálculo de cortocircuito.

[1,*]Department of Electrical Engineering (FCITEC), Universidad Autónoma de Baja California, Tijuana, México.
Corresponding author ✉: allen.castillo@uabc.edu.mx.

Suggested citation: A. A. Castillo Barrón, G. Ayala Jaimes, A. Jiménez Vega and F. J. Ramírez Arias "Analysis of closed-form ground-return impedances for short-circuit studies in overhead distribution systems," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 64-77, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.05.

1. Introduction

Distribution systems exhibit an inherently asymmetrical geometry and, unlike transmission lines, they are not transposed. This lack of symmetry leads to significant phase imbalances, increasing the current's circulation through the earth. Consequently, accurately modeling the ground-return impedance is essential to obtaining a precise representation of the system's electrical behavior [1–3].

The first model for earth-return impedance was developed by Carson in 1926, who derived the earliest analytical formulation by expressing the axial electric field in the soil as a Fourier-type improper integral under quasi-TEM assumptions. Carson's model represents a foundational contribution to the field; however, it relied on simplifying assumptions such as homogeneous soil, ground permeability equal to that of free space, and the neglect of displacement currents, which limited its applicability at higher frequencies and in soils with significant permittivity effects [4]. Subsequent researchers addressed these limitations.

Wise, in his early works from 1931 and 1934, introduced formulations that extended Carson's model by incorporating finite ground permeability and displacement currents, using the Hertzian vector potential and Fourier–Bessel expansions to generalize the wave equation for imperfect soils [5, 6]. Later, in 1968, Sunde further advanced Wise's work by incorporating the effects of layered ground structures into the earth-return impedance formulations [7].

Carson's formulation was originally expressed as improper integrals, which made its direct evaluation challenging. Over the years, researchers have proposed several methods to evaluate these integrals, including numerical integration techniques and infinite-series expansions [1], [8]. However, these approaches are often computationally intensive and require significant processing time. To address these limitations, multiple complex-image models have been developed. Sunde introduced a more rigorous physical basis for image placement and provided practical closed-form expressions [7]. Afterwards, Dubanton and Deri proposed more sophisticated complex-image depths derived from exponential or logarithmic approximations of Carson's integral [9, 10].

Alvarado further refined these results by introducing a more accurate closed-form model suitable for engineering applications [11], while Pizarro and subsequently Noda presented double complex-image models in which two optimized image conductors substantially improve approximation accuracy across wide frequency ranges and geometric configurations [12, 13]. Finally, it is important to note that Kersting proposed a model that relies only on the first few terms of Carson's series expansion [14], and that Carson himself, in his original work, had already derived a closed-form solution, although this contribution remained largely unrecognized until recent years [15].

Closed-form solutions for ground-return impedance have been extensively studied, and numerous works have analyzed their sensitivity to soil resistivity, conductor height, horizontal spacing, operating frequency, and ground stratification. These studies provide valuable insights into the accuracy and limitations of both classical and modern approximations, including Carson-based series, image-method formulations, and recent closed-form expressions. However, despite this extensive body of research, no previous work has evaluated the applicability or accuracy of these closed-form models for short-circuit studies in unbalanced distribution systems, in which mutual coupling and asymmetrical feeder geometries significantly influence the results [1], [8], [16–18].

For short-circuit studies in balanced systems, the symmetrical components method is widely employed. However, in distribution networks, which are inherently unbalanced due to their asymmetrical geometry, directly applying this method can lead to significant errors [19]. In such cases, a phase-domain analysis becomes essential, as it explicitly represents system unbalance and incorporates the mutual impedances among phases. Recent studies have integrated phase-domain representations directly into the symmetrical components framework, enabling a more accurate treatment of unbalance and mutual coupling effects.

These hybrid approaches are increasingly being used to analyze distribution systems with high penetration of distributed generation (DG), where the interaction between unbalanced network conditions and inverter-based resources necessitates detailed, phase-resolved short-circuit modeling [20–22].

Since short-circuit studies in distribution systems with distributed generation are now commonly performed using phase-domain analysis, which directly relies on an accurate representation of the earth-return impedance, it is essential to assess whether the most widely used closed-form formulations implemented in commercial software and research projects are suitable for this purpose [1], [17]. Evaluating their performance makes it possible to determine the reliability of these closed-form expressions for short-circuit calculations and, consequently, their appropriateness for analyzing unbalanced distribution systems with high DG penetration.

This study provides a comprehensive phase-domain short-circuit analysis of three distribution systems, incorporating the most widely used closed-form formulations for ground-return impedance. The novelty of this study lies in evaluating, for the first time, the direct impact of these formulations on short circuit results in unbalanced distribution networks, rather than focusing solely on their electromagnetic accuracy as in prior contributions.

The main contributions of this work are summarized as follows:

- To quantify the percentage error introduced in short-circuit calculations when different closed-form ground-return impedance formulations are employed.
- To identify the critical network configurations in which the selection of the impedance model exerts a significant influence on short-circuit results.

2. Materials and Methods

2.1. Closed-form impedance formulas

The principal closed-form expressions for earth-return impedance in overhead distribution lines are derived from Carson's formulations [16, 17], [23] and include the following components:

$$Z' = R_c + jX_c + jX_g + Z_e \tag{1}$$

Where R_c and X_c denote the conductor resistance and reactance, respectively, X_g represents the reactance associated with the geometric distribution of the conductors, and Z_e is the earth-return impedance. As is customary for short overhead lines (less than 50 miles), capacitance is neglected because it has a negligible impact at power frequencies [24].

The conductor resistance is typically provided by the manufacturer. The conductor reactance is given by:

$$X_c = \frac{\omega\mu_0\mu_r}{2\pi} ln \frac{r_{ext}}{GMR} \tag{2}$$

where

- ω is the complex frequency, rad/s.
- μ_0 is the vacuum permeability, H/mile.
- μ_r is the relative magnetic permeability.
- r_{ext} is the conductor radius, (ft),
- GMR is the geometric mean radius of the conductor, (ft).

The extended Carson model is presented in (3)-(6). In these expressions, Z_c denotes the conductor self-impedance and is given by the sum of R_c and jX_c. The second term on the right-hand side of equation (3) and the first term on the right-hand side of equation (4) correspond to the reactance associated with line geometry (jX_g). The last terms on the right-hand sides of (3) and (4) represent the earth-return impedance (Z_e).

$$Z'_{(self),carson} = Z_c + j\omega\frac{\mu_0}{2\pi} ln \frac{2h_s}{r_{ext}} + j\omega\frac{\mu_0\mu_r}{\pi} J_s \tag{3}$$

$$Z'_{(mutual),carson} = j\omega\frac{\mu_0}{2\pi} \ln \frac{\sqrt{(h_i + h_j)^2 + (x_{ij})^2}}{\sqrt{(h_i - h_j)^2 + (x_{ij})^2}} + j\omega\frac{\mu_0\mu_r}{\pi} J_m \tag{4}$$

where

$$J_s = P_s + jQ_s = \int_0^\infty \frac{e^{-2h_s\lambda}}{\lambda + \sqrt{\lambda^2 + j\omega\mu_0\mu_r/\rho}}\, d\lambda \tag{5}$$

$$J_m = P_m + jQ_m = \int_0^\infty \frac{e^{-2(h_i+h_j)\lambda}}{\lambda + \sqrt{\lambda^2 + j\omega\mu_0\mu_r/\rho}} Cos(\lambda x)\, d\lambda \tag{6}$$

and

- ρ is the ground resistivity, $\Omega \cdot ft$.
- h_s, h_i, h_j are the conductor heights above ground (ft).
- x is the horizontal separation between conductors (ft).

The solution to the infinite integrals in equations (5) and (6) was initially expressed by Carson as a closed-form solution. However, owing to the limited computational resources available at the time, he presented the result as an infinite-series expansion. Consequently, subsequent approximations became necessary to obtain practical formulations for earth-return impedance [1], [15]. This section summarizes the principal closed-form solutions for low frequencies, starting with the single- and double-logarithmic approximations to Carson's integrals [9–13], followed by the computational adaptations of the corresponding infinite-series expressions [14], [23].

2.1.1. Dubanton

The first approximation considered in this study for the self-impedance of an overhead conductor was proposed by Dubanton in 1969, based on the concept of complex depth. Subsequently, in 1976, Gary derived an expression for the mutual impedance between two overhead conductors, thereby completing Dubanton's formulation [9]. It is important to note that both expressions are empirical in nature.

$$Z_{\text{self}} = Z_c + j\omega\frac{\mu_0}{2\pi} \ln \frac{2(h_s + p)}{r_{\text{ext}}} \tag{7}$$

$$Z_{\text{mutual}} = j\omega\frac{\mu_0}{2\pi}\ln\frac{\sqrt{(h_i+h_j+2p)^2+(x_{ij})^2}}{\sqrt{(h_i-h_j)^2+(x_{ij})^2}} \quad (8)$$

where the complex depth (ft) is defined as:

$$p = \sqrt{\frac{\rho}{j\omega\mu_0\mu_r}} \quad (9)$$

2.1.2. Deri

In 1981, Deri et al. [10] provided a mathematical validation of the impedance formulas proposed by Dubanton and Gary. Their final closed-form solution is given by:

$$J_{s,\text{DERI}} = \frac{1}{2}\ln\frac{h_s+p}{h_s} \quad (10)$$

$$J_{m,\text{DERI}} = \frac{1}{2}\ln\frac{\sqrt{(h_i+h_j+2p)^2+(x_{ij})^2}}{\sqrt{(h_i+h_j)^2+(x_{ij})^2}} \quad (11)$$

2.1.3. Alvarado

A year later, in 1982, Alvarado et al. improved the impedance formulas developed by Deri [11]. Their main contribution was proposing an approximation that retained more terms than Deri's formulation. The resulting expression is given by:

$$J_{s,\text{ALVARADO}} = \frac{1}{2}\ln\frac{h_s+p}{h_s} - \frac{1}{24}\frac{1}{\left(\frac{h_s+p}{p}\right)^3} \quad (12)$$

$$J_{m,\text{ALVARADO}} = \frac{1}{4}\ln\frac{\left(1+\frac{p}{h}\right)^2+\tau^2}{1+\tau^2} - \frac{1}{48}\left(\frac{1}{\left[\frac{h}{p}(1+j\tau)+1\right]^3}+\frac{1}{\left[\frac{h}{p}(1-j\tau)+1\right]^3}\right) \quad (13)$$

where

$$\tau = \frac{x_{ij}}{h_i+h_j}, h = \frac{h_i+h_j}{2} \quad (14)$$

2.1.4. Pizarro and Eriksson

Pizarro and Eriksson [12] introduced a double logarithmic approach in 1991 to solve (3)-(6), as summarized below:

$$J_s = j\omega\frac{\mu_0}{2\pi}\left[\begin{array}{r} A\ln\left(\frac{\sqrt{(2h_s+2\alpha p)^2}}{r_{\text{ext}}}\right) \\ +(1-A)\ln\left(\frac{\sqrt{(2h_s+2\beta p)^2}}{r_{\text{ext}}}\right)\end{array}\right] \quad (15)$$

$$J_m = j\omega\frac{\mu_0}{2\pi}\left[\begin{array}{r} A\ln\frac{\sqrt{(h_i+h_j+2\alpha p)^2+x_{ij}^2}}{\sqrt{(h_i-h_j)^2+(x_{ij})^2}} \\ +(1-A)\ln\frac{\sqrt{(h_i+h_j+2\beta p)^2+x_{ij}^2}}{\sqrt{(h_i-h_j)^2+(x_{ij})^2}}\end{array}\right] \quad (16)$$

Where the constants were obtained using the least-squares method, yielding $A = 0.1159, \alpha = 0.2258$ and $\beta = 1.1015$. These values are valid for both self and mutual impedance.

2.1.5. Noda

In 2006, Noda [13] extended the work of Pizarro and Eriksson by approximating A, α and β as functions of typical distribution- and transmission-system parameters, including frequency, ground resistivity, and conductor height. This refinement improved the accuracy of Pizarro's impedance model; however, it required the introduction of an additional variable, θ.

The formulas for calculating self and mutual impedance are given in (15) and (16). The parameters A and α are defined as follows:

$$A = \begin{cases} 0.07360, & \theta \le 50.45°, \\ 0.002474\,\theta - 0.05127, & \theta > 50.45°. \end{cases} \quad (17)$$

$$\alpha = \begin{cases} 0.1500, & \theta \le 50.45°, \\ 0.004726\,\theta - 0.08852, & \theta > 50.45°. \end{cases} \quad (18)$$

$$\beta = \frac{1-A\alpha}{1-A} \quad (19)$$

Where, for self-impedance, $\theta = 0°$, whereas for mutual-impedance calculations, it is given by:

$$\theta = \tan^{-1}\frac{x_{ij}}{h_i+h_j} \quad (20)$$

2.1.6. Galloway

Galloway's impedance formulas [23] constitute a numerical adaptation of Carson's complete series expansion and are expressed as follows:

$$J_s = \frac{\pi}{8}(1-S_4)+\frac{1}{2}\left(\ln\frac{2}{\gamma r_{ij}}\right)S_2+\frac{1}{2}\theta S_2'-\frac{\sigma_1}{\sqrt{2}}+\frac{\sigma_2}{2}+\frac{\sigma_3}{\sqrt{2}} \quad (21)$$

$$J_m = \frac{1}{4} + \frac{1}{2}\left(\ln\frac{2}{\gamma r_{ij}}\right)(1 - S_4) - \frac{1}{2}\theta S_4' + \frac{\sigma_1}{\sqrt{2}} - \frac{\pi}{8}S_2 + \frac{\sigma_3}{\sqrt{2}} - \frac{\sigma_4}{2} \quad (22)$$

Where γ denotes Euler's constant (1.7811), and $S_2, S_2^{'}, S_4, S_4^{'}, \sigma_1, \sigma_2, \sigma_3$ and σ_4 are infinite series that are described in detail in [8].

And

$$r_{ij} = \sqrt{\frac{\omega\mu_0\mu_r}{\rho}}\, D_{ij} \quad (23)$$

$$D_{ij} = \sqrt{(h_i + h_j)^2 + (x_{ij})^2} \quad (24)$$

for self-impedance calculations, $D_{ij} = 2h_s$.

Finally, the upper limit k in the summations was set to 35 , since Ramos et al. [17] showed that, for typical 60-Hz distribution-line configurations, retaining 35 terms in the Galloway formulas yields a deviation of less than $1 \times 10 - 7$ relative to the numerical evaluation of Carson's integral. Accordingly, this work adopts the 35-term Galloway formulation as an exact numerical reference for Carson's integral, denoted as the Carson model.

2.1.7. Kersting

Kersting's formulas [14] are based on Carson's recommendations for r less than 1/4 [4]. Most distribution-line configurations under steady-state conditions ($f = 60Hz$) and with standard soil resistivity ($\rho = 328.084\ \Omega \cdot \text{ft}$) fall within this range.

$$J_s = \frac{\pi}{8} \quad (25)$$

$$J_m = -0.0386 + \frac{1}{2}\ln\frac{2}{r_{ij}} \quad (26)$$

Finally, to obtain the phase-impedance matrix (Z) for all line models, Kron reduction must be applied to the corresponding primitive impedance matrix (Z´). The complete procedure is described in detail in [18].

2.2. Distribution Line Configurations

After presenting the closed-form earth-return impedance formulas, the next step is to model the line configurations for the short-circuit studies. To this end, an algorithm was developed in Mathematica software to incorporate all the earth-impedance formulations. The algorithm was then validated against benchmark data reported in the original references.

Two distribution systems were selected for the studies. The IEEE 13-node test feeder was chosen because it is compact yet highly unbalanced, whereas the IEEE 34-node test feeder [25] was selected because it features long distribution lines. The test cases considered in this study are based on these IEEE test feeders, which were developed from actual distribution system configurations and are widely recognized as benchmark models for validating analysis methodologies in distribution networks. These feeders are now extensively used in short-circuit studies, phase-domain modeling, zero-sequence impedance evaluation, and analyses involving high DG penetration. Their broad adoption and sustained acceptance in recent literature support the representativeness and suitability of these configurations for the accuracy assessment conducted in this work [22], [26–28].

The IEEE 13-node test feeder includes five overhead line configurations (601–605); however, only three correspond to distinct conductor spacings. Accordingly, results are reported only for configurations 601, 603, and 605, which represent three-phase, two-phase, and single-phase lines, respectively. The principal data for these configurations are summarized in Table 1.

Table 1. Overhead line-configuration data for the IEEE 13-node test feeder

Config.	Phasing	Phase Size	Neutral Size	Spacing ID
601	B A C N	556.5 kcmil	4/0 AWG	500
603	C B N	1/0 AWG	1/0 AWG	505
605	C N	1/0 AWG	1/0 AWG	510

The IEEE 34-node test feeder includes four overhead line configurations. Configurations 300 and 301 are three-phase, configuration 305 is two-phase, and configuration 302 is single-phase. However, their conductor spacings are identical to those of the IEEE 13-node test feeder, as summarized in Table 2.

Table 2. Overhead line-configuration data for the IEEE 34-node test feeder

Config.	Phasing	Phase Size	Neutral Size	Spacing ID
300	B A C N	1/0 AWG	1/0 AWG	500
301	B A C N	#2 AWG	#2 AWG	500
302	A N	#4 AWG	#4 AWG	510
305	A C N	#4 AWG	#4 AWG	505

The conductor data for both systems are available in [25] and are summarized in Table 3. Conductor sizes are expressed in kcmil or AWG, and all conductors are of the ACSR type. Resistance values are reported in ohms per mile at 60 Hz and 50 °C. The conductor's external diameter is given in inches, and the geometric mean radius (GMR) is expressed in feet.

Table 3. Conductor data

Size (kcmil o AWG)	Type	Resistance (Ω/milla)	Diameter (inches)	GMR (ft)
556.5	ACSR	0.1859	0.927	0.03130
4/0		0.592	0.563	0.00814
1/0		1.12	0.398	0.00446
#2		1.69	0.316	0.00418
#4		2.55	0.257	0.00452

The configurations corresponding to the different spacings are illustrated in Figure 1, with all distances expressed in feet. Phase conductors are depicted in black, whereas the neutral conductor is shown in gold. The distance between the pole and the neutral conductor is 0.5 ft.

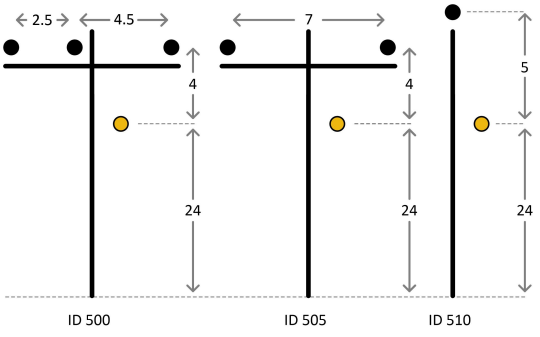


Figure 1. Line spacings

To properly assess the accuracy of the closed-form earth-return impedance formulas, Table 4 presents the maximum percentage error of the different impedance formulas relative to the Carson series for configurations 601, 603, and 605, corresponding to three-phase, two-phase, and single-phase cases, respectively. The modeling was performed at 60 Hz, assuming a soil resistivity of $100\Omega \cdot m$. This value was selected because typical soil resistivity ranges from 50 to $200\Omega \cdot m$, and $100\Omega \cdot m$ represents a moderately conservative condition for evaluating ground-return impedance behavior. Moreover, $100\Omega \cdot m$ is widely used as a benchmark in recent academic studies and IEEE technical reports that assess image-based methods, Carson-derived models, and frequency-dependent impedance formulations [1], [8], [29,30]. The percentage error is computed as follows:

Table 4. Maximum percentage error for the different distribution-line configurations

	Real part	Imaginary part
Configuration 601		
Deri-Dubanton	0.96243 %	0.66959 %
Kersting	0.24885 %	0.26336 %
Pizarro - Eriksson	0.11925 %	0.09400 %
Alvarado	0.06726 %	0.07141 %
Noda	0.02956 %	0.01086 %
Configuration 603		
Deri-Dubanton	118801%	0.76550 %
Kersting	0.19441 %	0.31422 %
Pizarro - Eriksson	0.14941 %	0.10769 %
Alvarado	0.08854 %	0.08228 %
Noda	0.03390 %	0.01212 %
Configuration 605		
Deri-Dubanton	0.18572 %	0.25734 %
Kersting	0.03160 %	0.11156 %
Pizarro - Eriksson	0.02340 %	0.03635 %
Alvarado	0.01389 %	0.02801 %
Noda	0.00532 %	0.00393 %

$$\text{error} = \max\left[\frac{|Z_{\text{Carson}} - Z_{\text{E-formulae}}|}{Z_{\text{Carson}}}\right] \times 100\% \quad (27)$$

It is important to emphasize that the Carson model is used as the reference in this work, since at low frequencies, there is no significant difference between the results obtained from Carson's formulation and those obtained from Wise's generalized earth-return impedance formula [6]. Wise's model is regarded as the most comprehensive because it accounts for soil permeability and displacement currents. Nevertheless, at power frequencies, these displacement currents are negligible, and both formulations converge to essentially the same results [1].

From Table 4, it is evident that Noda's formulas yield the smallest percentage errors in both the real and imaginary components. Alvarado's formula provides the next most accurate approximation, with maximum percentage errors below 0.09%. Pizarro's approximation follows, exhibiting, with a maximum error of 0.15% in two-phase configurations. In contrast, the Kersting and Deri formulas produce larger discrepancies, with maximum errors of approximately 0.32% and 1.19%, respectively. These findings are consistent with the conclusions drawn by Martins [16] and Ramos-Leaños [17], who identified Noda's formulations as the most accurate closed-form approximation to the Carson series for typical distribution-line configurations. It is also worth noting that the procedure used in this study to compute percentage differences is identical to that adopted by Papadopoulos [1].

From these tables, it is evident that the Deri and Dubanton formulas yield identical impedance matrices for the line configurations analyzed in this study. Accordingly, they are combined and treated as a single formulation in the subsequent analysis.

Finally, it should be noted that the smallest percentage errors occur in the single-phase configurations (605). Moreover, because single-phase lines are typically short, these configurations are expected to exert only a limited influence on short-circuit calculations. In contrast, the two-phase configurations (603) exhibit larger percentage errors and, since such lines are generally of medium length, they are more likely to have a pronounced impact on short-circuit study results.

3. Results and discussion

This section presents the principal results of the short-circuit studies conducted on the IEEE 13-node and 34-node test feeders. Three case studies are examined to represent distinct distribution-system conditions: a 13-bus feeder with short lines (less than 1 mile), a 34-bus feeder with long lines (up to 15 miles), and a third scenario characterized by extremely long lines (greater than 15 miles).

The algorithm used for these studies was implemented in Mathematica, and its results match exactly the reference data published by the IEEE Power and Energy Society [31]. Phase-domain analysis was adopted for the short-circuit calculations, since it is widely recognized as a benchmark approach for distribution-system studies [20], [27]. Detailed descriptions of this method are available in [32, 33]. Standard assumptions were applied throughout the short-circuit simulations: the pre-fault voltage was set to 1 pu, and static loads were neglected.

Before presenting the results of the short-circuit studies, it is important to clarify that, for ease of analysis, all outcomes are reported as the maximum percentage error, computed using equation (28). In this expression, I_{Carson} denotes the short-circuit current obtained with the Carson model, and $I_{(E-formulae)}$ denotes the current obtained with the corresponding closed-form earth-impedance formulations. Short-circuit current values in amperes are available from the authors upon request.

$$\text{error} = \max\left[\frac{|I_{\text{Carson}} - I_{\text{E-formulae}}|}{I_{\text{Carson}}}\right] \times 100\% \quad (28)$$

The results for the different fault types are presented below. It is important to note that three-phase and line-to-line fault outcomes are not reported, since no significant differences were observed among the impedance models. This behavior is expected because these faults do not involve an earth-return path; consequently, all formulations converge, resulting in essentially identical short-circuit current magnitudes.

3.1. IEEE 13-Node Test Feeder

The first short-circuit study is conducted on the IEEE 13-node test feeder, whose one-line diagram is shown in Figure 2. In the diagram, solid, dashed, and dotted lines denote three-phase, two-phase, and single-phase line configurations, respectively.

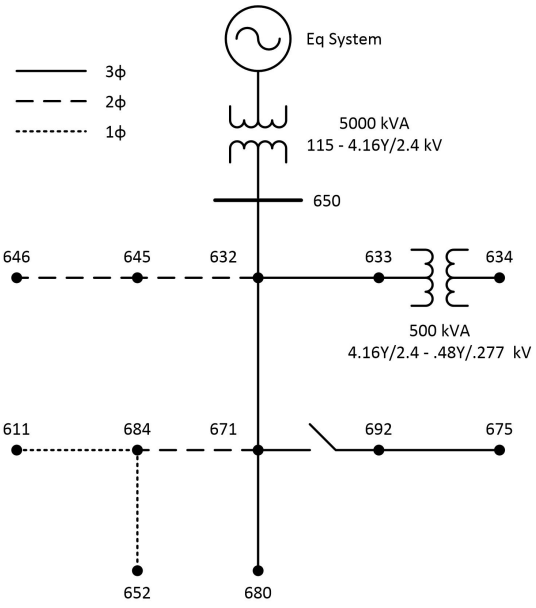


Figure 2. IEEE 13-node test feeder

The complete system data is available in [31]. Table 5 summarizes the main line-segment data.

Table 5. Line-segment data for Case Study1

From Node	To Node	Config	Phases	Length (ft)	Overhead or Underground
632	633	602	3	500	A
632	645	603	2	500	A
632	671	601	3	2000	A
645	646	603	2	300	A
650	632	601	3	2000	A
671	680	601	3	1000	A
671	684	604	2	300	A
684	611	605	1	300	A
684	652	607	1	800	U
692	675	606	3	500	U

Segment 671–692 is not included in the table because it corresponds to a switch with zero length. The impedance matrices for the underground line segments 684-652 and 692-675 were taken from [31]. These matrices were kept unchanged, because their differences relative to the Carson model are less than 0.01% [34].

3.1.1. Three-Phase-to-Ground Fault

Figure 3 presents the maximum percentage differences in short-circuit current between the Carson Model and the closed-form earth-impedance formulations for a three-phase-to-ground fault.

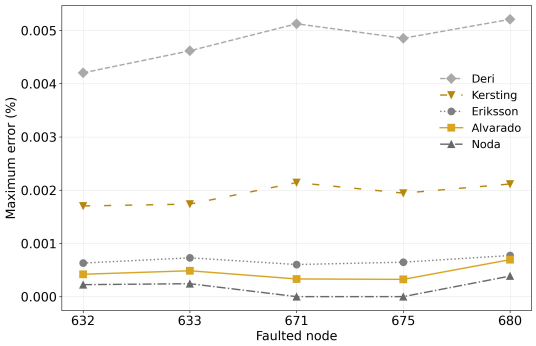


Figure 3. Maximum percentage error for a three-phase-to-ground fault

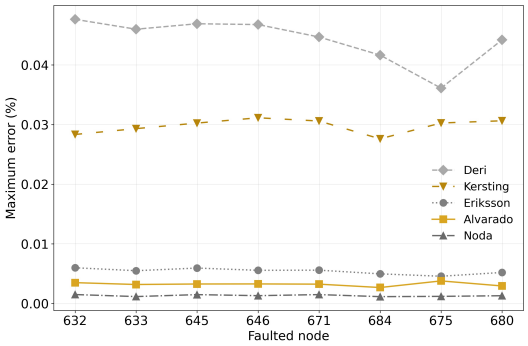


Figure 4. Maximum percentage error for a line-to-line-to-ground fault

From Figure 3, it is evident that all closed-form impedance formulations yield errors below 0.006% for this fault type. Among them, Noda's formulation provides the closest agreement with the Carson model, followed in decreasing accuracy by Alvarado, Eriksson, Kersting, and Deri. The largest percentage error occurs at node 680, which lies along the longest three-phase path from the short-circuit source, approximately 5000 ft.

3.1.2. Line-to-Line-to-Ground Fault

Figure 4 depicts the maximum percentage errors in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a line-to-line-to-ground fault.

As Figure 4 shows, Noda's formulations provide the closest agreement, followed by Alvarado's and Eriksson's. The Kersting and Deri formulations also perform well, with errors remaining below 0.05%.

The highest average error occurs at node 680, which is located along the longest two-phase path from the short-circuit source (approximately 5000 ft). Notably, nodes 645 and 646 also exhibit relatively large errors because they are two-phase nodes; however, since the associated line sections are short, their overall effect remains limited.

3.1.3. Single-Line-to-Ground Fault

Figure 5 presents the maximum percentage differences in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a single-line-to-ground fault.

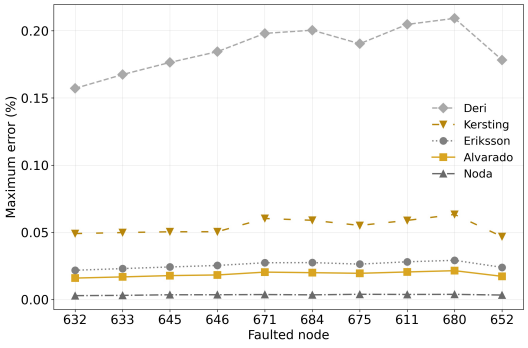


Figure 5. Maximum percentage error for a single-line-to-ground fault

From Figure 5, it is evident that, for this fault type, all earth-impedance formulations yield errors below 0.21%. Consistent with the other fault cases, Noda's formulation provides the closest agreement with the Carson model, followed closely by Alvarado's and Eriksson's, and then by Kersting's and Deri's.

3.1.4. Analysis for Case Study 1

Based on the short-circuit results, it can be concluded that the largest discrepancies generally occur for single-line-to-ground faults (0.21%), followed by line-to-line-to-ground faults (0.05%), and, finally, three-phase-to-ground faults (0.006%). Across all fault types, Noda's formulation provides the most accurate approximation,

closely followed by Alvarado's and Eriksson's. The remaining earth-impedance formulations introduce errors that may be non-negligible in certain applications, particularly when longer line sections are involved.

An updated analysis is conducted to examine the relationship between the modeling errors associated with the line configurations and the errors observed in the short-circuit studies, with the aim of determining whether a correlation exists between these parameters. Because node 680 exhibits the largest short-circuit error and configuration 601 connects the source node (650) to node 680, the assessment focuses on the three-phase configuration 601. For this configuration, the reactive component is approximately three times larger than the resistive component; therefore, the imaginary part is adopted as the reference for the correlation analysis. Table 6 summarizes the relationship between impedance-modeling errors, and the maximum short-circuit errors at node 680. The reported values are obtained by dividing the maximum short-circuit errors by the maximum errors in the imaginary part of the corresponding impedances, so that larger ratios indicate a stronger correlation.

Table 6. Relationship between impedance errors and short-circuit results

Formulations	Three--Phase-to-Ground Fault	Line-to-Line-to-Ground Fault	Single-Line-to-Ground Fault
Deri-Dubanton	0.008	0.066	0.313
Kersting	0.008	0.116	0.240
Eriksson	0.008	0.055	0.311
Alvarado	0.010	0.041	0.301
Noda	0.036	0.120	0.355

From Table 6, it can be concluded that impedance-modeling errors exert a stronger influence on line-to-ground faults, as reflected by the larger ratios. It should be noted that this assessment was performed only for node 680, which is a three-phase node. A more comprehensive understanding requires extending the same analysis to two-phase and single-phase nodes associated with longer line sections. This extended evaluation is addressed in the following cases.

3.2. Long distribution lines

The objective of this case study is to examine how line length affects the magnitude of short-circuit fault currents. The short-circuit analysis is performed on a section of the IEEE 34-node test feeder, as shown in the one-line diagram in Figure 6. In the diagram, solid, dashed, and dotted lines denote three-phase, two-phase, and single-phase line configurations, respectively.

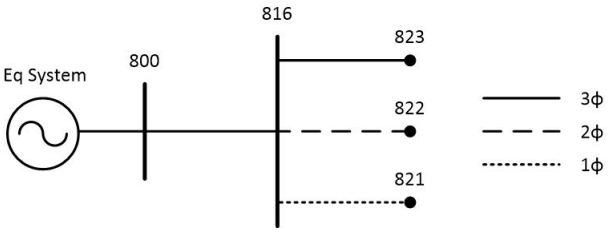


Figure 6. IEEE 34-nodetest feeder

The main feeder section up to node 816 is modeled as a three-phase line with configuration 300 and a length of 105,060 ft. From node 816 onward, three branches are considered: a three-phase branch with configuration 301, a two-phase branch with configuration 305, and a single-phase branch with configuration 302. The key line-segment data are summarized in Table 7.

Table 7. Line-segment data for Case Study 2

From Node	To Node	Config	Phases	Length (ft)	Overhead or Underground
800	816	300	3	105060	A
816	823	301	3	79200	A
816	822	305	2	79200	A
816	821	302	1	79200	A

The results are presented for terminal nodes located 3, 6, 9, 12, and 15 miles downstream from the three-phase node 816.

3.2.1. Three-Phase-to-Ground Fault

Figure 7 depicts the maximum percentage differences in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a three-phase-to-ground fault at node 823.

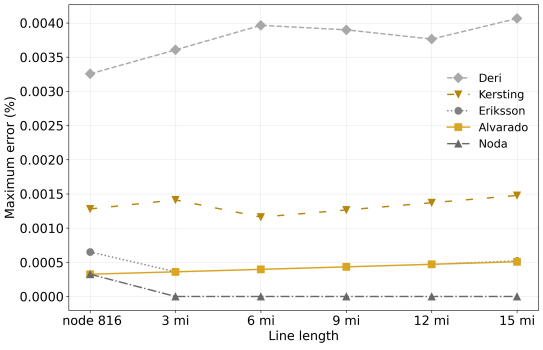


Figure 7. Maximum percentage error for a three-phase-to-ground fault in Case Study2 (node 823)

From Figure 7, it is evident that all closed-form formulations yield errors below 0.005% for this fault type.

3.2.2. Line-to-Line-to-Ground Fault

Figure 8 depicts the maximum percentage error in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a line-to-line-to-ground fault, corresponding to the two-phase case.

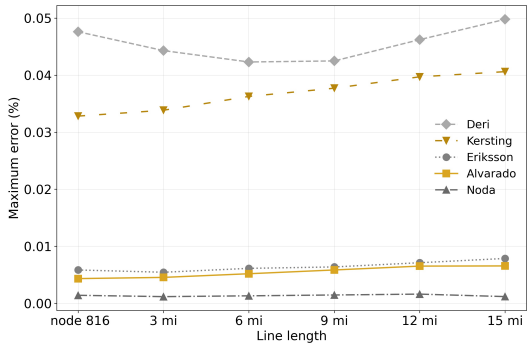


Figure 8. Maximum percentage error for a line-to-line-to-ground fault in Case Study 2 (node 822, two-phase)

The three-phase case is not reported because it is essentially identical to the two-phase case; both exhibit a maximum error of 0.05% for the Deri formulation.

3.2.3. Single-Line-to-Ground Fault

Figure 9 depicts the maximum percentage differences in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a single-line-to-ground fault, corresponding to the single-phase case.

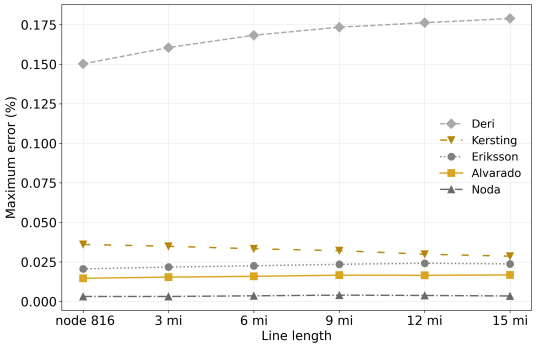


Figure 9. Maximum percentage error for a single-line-to-ground fault in Case Study 2 (node 821, single-phase)

From Figure 9, it is evident that Noda's formulations provide the closest agreement with the Carson model, followed by Alvarado's, Eriksson's, and Kersting's. The Deri formulations also perform satisfactorily, with errors remaining below 0.18%. It should be noted that the three-phase and two-phase cases exhibit trends comparable to those of the single-phase configuration; therefore, they are not reported.

3.2.4. Analysis for Case Study 2

Based on the short-circuit results, it can be concluded that the largest discrepancies generally occur for single-line-to-ground faults (0.19%), followed by double line-to-ground faults (0.05%), and, finally, three-phase-to-ground faults (0.005%). Across all fault types, Noda's formulations provide the most accurate approximation, closely followed by the Alvarado and Eriksson methods. The remaining formulations exhibit errors that may be non-negligible in line-to-earth fault calculations.

After analyzing the results of the first and second case studies, it is apparent that the line-to-ground fault constitutes the most critical scenario. Accordingly, a correlation analysis is conducted to examine the relationship between line-modeling errors and the outcomes of the line-to-ground fault studies. For configurations 302 and 305, which correspond to single-phase and two-phase lines, the resistive component is approximately twice the reactive component; therefore, these configuration are adopted as the reference cases for this assessment.

Table 8 summarizes the relationship between line-modeling errors and the maximum errors obtained in the short-circuit studies at a distance of 15 miles from node 816. The values reported in Table 8 are computed as the ratio of the maximum short-circuit errors to the maximum errors in the real part of the corresponding impedances. Larger values indicate a stronger correlation.

Table 8. Relationship between impedance-modeling error and line-to-earth short-circuit results

Formulations	Config 301 (3ϕ)	Config 305 (2ϕ)	Config 302 (1ϕ)
Deri-Dubanton	0.137	0.125	1,507
Kersting	0.163	0.106	1,245
Eriksson	0.148	0.137	1,634
Alvarado	0.180	0.168	2,039
Noda	0.107	0.093	1,177

From Table 8, it can be concluded that impedance-modeling error is most influential in the single-phase configuration, as evidenced by the larger ratios. This indicates a close relationship between line-modeling error and line-to-ground fault error for single-phase lines. Regarding the three-phase and two-phase configurations, the influence of modeling error on short-circuit calculations also remains significant, consistent with the findings of Case Study 1; nevertheless, the single-phase configuration continues to represent the critical case. Finally, to encompass all relevant operating conditions, extremely long lines must be considered, and this scenario is examined in the final case study.

3.3. Very long distribution lines

The objective of this third case study is to examine short-circuit fault magnitudes in extremely long distribution lines (greater than 50 miles), thereby covering all relevant scenarios in distribution-system modeling. Figure 10 illustrates the system adopted for this analysis, which is based on the IEEE 13-node test feeder. Specifically, the overhead line section between nodes 650 and 632 is extended to a total length of 50 miles. This line uses the three-phase configuration 601, with parameters provided in Table 1. Only the three-phase case is considered, since extremely long distribution lines are typically implemented using three-phase configurations.

3.3.1. Short-circuit results

Figure 10 illustrates the maximum percentage differences in short-circuit current between the Carson model and the closed-form earth-impedance formulations for a single line-to-ground fault.

From Figure 10, it is evident that Noda's formulations provide the closest agreement with the Carson model, followed by Alvarado's, Eriksson's, and Kersting's. The Deri formulations also perform satisfactorily, with errors remaining below 0.27%.

The results for the three-phase-to-ground and double line-to-ground faults are not reported because they exhibit behavior similar to that shown in Figures 7 and 8 for Case Study 2. The maximum percentage error is 0.007% for the three-phase-to-ground fault and 0.04% for the double line-to-ground fault.

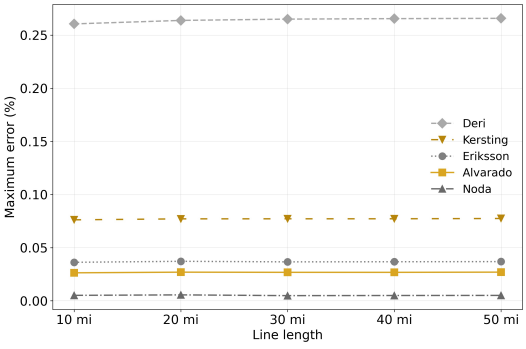


Figure 10. Maximum percentage error for a single-line-to-ground fault in Case Study 3.

3.3.2. Analysis for Case Study 3

The three-phase-to-ground fault study shows that all closed-form earth-impedance formulations yield maximum errors below 0.007%. A consistent trend emerges when these results are compared with those of Case Studies 1 and 2. Specifically, Case Study 1 reported a maximum error of 0.006% at 1 mile from the source, whereas Case Study 2 showed a maximum error of 0.005% at 15 miles. Collectively, these findings indicate that the closed-form earth-impedance formulations considered are sufficiently accurate for modeling three-phase-to-ground faults.

For line-to-line-to-ground faults, all closed-form earth-impedance formulations exhibit maximum errors of approximately 0.04%. This result is consistent with Case Studies 1 and 2, which reported maximum errors of 0.05% at 2000 ft and 15 miles, respectively. Across all cases, the maximum error remains below 0.05%, indicating that the closed-form earth-impedance formulations considered are sufficiently accurate for modeling this fault type.

For line-to-earth (single-line-to-ground) faults, all closed-form earth-impedance formulations exhibit maximum errors of 0.27%. Case Study 1 reported a maximum error of 0.21% at 5000 ft, whereas Case Study 2 showed a maximum error of 0.19% at 15 miles. Across all three case studies, the maximum error remains below 0.27%.

Based on the results of this short-circuit study, it can be concluded that the largest discrepancies generally occur for line-to-ground faults (0.27%), followed by double line-to-ground faults (0.04%), and, finally, three-phase-to-ground faults (0.007%). Among the evaluated formulations, Noda's model provides the most accurate approximation across all fault types, closely followed by those of Alvarado and Eriksson. By contrast, the remaining formulations may introduce non-negligible errors in line-to-earth fault calculations, particularly in applications that demand high accuracy. These conclusions are consistent with the trends observed in the previously analyzed case studies and are directly relevant to ongoing research examining the influence of distributed generation on short-circuit current magnitudes [22], [35, 36].

A correlation analysis is conducted to examine the relationship between line-modeling error magnitudes and the outcomes of the short-circuit studies. For configuration 601, the reactive component is approximately three times larger than the resistive component; therefore, the reactive part is adopted as the reference for this assessment. Table 9 presents the relationship between line-modeling errors, and the maximum short-circuit errors observed at a distance of 50 miles from node 650. The values reported in Table 9 are computed by dividing the maximum short-circuit errors by the maximum modeling errors associated with the reactive part of the corresponding impedances.

Table 9. Relationship between impedance-modeling error and short-circuit results

Formulations	Three--Phase-to-Ground Fault	Line-to-Line-to-Ground Fault	Single-Line-to-Ground Fault
Deri-Dubanton	0.009	0.056	0.397
Kersting	0.010	0.116	0.294
Eriksson	0.009	0.046	0.392
Alvarado	0.009	0.040	0.378
Noda	0.013	0.120	0.467

From Table 9, it can be concluded that impedance-modeling errors are most consequential for line-to-ground faults, as evidenced by the strong correlation observed for most closed-form earth-impedance formulations. By contrast, for three-phase and line-to-line faults, no clear relationship emerges between impedance-modeling errors and short-circuit results. Accordingly, for very long distribution lines, the choice of earth-impedance formulation has a negligible influence on short-circuit calculations under these fault conditions.

4. Conclusions

This study presented, in a comprehensive and accessible manner, the most widely used closed-form earth-return impedance formulations for short-circuit analysis in distribution systems. These formulations were applied to model single-phase, two-phase, and three-phase line configurations representative of practical distribution networks. Their impedance accuracy was quantified through percentage-error comparisons against the Carson reference model.

After the line-modeling stage, three short-circuit case studies were conducted. The first examined a feeder with short lines (less than 1 mile), the second considered medium-length lines (up to 15 miles), and the third evaluated very long lines (up to 50 miles). In each case, the percentage error between short-circuit currents computed using closed-form formulations and those obtained with the Carson model was determined.

Across all studies, line-to-ground faults consistently produced the largest percentage errors, followed by line-to-line-to-ground faults, whereas three-phase-to-ground faults exhibited the smallest errors. Notably, three-phase and line-to-line faults exhibited negligible differences between the studied formulations and the Carson model.

For all cases, the maximum error for three-phase-to-ground faults remained below 0.007%, and for line-to-line-to-ground faults below 0.04%, confirming that all closed-form formulations considered are adequate for modeling these fault categories. In contrast, line-to-ground faults reached a maximum error of 0.27%, indicating that the selection of the earth-return impedance formulation can be consequential in studies where high short-circuit accuracy is required.

Among the evaluated formulations, Noda's model consistently provided the most accurate approximation to the Carson reference, closely followed by Alvarado's and Eriksson's formulations, while Kersting's and Deri's yielded the largest deviations. In general, formulations with more elaborate complex-image representations tended to achieve higher accuracy. Nevertheless, except for line-to-ground faults, simpler formulations appear sufficient for short-circuit studies and may be preferred when reducing computational effort is a priority.

The correlation analysis further showed that, for three-phase and line-to-line-to-ground faults, impedance-modeling errors do not translate into noticeable short-circuit calculation errors. By contrast, a strong correlation was observed for line-to-ground faults, particularly in single-phase configurations, underscoring that accurate earth-return impedance modeling is critical in these scenarios.

Overall, the closed-form earth-return impedance formulations investigated are suitable for short-circuit analysis in distribution systems. However, for single-line-to-ground faults in single-phase lines, the choice of closed-form formulation is crucial to ensuring accurate short-circuit results.

Contributor Roles

- **Allen A. Castillo Barrón:** Conceptualization, formal analysis, methodology, project administration, writing – original draft and writing – review & editing.
- **Gerardo Ayala Jaimes:** Resources, software, visualization, writing – original draft and writing – review & editing.
- **Alejandra Jiménez Vega:** Investigation, supervision and validation.
- **Francisco J. Ramírez Arias:** Data curation, software and visualization.

References

[1] T. A. Papadopoulos, A. I. Chrysochos, C. K. Traianos, and G. Papagiannis, "Closed-form expressions for the analysis of wave propagation in overhead distribution lines," *Energies*, vol. 13, no. 17, p. 4519, Sep. 2020. [Online]. Available: https://doi.org/10.3390/en13174519

[2] R. Cleenwerck, H. Azaioud, R. Claeys, T. Coosemans, J. Knockaert, and J. Desmet, "An approach to the impedance modelling of low-voltage cables in digital twins," *Electric Power Systems Research*, vol. 210, p. 108075, Sep. 2022. [Online]. Available: http://doi.org/10.1016/j.epsr.2022.108075

[3] J. Zhang, F. Geth, R. Heidari, and G. Verbič, "Beyond simplifications: Evaluating assumptions for low-voltage network modelling in the DER era," *Sustainable Energy, Grids and Networks*, vol. 43, p. 101860, Sep. 2025. [Online]. Available: http://doi.org/10.1016/j.segan.2025.101860

[4] J. R. Carson, "Wave propagation in overhead wires with ground return," *Bell System Technical Journal*, vol. 5, no. 4, pp. 539–554, Oct. 1926. [Online]. Available: http://doi.org/10.1002/j.1538-7305.1926.tb00122.x

[5] W. H. Wise, "Effect of ground permeability on ground return circuits," *Bell System Technical Journal*, vol. 10, no. 3, pp. 472–484, Jul. 1931. [Online]. Available: http://doi.org/10.1002/j.1538-7305.1931.tb01287.x

[6] W. Wise, "Propagation of high-frequency currents in ground return circuits," *Proceedings of the IRE*, vol. 22, no. 4, pp. 522–527, Apr. 1934. [Online]. Available: http://doi.org/10.1109/JRPROC.1934.225868

[7] E. D. Sunde, *Earth Conduction Effects in Transmission Systems*. Dover Publications, 1968. [Online]. Available: https://upsalesiana.ec/ing35ar5r7

[8] H. Wohlfarth, "Calculation of ground impedances," *IEEE Transactions on Power Delivery*, vol. 39, no. 4, pp. 2113–2124, Aug. 2024. [Online]. Available: https://doi.org/10.1109/TPWRD.2024.3388570

[9] C. Dubanton, "Calcul approché des parametres primaires et secondaires d'une ligne de transport," *EDF Bulletin de la Direction des Études et Recherches*, no. 1, pp. 53–62, 1969, cited by: 16. [Online]. Available: https://upsalesiana.ec/ing35ar5r10

[10] A. Deri, G. Tevan, A. Semlyen, and A. Castanheira, "The complex ground return plane a simplified model for homogeneous and multi-layer Earth return," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 8, pp. 3686–3693, Aug. 1981. [Online]. Available: https://doi.org/10.1109/TPAS.1981.317011

[11] F. Alvarado and R. Betancourt, "An accurate closed-form approximation for ground return impedance calculations," *Proceedings of the IEEE*, vol. 71, no. 2, pp. 279–280, 1983. [Online]. Available: https://doi.org/10.1109/PROC.1983.12573

[12] M. Pizarro and R. Eriksson, "Modelling of the ground mode of transmission lines in time domain simulations," *7th Int. Symp. on High Voltage Engineering*, pp. 179–182, 1991.

[13] T. Noda, "A double logarithmic approximation of Carson's ground-return impedance," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 472–479, Jan. 2006. [Online]. Available: https://doi.org/10.1109/TPWRD.2005.852307

[14] W. Kersting and W. Phillips, "Distribution feeder line models," *IEEE Transactions on Industry Applications*, vol. 31, no. 4, pp. 715–720, 1995. [Online]. Available: https://doi.org/10.1109/28.395276

[15] T. Theodoulidis, "On the closed-form expression of Carson's integral," *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 59, no. 1, pp. 26–29, 2015. [Online]. Available: https://doi.org/10.3311/PPee.7894

[16] T. Martins, A. Lima, and S. Carneiro, "Effect of approximate impedance formulae on the accuracy of transmission line modelling," *IET Generation, Transmission & Distribution*, vol. 1, no. 4, pp. 534–539, Jul. 2007. [Online]. Available: https://doi.org/10.1049/iet-gtd:20060289

[17] O. Ramos-Leanos, J. L. Naredo, F. A. Uribe, and J. L. Guardado, "Accurate and approximate evaluation of power-line Earth impedances through the Carson integral," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 5, pp. 1465–1473, Oct. 2017. [Online]. Available: https://doi.org/10.1109/TEMC.2017.2679213

[18] S. Li and D. Tylavsky, "Analytic continuation as the origin of complex distances in impedance approximations," *International Journal of Electrical Power & Energy Systems*, vol. 105, pp. 699–708, Feb. 2019. [Online]. Available: https://doi.org/10.1016/j.ijepes.2018.09.022

[19] K. Gampa, S. A. Vemprala, and S. M. Brahma, "Errors in fault analysis of power distribution systems using sequence components approach," in *IEEE PES T&D 2010*. IEEE, 2010, pp. 1–6. [Online]. Available: https://doi.org/10.1109/TDC.2010.5484378

[20] R. S. F. Ferraz, R. S. F. Ferraz, A. C. Rueda-Medina, and M. H. M. Paiva, "Power flow and fault analysis using graph theory," in *2021 IEEE URUCON*. IEEE, Nov. 2021, pp. 6–11. [Online]. Available: https://doi.org/10.1109/URUCON53396.2021.9647053

[21] A. Suresh, K. Murari, S. Kamalasadan, and S. Paudyal, "Steady-state fault analysis of unbalanced power distribution network utilizing a novel sequence component based methodology," in *2023 IEEE Industry Applications Society Annual Meeting (IAS)*. IEEE,

Oct. 2023, pp. 1–6. [Online]. Available: https://doi.org/10.1109/IAS54024.2023.10406399

[22] J. He, Z. Li, W. Li, J. Zou, X. Li, and F. Wu, "Fast short-circuit current calculation of unbalanced distribution networks with inverter-interfaced distributed generators," *International Journal of Electrical Power & Energy Systems*, vol. 146, p. 108728, Mar. 2023.

[23] R. Galloway, W. Shorrocks, and L. Wedepohl, "Calculation of electrical parameters for short and long polyphase transmission lines," *Proceedings of the Institution of Electrical Engineers*, vol. 111, no. 12, p. 2051, 1964. [Online]. Available: https://doi.org/10.1049/piee.1964.0331

[24] M. Cenký, J. Bendík, B. Cintula, P. Janiga, A. Belan, and Z. Eleschová, "Three-phase and single-phase measurement of overhead power line capacitance evaluation," *Electrical Engineering*, vol. 105, no. 2, pp. 1045–1065, Jan. 2023. [Online]. Available: https://doi.org/10.1007/s00202-022-01714-1

[25] W. Kersting, "Radial distribution test feeders," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, ser. PESW-01, vol. 2. IEEE, 2001, pp. 908–912. [Online]. Available: https://doi.org/10.1109/PESW.2001.916993

[26] K. P. Schneider, B. A. Mather, B. C. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3181–3188, May 2018. [Online]. Available: https://doi.org/10.1109/TPWRS.2017.2760011

[27] I. Kim, "A short-circuit analysis algorithm capable of analyzing unbalanced loads and phase shifts through transformers using the Newton-Raphson power-flow calculation, sequence, and superposition methods," *International Transactions on Electrical Energy Systems*, vol. 31, no. 4, Oct. 2020. [Online]. Available: https://doi.org/10.1002/2050-7038.12653

[28] G. Meena, A. Mathur, V. P. Singh, K. Murari, and S. Kamalasadan, "Efficient evaluation of short-circuit faults in active distribution networks," in *2024 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*. IEEE, Dec. 2024, pp. 1–6. [Online]. Available: https://doi.org/10.1109/PEDES61459.2024.10961501

[29] T. F. Garbelim Pascoalato, A. R. Justo de Araújo, S. Kurokawa, and J. P. Filho, "Effects of frequency-dependent soil electrical parameters on sequence parameters of double-circuit transmission lines," *Electric Power Systems Research*, vol. 247, p. 111724, Oct. 2025. [Online]. Available: https://doi.org/10.1016/j.epsr.2025.111724

[30] A. A. Castillo, A. L. López, N. A. Barboza Tello, I. Uriarte, and A. Jiménez, "A comparison of models for calculating distribution lines impedance," in *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. IEEE, Nov. 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ROPEC.2018.8661470

[31] IEEE. Resources – IEEE PES Test Feeder. Institute of Electrical and Electronics Engineers. [Online]. Available: https://upsalesiana.ec/ing35ar5r32

[32] W. H. Kersting, *Distribution System Modeling and Analysis, Third Edition*. CRC Press, 2012. [Online]. Available: https://upsalesiana.ec/ing35ar5r33

[33] W. H. Kersting and G. Shirek, "Short circuit analysis of IEEE test feeders," in *PES T&D 2012*. IEEE, May 2012, pp. 1–9. [Online]. Available: http://doi.org/10.1109/TDC.2012.6281539

[34] W. H. Kersting and R. K. Green, "The application of Carson's equation to the steady-state analysis of distribution feeders," in *2011 IEEE/PES Power Systems Conference and Exposition*. IEEE, Mar. 2011, pp. 1–6. [Online]. Available: http://doi.org/10.1109/PSCE.2011.5772579

[35] F. Rongqi, L. Kuan, H. Qiang, J. Panpan, and W. Anning, "Calibration of distributed PV setting calculation model and simulation calculation model based on short-circuit test of distribution network," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. IEEE, Dec. 2023, pp. 603–607. [Online]. Available: http://doi.org/10.1109/ITAIC58329.2023.10408848

[36] J. Hoeksema, R. Torkzadeh, J. Van Waes, S. Cobben, and V. Ćuk, "A case study on the future changes in short circuit power to analyze the impact on voltage dips," *IET Conference Proceedings*, vol. 2023, no. 6, pp. 1330–1334, Jul. 2023. [Online]. Available: http://doi.org/10.1049/icp.2023.0707

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.06
pISSN: 1390-650X / eISSN: 1390-860X

Automatic classification of electrical complaints using decision trees and random forest: an applied study at CNEL EP

Clasificación automática de reclamos eléctricos mediante árboles de decisión y random forest: un estudio aplicado a CNEL EP

Francisco Javier Carpio Velasco[1,*], Gloria Margarita Garcés Beltrán[1]

Received: 03-10-2025, Received after review: 11-11-2025, Accepted: 18-11-2025, Published: 01-01-2026

Abstract

This study presents an automated classification system to prioritize electrical service complaints at CNEL EP. A total of 143,113 real records were processed through data cleaning, missing-value imputation, and the engineering of predictive variables reflecting complaint urgency and recurrence. Based on these criteria, the target variable "Priority" was defined to distinguish high-priority from normal complaints. Supervised learning models, specifically Decision Tree and Random Forest, were then trained using one-hot encoding and cross-validation. Random Forest delivered the best performance, achieving 91% accuracy and an AUC-ROC of 0.89. These results indicate that the proposed system can significantly improve technical resource allocation and reduce response times for electrical complaints. Moreover, the study demonstrates the feasibility of integrating machine-learning techniques into the operational management of electric distribution companies, enabling future enhancements and real-time deployment.

Keywords: electrical complaints, automatic classification, supervised learning, Decision Tree, Random Forest, machine learning

Resumen

Este estudio presenta un sistema de clasificación automática destinado a priorizar reclamos eléctricos en CNEL EP. Para ello se procesaron 143 113 registros reales, incluyendo limpieza de datos, imputación de valores faltantes y la creación de variables predictivas relacionadas con la urgencia y la recurrencia de los reclamos. A partir de estos criterios se construyó la variable objetivo "prioridad", que permite distinguir entre reclamos prioritarios y normales. Con esta información se entrenaron modelos de aprendizaje supervisado —árbol de decisión y random forest— aplicando técnicas como one-hot encoding y validación cruzada. El modelo random forest obtuvo los mejores resultados, con una exactitud del 91 % y un AUC-ROC de 0.89. Los resultados muestran que el sistema propuesto puede mejorar significativamente la asignación de recursos técnicos y los tiempos de respuesta ante reclamos eléctricos. Además, demuestra la viabilidad de incorporar técnicas de aprendizaje automático en la gestión operativa de empresas distribuidoras de electricidad, abriendo posibilidades para futuras mejoras y aplicaciones en tiempo real.

Palabras clave: reclamos eléctricos, clasificación automática, aprendizaje supervisado, árbol de decisión, random forest, aprendizaje automático

[1,*]Facultad de Ciencias de la Ingeniería, Universidad UTE, Santo Domingo de los Tsáchilas, Ecuador.
Corresponding author ✉: francisco.carpio@ute.edu.ec.

Suggested citation: F. J. Carpio Velasco and G. M. Garcés Beltrán "Automatic classification of electrical complaints using decision trees and random forest: an applied study at CNEL EP," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 78-89, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.06.

1. Introduction

The growing complexity of the electrical sector requires the adoption of knowledge-management tools and intelligent technologies that enhance operational decision-making and optimize customer service [1]. In this context, multiple studies have shown that structured service classification can guide sustainability and efficiency strategies in both industrial and social settings [2].

In the electrical sector, customer classification through clustering techniques has been widely investigated, with studies highlighting algorithms such as k-means, fuzzy clustering, and self-organizing maps. These methods enable the identification of consumption patterns and the segmentation of technical services [3, 4]. Their effectiveness has been validated in real-world applications, where they have demonstrated a strong capacity to improve resource allocation and operational response.

The prioritization of electrical complaints can be guided by studies that estimate customer interruption costs, distinguishing economic impacts according to the type of service disruption [5]. Similarly, the use of electrical data for automated inference has been examined in studies showing that relevant properties can be detected from consumption traces using supervised classification algorithms [6].

Efficient handling of electrical complaints directly influences customer satisfaction, as evidenced by recent studies in Latin American contexts [7]. In Ecuador, substantial concentrations of complaints have been documented in provinces such as Manabí, where consumers reported tariff increases higher than those observed in other cities such as Guayaquil [8]. Additionally, applied research in CNEL EP has identified deficiencies in service perception, underscoring the need to strengthen mechanisms for addressing external users' requests [9].

During the pandemic, excessive billing practices intensified, motivating proposals for agile service mechanisms [10]. In Ecuador, the quality of technical service has been evaluated using indicators such as FMIk and TTIk, which enable the identification of critical areas and justify the need for operational prioritization [11]. At the international level, cases such as Cartagena, Colombia, have shown that deficiencies in electrical service can trigger social tensions and community conflicts [12].

The quality of electrical service has been identified as a key determinant of customer satisfaction and loyalty, particularly in public monopoly contexts [13]. However, recent evidence indicates that perceived quality, evaluated through SERVQUAL dimensions, does not always translate into customer satisfaction, as shown in the case of IBEDC in Nigeria [14]. Similarly, staff customer orientation has been found to significantly shape service perception, as evidenced by studies conducted in Ghana [15].

In modern electrical networks, fault identification and classification are critical for ensuring operational stability. Advances in intelligent sensors and communication technologies have transformed traditional systems into smart grids capable of integrating distributed generation and automation. Sarathkumar et al. [16] note that current fault-detection and mitigation methods require a thorough review to address the limitations of intelligent systems, and they propose advanced technologies that enhance transient stability, self-recovery capability, and power-supply quality.

The incorporation of deep neural networks into electrical systems has proven an effective strategy for enhancing fault detection and classification. Alhanaf et al. [17] propose methods based on artificial neural networks (ANN) and one-dimensional convolutional neural networks (1D-CNN), which, using sensor data such as voltage and current, achieve accuracy rates above 99% in identifying faulty lines and fault types. These findings underscore the potential of supervised models to automate critical processes in electrical networks, a capability that can be extended to the technical classification of institutional complaints.

The reliability of modern electrical networks depends on the ability to identify and classify faults across multiple system layers. Krivohlava et al. [18] conducted a systematic review of 30 fault types in smart grid infrastructures, encompassing physical, software, communication, and business components. Their SGAM-based classification provides a robust technical framework that can be adapted to institutional contexts to improve the management of electrical complaints and strengthen operational resilience.

The automatic classification of electrical faults has advanced through hybrid algorithms that combine dimensionality reduction with supervised discrimination. Hosseinzadeh et al. [19] propose a model that integrates PCA, LDA, and k-NN, improving the accuracy of fault detection and classification in smart grids. This approach demonstrates that, given load variability and the limitations of traditional models, supervised learning algorithms provide an effective and adaptable solution that is also applicable to the technical management of electrical complaints.

The prediction of extreme events in electrical systems requires models that integrate both technical and operational criteria. Mohammadian et al. [20] propose a classifier based on Bayesian decision theory that not only minimizes prediction errors but also accounts for the cost of each preventive action. This approach enables resilience-oriented decision-making and can be extrapolated to the classification of institutional electrical complaints, particularly when prioritization must be based on impact and urgency.

Rana [21] presents a systematic review of 180 studies demonstrating that artificial intelligence–based ap-

proaches, including machine learning, digital twins, and self-healing grids, outperform traditional methods in accuracy, adaptability, and operational efficiency. These findings reinforce the feasibility of applying supervised models to classify institutional electrical complaints, anticipate faults, and optimize technical response.

In recent years, several enhancements to the random forest algorithm have been proposed to address limitations associated with tree correlation and reduced accuracy in certain decision-tree configurations. Sun et al. [22] developed an improved random forest that incorporates tree-selection mechanisms based on accuracy and correlation reduction, achieving superior performance over conventional implementations. Jing et al. [23] proposed a hybrid predictive-maintenance approach for electrical meters that combines LSTM and XGBoost with an improved optimization algorithm, achieving significant improvements in accuracy and operational efficiency. Hussain et al. [24] assessed multiple machine-learning algorithms, including random forest and gradient boosting, for predicting energy consumption in electric vehicles, confirming the effectiveness of ensemble methods for capturing complex patterns. Khalili et al. [25] applied CatBoost and ExtraTrees to the predictive maintenance of submersible pumps, achieving high accuracy and sensitivity in identifying critical faults. Collectively, these studies reinforce the applicability of ensemble models for prioritizing electrical complaints in institutional contexts.

Multiple studies have shown that machine-learning techniques can substantially optimize operational processes in electrical distribution companies. For example, Ahmad and Chen [26] developed ML-based models, including CDT, k-NN, and LRM, for short- and medium-term electricity-demand forecasting, achieving marked improvements in accuracy even when only incomplete climate data are available. Such contributions reinforce the relevance of applying machine-learning models to critical tasks in the electrical sector, such as the automated prioritization of complaints addressed in this study.

Zhang et al. [27] conduct a comprehensive review of machine-learning techniques for detecting non-technical losses in electrical distribution companies, showing that supervised algorithms can identify irregular consumption patterns and effectively support energy-fraud reduction.

According to Ausmus et al. [28], big data analytics has become a key tool in the electrical sector, enabling the processing of large data volumes from SCADA systems and phasor measurement units (PMUs) to improve real-time operational decision-making. This approach facilitates the integration of machine-learning techniques to optimize the prioritization of electrical complaints and promote more efficient management of technical resources.

Recent studies show that supervised decision-tree–based models enable the automatic identification of event zones in electrical systems, thereby improving outage management and asset maintenance [29]. This evidence supports the use of similar techniques for automated prioritization of electrical complaints, optimizing operational response and resource allocation.

In the field of incident classification in electrical distribution systems, several studies have proposed relevant methodologies. For example, Shen et al. [30] demonstrated that integrating dimensionality reduction techniques with convolutional neural networks (CNN) significantly improves the automatic detection of disturbances in electrical systems, reinforcing the value of machine-learning-based approaches for fault analysis and diagnosis.

Xu et al. [31] demonstrated that logistic regression and neural networks can be effectively applied to automatically identify fault causes in distribution systems, even under imbalanced data conditions and complex real-world scenarios.

Kumwenda et al. [32] developed a classification model for low-voltage networks using synthetic and real data, demonstrating that ensemble-based methods, particularly bagged trees, achieve high accuracy in identifying system disturbances.

In this context, the present work proposes and validates a supervised classification model for the automated prioritization of electrical complaints using a real dataset of 143,113 records from CNEL EP, Ecuador. The methodology includes data cleaning, the construction of derived variables capturing urgency and recurrence by zone and month, and the implementation of classifiers such as Decision Tree and Random Forest. The results show high accuracy in identifying priority complaints, demonstrating the feasibility of integrating machine-learning techniques into the operational management of the electrical sector.

2. Materials and Methods

In this study, an automated classification system for technical complaints in the Ecuadorian electrical system was developed using supervised learning algorithms. The methodology comprised six stages: collection and exploratory analysis of institutional data from CNEL EP; database preprocessing; feature engineering; dataset preparation; design and training of the classification models (Decision Tree and Random Forest); and, finally, performance evaluation using standard classification metrics. Figure 1 summarizes these stages.

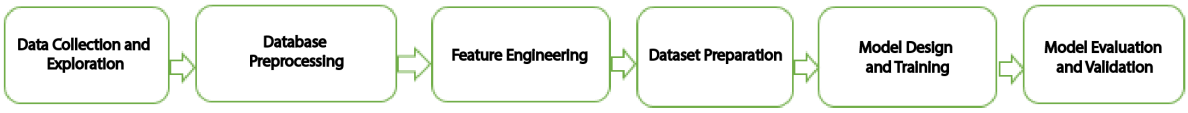


Figure 1. Overall methodological framework used in this study

2.1. Data Collection

An institutional database from CNEL EP was compiled, comprising 143,113 historical records of technical complaints. Each record contains variables such as business unit, month, process status, complaint type, and incident classification. This dataset provides a representative collection of real cases for training supervised models. Table 1 presents the first ten records, illustrating the structure and diversity of the data.

Table 1. Sample records from the CNEL EP electrical complaint database

Case ID	Business Unit	Month	Status	Complaint Type	Classification
01521	MAN	Jul-21	Closed	Technical	Voltage variation
01522	STE	Jul-21	Closed	Technical	No power in the area / short circuit
01523	GYE	Jul-21	Closed	Technical	Streetlight off at night
01524	EOR	Jul-21	Closed	Technical	No power in the area / short circuit
01525	EOR	Jul-21	Closed	Technical	No power in the area / short circuit
01526	MAN	Jul-21	Closed	Technical	No power in the area / short circuit
01527	MAN	Jul-21	Closed	Technical	No power in the area / short circuit
01528	GYE	Jul-21	Closed	Technical	Line on the ground
01529	EOR	Jul-21	Closed	Technical	No power in the area / short circuit
01530	GYE	Jul-21	Closed	Technical	Sparks at service drop/network/meter

2.2. Data Preprocessing

Data preprocessing involved removing of unnecessary spaces in text columns and correcting character-encoding issues by converting the "type of complaint" and "classification" fields from Latin1 to UTF-8. These steps ensured consistent textual values and prepared the data for subsequent analysis.

In addition, null values were verified, and appropriate transformations were applied to preserve data integrity. This preprocessing ensured consistent textual values and prepared the dataset for subsequent analysis.

2.3. Feature Engineering

To enhance model predictive performance and define a classification label consistent with the operational reality of electrical complaints, the original dataset was transformed to construct the target variable Priority. This variable was defined using two analytical criteria derived from the data: urgency and recurrence by zone and month.

Urgency

The urgency variable classifies each complaint based on keywords identified in the "Classification" column of Table 1. This categorization distinguishes complaints with immediate user impact from those requiring less urgent attention. Table 2 presents representative examples from the CNEL EP complaint database and assigns high-, medium-, or low-urgency levels according to the severity of the reported incident.

Table 2. Urgency-level assignment for electrical complaints in CNEL EP

Keyword(s)	Urgency level
"no power", "short circuit"	High
"voltage variation"	Medium
"streetlight off", "sparks at service drop"	Low

Recurrence by zone and month ($\mathbf{recurrence_{zone_{month}}}$)

This variable represents the frequency of complaints in each business unit during a specific month. It is computed as:

$(\text{recurrence}_{\text{zone}_{\text{month}}})$ = Number of complaints in the business unit during the month.

Example: if 120 complaints were recorded in the "MAN" business unit in July, then all complaints from "MAN" in that month are assigned $(\text{recurrence}_{\text{zone}_{\text{month}}}) = 120$. This allows critical zones with high complaint concentration to be identified.

Priority (target variable)

This variable combines urgency and recurrence to classify complaints into two categories:

The threshold was defined as the mean recurrence plus one standard deviation across all complaints, enabling automatic identification of the most critical cases.

This process yielded a representative label consistent with institutional criteria for priority attention.

The resulting "Priority" variable was added to the dataset as the model's target label, as shown in Table 3.

$$\text{Priority} = \begin{cases} \text{Priority}, & \text{yes Urgency} = \text{High or recurrence}_{\text{zone, month}} > \text{threshold} \\ \text{Normal}, & \text{otherwise} \end{cases} \quad (1)$$

Table 3. Sample complaints with derived variables

Case ID	Business Unit	Month	Status	Complaint Type	Classification	Priority
10901521	MAN	Jul-21	Closed	Technical	Voltage variation	Normal
10901522	STE	Jul-21	Closed	Technical	No power in the area / short circuit	High-priority
10901523	GYE	Jul-21	Closed	Technical	Streetlight off at night	Normal

2.4. Dataset Preparation

With the variables defined, the dataset was prepared for supervised learning. Categorical features such as business unit and complaint type were encoded using one-hot encoding (pd.get_dummies), transforming each category into a binary indicator column. This approach enables the classification models to process categorical information appropriately without imposing an artificial numerical order among categories.

The dataset was split into training (80%) and testing (20%) subsets to ensure evaluation on unseen data and to assess model generalization.

In addition, stratified cross-validation ($k = 5$) was applied to maintain class proportions. The value $k = 5$ was selected because it provides an adequate balance between metric stability and computational cost, given the dataset size. Temporal validation by month was also performed by training on earlier periods and evaluating on subsequent ones to verify the model's robustness with respect to complaint seasonality.

2.5. Model Design and Training

To address the automated prioritization of electrical complaints, two supervised models were implemented: Decision tree and Random forest. Both algorithms classify complaints into two priority categories ("high-priority" or "normal"), using business unit, month, complaint type, and incident classification as input variables.

In this way, the system learns to identify patterns associated with high-priority complaints by considering the incident type, its temporal context, and the operational zone in which it occurs.

The Decision Tree model was selected for its interpretability, which allows supervisors to understand how input variables influence the final decision and to justify complaint prioritization.

This model constructs decision rules as a hierarchy of nodes, branches, and leaves: each node represents a condition on a variable, and each leaf corresponds to the assigned class. The configuration used random_state=42 to ensure reproducibility, no depth restriction (max_depth=None) to allow the tree to capture all relationships present in the data, and the Gini impurity criterion for split evaluation. The model was trained on the training set (X_train, y_train) and evaluated on the test set (X_test).

To improve robustness and generalization, a Random Forest model was implemented as an ensemble of 100 decision trees (n_estimators=100), with predictions combined through majority voting. Bootstrap=True was used to enable sampling with replacement, and the Gini index was retained as the impurity measure for each tree. This approach reduces the risk of overfitting and accommodates the variability and correlation inherent in complaint data. As with the Decision Tree model, the Random Forest was trained on (X_train, y_train) and evaluated on (X_test).

2.6. Model Evaluation and Validation

The models were assessed using standard classification metrics to evaluate their performance in predicting high-priority and normal complaints. Accuracy represents the overall proportion of correctly classified complaints. Precision indicates the share of complaints predicted as high-priority that are truly high-priority. Recall measures the proportion of high-priority complaints correctly identified among all high-priority cases in the dataset. The F1-score, defined as the harmonic mean of precision and recall, provides a balanced measure that is particularly informative when

the high-priority class is less frequent. Together with the confusion matrix, these metrics enable a reliable evaluation of each model's ability to prioritize complaints effectively.

For a more detailed assessment, traditional metrics were complemented with advanced tools. The PR-AUC (precision-recall area under the curve) is particularly informative in imbalanced-class settings because it reflects the model's ability to correctly identify priority complaints. Decision-curve analysis enables visualization of the operational benefit associated with different decision thresholds, supporting the selection of an optimal prioritization rule. Calibration analysis verifies that predicted probabilities correspond to the observed likelihood of a priority complaint. Finally, confidence intervals provide statistical robustness to the reported metrics and reduce the risk of overestimating model performance.

The combination of traditional and complementary metrics ensures a comprehensive and reliable evaluation, supporting both the selection of the best-performing model and its practical implementation in electrical complaint management.

3. Results and Discussion

This section presents the results of applying supervised classification models to institutional electrical complaints, distinguishing between the normal and high-priority categories. Model performance was assessed using standard metrics, including precision, recall, F1-score, and overall accuracy, which enabled evaluation of each model's ability to correctly identify cases requiring urgent attention.

The performance of the Decision Tree model is summarized in Table 4, indicating satisfactory classification of complaints. The model achieved an overall accuracy of 87%, with balanced precision, recall, and F1-score values. For the normal class, precision, recall, and F1-score were 0.88, 0.90, and 0.89, respectively. For the high-priority class, the corresponding values were 0.85, 0.83, and 0.84.

Table 4. Classification report for the Decision Tree model

Normal	0.88	0.90	0.89
High-priority	0.85	0.83	0.84
Accuracy	0.87	0.87	0.87
Macro avg	0.87	0.86	0.86
	Precision	**Recall**	**F1-score**

These results indicate that the model reliably identifies the most urgent complaints while maintaining an appropriate balance between correctly classifying high-priority cases and minimizing false positives.

Figure 2 shows the confusion matrix for the Decision Tree model trained on the cleaned electrical complaint dataset. The model correctly classified 74,961 normal complaints and 49,666 high-priority complaints, demonstrating a solid ability to distinguish between the two categories. However, some misclassifications occurred: 8,329 normal complaints were incorrectly labeled as high-priority, while 10,157 high-priority complaints were classified as normal. These values correspond to the model's false positives and false negatives, respectively.

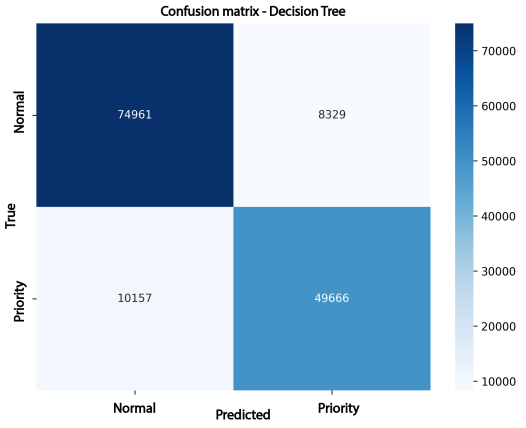


Figure 2. Confusion matrix for the Decision Tree model

These results confirm solid overall performance (87% accuracy) with balanced behavior across both classes. A slight tendency to underestimate some truly high-priority complaints persists, which is expected in scenarios where the priority class is underrepresented.

The Random Forest model outperformed the Decision Tree, confirming its stronger ability to capture nonlinear patterns and mitigate overfitting. Table 5 summarizes the results, showing an overall accuracy of 91% with balanced performance across both classes. The normal category achieved a precision of 0.92, a recall of 0.93, and an F1-score of 0.92, while the high-priority category recorded a precision of 0.90, a recall of 0.89, and an F1-score of 0.89.

Table 5. Classification report for the Random Forest model

Normal	0.92	0.94	0.93
High-priority	0.90	0.89	0.89
Accuracy	0.91	0.91	0.91
Macro avg	0.91	0.91	0.91
	Precision	**Recall**	**F1-score**

These results show that the model achieves a better balance between sensitivity and precision, reducing the classification errors observed in the Decision Tree. In practical terms, the Random Forest improves the identification of high-priority complaints while maintaining a low false-positive rate, making it a more reliable tool for supportingoperational prioritization.

Figure 3 shows the confusion matrix for the Random Forest model. This model exhibits more balanced and accurate performance, correctly classifying 77,495 normal complaints and 53,242 high-priority complaints.

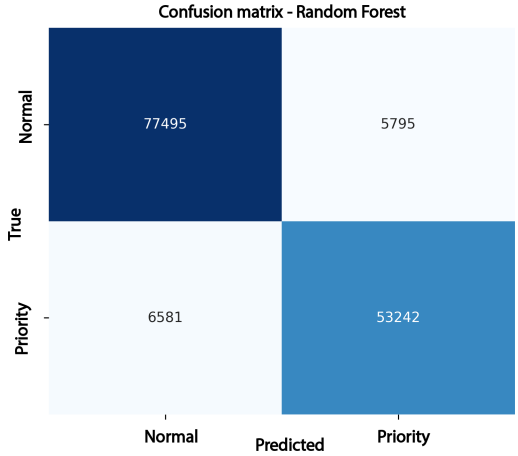


Figure 3. Confusion matrix for the Random Forest model

Classification errors were relatively small: 5,795 normal complaints were incorrectly identified as high-priority, while 6,581 high-priority complaints were classified as normal. This behavior confirms the model's consistency and strong generalization capability, as it substantially reduces critical errors compared to the Decision Tree. Consequently, the Random Forest emerges as one of the most reliable alternatives for automated prioritization of electrical complaints.

To broaden the comparison and validate the robustness of the results, two additional benchmark models, Logistic Regression and XGBoost, were trained using the same dataset and validation scheme. The results are summarized in Table 6, showing that Random Forest consistently achieved the best performance, followed by the Decision Tree. XGBoost and Logistic Regression produced acceptable results, but with lower recall in detecting priority complaints.

Table 6. Comparison of performance metrics across supervised models

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.84	0.81	0.82	0.83
XGBoost	0.86	0.83	0.84	0.85
Decision Tree	0.88	0.86	0.87	0.87
Random forest	0.92	0.91	0.91	0.91

The Random Forest model achieved 91% accuracy and an average F1-score of 0.91, making it the most robust and stable alternative. The Decision Tree showed slightly lower performance, with 87% accuracy and an F1-score of 0.87, while still providing solid classification capability but with greater variability between classes. XGBoost achieved 85% accuracy and an F1-score of 0.84, exhibiting a mild tendency toward overfitting in the minority class. Finally, Logistic Regression, although simpler, delivered reasonable performance, achieving 83% accuracy and an F1-score of 0.82, confirming its value as a baseline but revealing limitations in capturing nonlinear relationships among variables.

These results confirm that ensemble-based models, particularly Random Forest, achieve stronger generalization by reducing errors in identifying high-priority complaints. By contrast, individual or linear models exhibit lower performance when confronted with complex or nonlinear patterns such as those present in institutional electrical claims.

To evaluate the discriminative capacity of the classification models in greater detail, receiver operating characteristic (ROC) and precision–recall (PR) curves were generated for the four analyzed algorithms: Decision tree, Random forest, Logistic Regression, and XGBoost.

Figure 4a shows that the Random Forest model achieved the highest AUC (0.89), followed by the Decision Tree (0.85), XGBoost (0.82), and Logistic Regression (0.80). These values indicate that Random Forest has the strongest ability to discriminate between high-priority and normal complaints.

Figure 4b shows the precision–recall curves, which are particularly relevant given the dataset's class imbalance. Random Forest again stands out achieving a PR-AUC of 0.88, while the Decision Tree, XGBoost, and Logistic Regression reached 0.84, 0.81, and 0.78, respectively. This result confirms that Random Forest maintains higher precision and recall even when the prevalence of the positive class is low.

Regarding the reliability of predicted probabilities, the calibration plot in Figure 4c shows that Random Forest and XGBoost align more closely with the ideal diagonal, indicating probability estimates that

are more consistent with observed frequencies. In contrast, the Decision Tree and Logistic Regression exhibit slight deviations, suggesting poorer calibration at certain thresholds.

The decision-curve analysis in Figure 4d confirms that Random Forest provides the highest net benefit across a wide range of thresholds, making it the most advantageous model for automated prioritization of electrical complaints from an operational standpoint. This type of curve is included because it evaluates performance while accounting for the real costs and benefits associated with each decision. Unlike purely statistical metrics, decision-curve analysis quantifies the model's practical utility by incorporating the consequences of correct and incorrect classifications. This is essential in operational settings, where misclassification may lead to service delays or inefficient resource use.

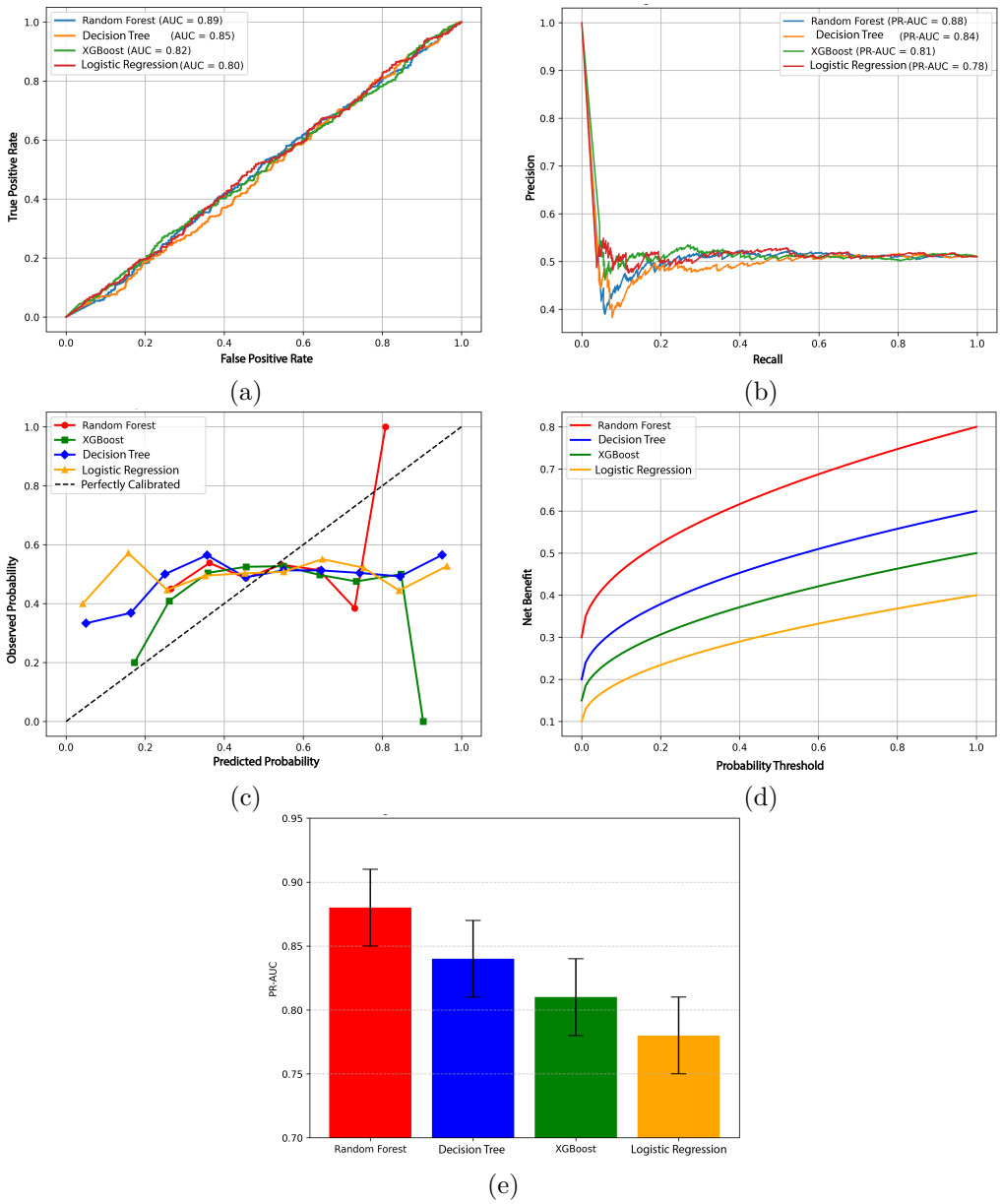


Figure 4. a) ROC curves of the classification models. b) Precision–recall curves for the models. c) Calibration curves for the models. d) Decision-curve analysis. e) 95% confidence intervals for PR-AUC

Additionally, 95% confidence intervals for the PR-AUC were estimated using bootstrap resampling (1000 iterations). Random Forest showed a narrow, high interval (0.85–0.91), followed by the Decision Tree (0.81–0.87), XGBoost (0.78–0.84), and Logistic Regression (0.75–0.81). Confidence intervals quantify performance variability and assess each model's stability under potential dataset fluctuations. Whereas a single PR-AUC value reflects average performance, the intervals indicate how consistently a model performs under perturbations, which is essential when selecting reliable methods for real-world deployment. Figure 4e summarizes these intervals, highlighting the superior statistical stability of Random Forest.

Taken together, Figures 4a–4e show that Random Forest provides the best overall performance, combining strong discrimination, good calibration, high operational benefit, and robust statistical stability.

The Decision Tree remains an acceptable lower-complexity alternative, whereas XGBoost and Logistic Regression show competitive performance but weaker generalization in this context.

Unlike previous studies such as Hosseinzadeh et al. (2021), which validated their fault prediction model using simulated data, and Alhanaf et al. (2021), which relied on smaller-scale experimental samples, the present work is distinguished by the use of a real institutional database from CNEL EP. This dataset comprises more than 143,000 electrical complaints recorded in the province of Santo Domingo. Using real operational records provides a level of realism and immediate applicability that is not always present in international research, as it directly reflects the nature of the complaints submitted by users to the electric utility company.

Furthermore, the methodological approach proposed here is not limited to evaluating a single algorithm; instead, it systematically compares the performance of Decision Tree and Random Forest models using precision, recall, F1-score, and AUC-ROC metrics, thereby strengthening the validity of the results. Accordingly, the primary contribution of this study is to demonstrate how machine-learning techniques can be practically integrated into the complaint-management processes of an electricity distribution company in Ecuador. This work overcomes the limitations of studies that rely solely on synthetic data or controlled settings and provides findings that are directly applicable to daily operations.

The results show that the Random Forest model consistently outperforms the Decision Tree in automated prioritization of electrical complaints, achieving an overall accuracy of 91% and an AUC-ROC of 0.89, compared with 87% accuracy and an AUC-ROC of 0.85 for the Decision Tree. This improvement indicates a stronger ability to discriminate between high-priority and normal cases [1], consistent with the robustness and generalization capacity that characterize ensemble models [22], [24, 25].

While Hussain et al. [24] applied ensemble models to energy prediction in electric vehicles, the present work uses Random Forest to prioritize electrical complaints at CNEL EP, demonstrating that comparable robustness and generalization capacity can enhance operational management and resource allocation in institutional settings.

Similarly, Jing et al. [23] focused on predicting and optimizing electric meter maintenance through a hybrid LSTM-XGBoost approach. By contrast, this study applies Random Forest to the prioritization of institutional electrical complaints, incorporating derived variables such as urgency and recurrence by zone and month. These results illustrate how supervised approaches can optimize resource allocation and improve service delivery to end users within CNEL EP.

Unlike the study by Sun et al. [22], which focused on improving the internal architecture of Random Forest to reduce correlation and increase accuracy, the present work emphasizes the practical integration of the model into the institutional management of electrical complaints. Incorporating derived variables enables more effective operational prioritization, thereby complementing methodological advances in algorithm refinement with a concrete application to the Ecuadorian electricity sector.

The analysis of the confusion matrices indicates that Random Forest substantially reduces critical errors by minimizing the misclassification of high-priority complaints as normal. This improvement has direct operational implications, enabling shorter response times, more efficient scheduling of technical crews, and enhanced user perception of service quality [1], [7]. Taken together, these findings validate the applicability of ensemble models for prioritizing electrical complaints in real institutional settings and support their integration into operational management processes.

From a comprehensive perspective, analyses based on confusion matrices, ROC and precision–recall curves, calibration plots, decision-curve analysis, and confidence intervals indicate that the superior performance of Random Forest is consistent and not driven by model-specific adjustments. This combination of strong discrimination, statistical stability, and operational utility is critical for managing electrical complaints, where misclassifications can cause service delays, inefficient crew allocation, and reduced perceived service quality. Accordingly, Random Forest emerges as the most suitable model for automated prioritization at CNEL EP, offering a robust balance between technical performance and operational relevance.

Although XGBoost and Logistic Regression exhibit acceptable performance, limitations in calibration and stability constrain their direct applicability. The Decision Tree, despite its simplicity and interpretability,

also shows reduced effectiveness under high variability and class-imbalance conditions. Taken together, these findings confirm that ensemble approaches such as Random Forest provide a robust and reliable alternative for optimizing electrical complaint management in the Ecuadorian electricity sector.

4. Conclusiones

This study demonstrates the viability and effectiveness of applying supervised learning models for the automated prioritization of electrical complaints within the institutional context of CNEL EP. The results show that both Decision Tree and Random Forest can accurately classify complaints by priority level. However, Random Forest achieved superior performance, reaching 91% overall accuracy and an AUC-ROC of 0.89. This superiority is reflected in reduced critical errors and stronger discrimination between high-priority and normal complaints, which is essential for improving operational efficiency and enhancing user satisfaction.

Incorporating derived variables such as urgency, recurrence by zone and month, along with careful preprocessing and categorical-variable encoding, enhanced the models' predictive capacity and ensured robust performance under the variability of real-world data. Complementary analyses using confusion matrices, ROC and precision–recall curves, calibration plots, and decision-curve analysis confirmed that the superior performance of Random Forest is consistent and not driven by model-specific adjustments, demonstrating its statistical stability and operational utility.

From an operational perspective, implementing this system provides a reliable and scalable tool for intelligent allocation of technical resources and improved customer service, contributing to more efficient and proactive management of electrical services.

As future work, it is recommended to expand the dataset by incorporating additional contextual variables such as weather conditions, incidence by user type, and infrastructure-related events. It is also advisable to explore more advanced or hybrid machine-learning algorithms and to evaluate real-time deployment to further enhance operational efficiency. These actions will consolidate the model's applicability, strengthen institutional decision-making, and broaden its impact on optimizing electrical service management in Ecuador.

For operational integration at CNEL EP, the model can be embedded in an interactive control dashboard that receives complaints in real time, automatically preprocesses the data, and classifies each case as normal or high-priority using the Random Forest model. The outputs would be displayed on a dashboard accessible to supervisors, enabling more efficient assignment of technical crews and real-time monitoring of performance indicators. This workflow ensures that critical complaints are addressed promptly, optimizing both human-resource utilization and response times.

Contributor Role

- **Francisco Javier Carpio Velasco:** conceptualization, methodology, formal analysis, software, visualization, writing – original draft and validation.
- **Gloria Margarita Garcés Beltrán:** supervision, formal analysis and writing – review & editing.

References

[1] R. Walters, E. J. Jaselskis, and J. M. Kurtenbach, "Classification of Knowledge within the Electrical Contracting Industry: A Case Study," *Leadership and Management in Engineering*, vol. 7, no. 1, pp. 11–17, 2007. [Online]. Available: https://doi.org/10.1061/(ASCE)1532-6748(2007)7:1(11)

[2] P. Gaiardelli, B. Resta, V. Martínez, R. Pinto, and P. Albores, "A classification model for product-service offerings," *Journal of Cleaner Production*, vol. 66, pp. 507–519, Mar. 2014. [Online]. Available: https://doi.org/10.1016/j.jclepro.2013.11.032

[3] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Emergent electricity customer classification," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 152, no. 2, p. 164, 2005. [Online]. Available: https://upsalesiana.ec/ing35ar6r3

[4] ——, "Load Pattern-Based Classification of Electricity Customers," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232–1239, May 2004. [Online]. Available: https://doi.org/10.1109/TPWRS.2004.826810

[5] G. Wacker and R. Billinton, "Customer cost of electric service interruptions," *Proceedings of the IEEE*, vol. 77, no. 6, pp. 919–930, Jun. 1989. [Online]. Available: https://doi.org/10.1109/5.29332

[6] C. Beckel, L. Sadamori, and S. Santini, "Towards automatic classification of private households using electricity consumption data," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ser. SenSys '12. ACM, Nov. 2012, pp. 169–176. [Online]. Available: https://doi.org/10.1145/2422531.2422562

[7] H. M. Ahmed, "The effect of customer complaint handling practices on customer satisfaction

in ethiopian electric utility customer service center," *International Journal of Management (IJM)*, vol. 15, no. 1, 2024. [Online]. Available: https://upsalesiana.ec/ing35ar6r7

[8] G. F. Calderón Intriago, J. C. Mera Macías, and J. C. Guamán Segarra, "El mercado eléctrico ecuatoriano y su incidencia en los altos costos de las tarifas de energía eléctrica a los consumidores residenciales y comerciales en la provincia de manabí comprendidos en el periodo de enero a diciembre del 2018." *Revista de Investigaciones en Energía, Medio Ambiente y Tecnología: RIEMAT*, vol. 3, no. 2, p. 24, Dec. 2018. [Online]. Available: https://doi.org/10.33936/riemat.v3i2.1627

[9] A. K. Montesdeoca Rivas and M. Cedillo Fajardo, "Análisis de la calidad de servicio y la satisfacción del usuario externo en la empresa eléctrica pública estratégica corporación nacional de electricidad CNEL EP, Guayaquil–2020," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, no. 3, pp. 9840–9858, Jul. 2023. [Online]. Available: https://doi.org/10.37811/cl_rcm.v7i3.7009

[10] C. A. Vásconez-Mejía, J. F. Molina-Mora, and M. E. Gaspar-Santos, "Derechos de los consumidores del servicio eléctrico en tiempos de pandemia en Ecuador," *CIENCIAMATRIA*, vol. 8, no. 2, pp. 80–91, Mar. 2022. [Online]. Available: https://doi.org/10.35381/cm.v8i2.698

[11] M. Neira and R. Quintanilla, "La calidad del servicio técnico de electricidad," *Revista Técnica "Energía"*, vol. 8, no. 1, Jan. 2012. [Online]. Available: https://doi.org/10.37116/revistaenergia.v8.n1.2012.189

[12] C. C. Orozco Domínguez, A. F. Sánchez Hernández, and M. J. Santiago Camacho, "Deficiencias en la prestación del servicio de energía eléctrica en la ciudad de Cartagena," *Vis Iuris*, pp. 55–74, Nov. 2017. [Online]. Available: https://doi.org/10.22518/vis.v0i00.1174

[13] G. E. Chodzaza and H. S. Gombachika, "Service quality, customer satisfaction and loyalty among industrial customers of a public electricity utility in Malawi," *International Journal of Energy Sector Management*, vol. 7, no. 2, pp. 269–282, Jun. 2013. [Online]. Available: https://doi.org/10.1108/IJESM-02-2013-0003

[14] S. Abideen, O. Joseph, J. Surajudeen, and A. Adeyinka, "Service quality and customer satisfaction of electricity distribution company: A study of Olumo Business Hub, Abeokuta, Ogun State," *SSRN Electronic Journal*, 2018. [Online]. Available: https://doi.org/10.2139/ssrn.3391404

[15] W. Li, W. W. Kwabla Pomegbe, C. S. Kofi Dogbe, and J. D. Novixoxo, "Employees' customer orientation and customer satisfaction in the public utility sector: The mediating role of service quality," *African Journal of Economic and Management Studies*, vol. 10, no. 4, pp. 408–423, Dec. 2019. [Online]. Available: https://doi.org/10.1108/AJEMS-10-2018-0314

[16] D. Sarathkumar, M. Srinivasan, A. A. Stonier, R. Samikannu, N. R. Dasari, and R. A. Raj, "A technical review on classification of various faults in smart grid systems," *IOP Conference Series: Materials Science and Engineering*, vol. 1055, no. 1, p. 012152, Feb. 2021. [Online]. Available: https://doi.org/10.1088/1757-899X/1055/1/012152

[17] A. S. Alhanaf, H. H. Balik, and M. Farsadi, "Intelligent fault detection and classification schemes for smart grids based on deep neural networks," *Energies*, vol. 16, no. 22, p. 7680, Nov. 2023. [Online]. Available: https://doi.org/10.3390/en16227680

[18] Z. Krivohlava, S. Chren, and B. Rossi, "Failure and fault classification for smart grids," *Energy Informatics*, vol. 5, no. 1, Oct. 2022. [Online]. Available: https://doi.org/10.1186/s42162-022-00218-3

[19] J. Hosseinzadeh, F. Masoodzadeh, and E. Roshandel, "Fault detection and classification in smart grids using augmented K-NN algorithm," *SN Applied Sciences*, vol. 1, no. 12, Nov. 2019. [Online]. Available: https://doi.org/10.1007/s42452-019-1672-0

[20] M. Mohammadian, F. Aminifar, N. Amjady, and M. Shahidehpour, "Data-driven classifier for extreme outage prediction based on Bayes decision theory," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 4906–4914, Nov. 2021. [Online]. Available: https://doi.org/10.1109/TPWRS.2021.3086031

[21] S. Rana, "AI-driven fault detection and predictive maintenance in electrical power systems: A systematic review of data-driven approaches, digital twins, and self-healing grids," *American Journal of Advanced Technology and Engineering Solutions*, vol. 1, no. 01, pp. 258–289, Feb. 2025. [Online]. Available: https://doi.org/10.63125/4p25x993

[22] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Systems with Applications*, vol. 237, p. 121549, Mar. 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.121549

[23] H. Jing, H. Cao, and S. Sun, "Multi-objective predictive maintenance optimization of electric energy meters based on LSTM-XGBoost and modified firefly algorithm," *International Journal of Emerging Electric Power Systems*, Aug. 2025. [Online]. Available: https://doi.org/10.1515/ijeeps-2025-0184

[24] I. Hussain, K. B. Ching, C. Uttraphan, K. G. Tay, and A. Noor, "Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study," *Scientific Reports*, vol. 15, no. 1, May 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-94946-7

[25] Y. Khalili, M. Ahmadi, and M. K. Moraveji, "Time-aware predictive maintenance of electrical submersible pumps using catboost ensemble learning and trend-based labeling," *Journal of Petroleum Exploration and Production Technology*, vol. 15, no. 9, Aug. 2025. [Online]. Available: https://doi.org/10.1007/s13202-025-02070-z

[26] T. Ahmad, H. Chen, R. Huang, G. Yabin, J. Wang, J. Shair, H. M. Azeem Akram, S. A. Hassnain Mohsan, and M. Kazim, "Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment," *Energy*, vol. 158, pp. 17–32, Sep. 2018. [Online]. Available: https://doi.org/10.1016/j.energy.2018.05.169

[27] T. Zhang, R. Gao, and S. Sun, "Theories, applications and trends of non-technical losses in power utilities using machine learning," in *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE, May 2018, pp. 2324–2329. [Online]. Available: https://doi.org/10.1109/imcec.2018.8469410

[28] J. Ausmus, R. S. de Carvalho, A. Chen, Y. N. Velaga, and Y. Zhang, "Big data analytics and the electric utility industry," in *2019 International Conference on Smart Grid Synchronized Measurements and Analytics (SGSMA)*. IEEE, May 2019. [Online]. Available: https://doi.org/10.1109/SGSMA.2019.8784657

[29] M. J. Mousavi, J. Stoupis, and K. Saarinen, "Event zone identification in electric utility systems using statistical machine learning," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, Apr. 2018, pp. 1–9. [Online]. Available: https://doi.org/10.1109/TDC.2018.8440570

[30] Y. Shen, M. Abubakar, H. Liu, and F. Hussain, "Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems," *Energies*, vol. 12, no. 7, p. 1280, Apr. 2019. [Online]. Available: https://doi.org/10.3390/en12071280

[31] L. Xu and M.-Y. Chow, "A classification approach for power distribution systems fault cause identification," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 53–60, Feb. 2006. [Online]. Available: https://doi.org/10.1109/TPWRS.2005.861981

[32] B. Kumwenda, E. Zulu, and M. Ndiaye, "System disturbance classification model for a low voltage distribution network," in *2024 IEEE PES/IAS PowerAfrica*. IEEE, Oct. 2024, pp. 01–05. [Online]. Available: https://doi.org/10.1109/PowerAfrica61624.2024.10759455

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.07
pISSN: 1390-650X / eISSN: 1390-860X

Estimation of emissions from failures in Otto engines using convolutional neuronal networks

Estimación de emisiones por fallos en motores Otto mediante redes neuronales convolucionales

Elmer I. Arias-Montaño[1], Rogelio S. León-Japa[1,*], Pedro García-Jaramillo[1], José Maldonado-Ortega[2]

Received: 24-05-2024, Received after review: 05-11-2025, Accepted: 20-11-2025, Published: 01-01-2026

Abstract

This study applies a machine learning technique, specifically Convolutional Neural Networks (CNNs), to predict pollutant emissions resulting from failures in actuators and components of Otto engines. The work addresses the current lack of non-intrusive methods that exploit signals already available in the vehicle to estimate, with high accuracy, emissions associated with failures in the injection, ignition, and air intake systems. Concentrations of CO (% carbon monoxide), CO_2 (% carbon dioxide), HC (unburned hydrocarbons, ppm), and O_2 (% oxygen) are quantified by analyzing the Manifold Absolute Pressure (MAP) sensor signal under a rigorous sampling and signal-processing protocol. Optimal features are extracted from the MAP signal based on their informational relevance and discriminative capacity. These features are obtained through spectrographic transformation, enabling the construction of a robust database. The resulting dataset serves as an effective input for CNN training, achieving emission prediction errors below 1%.

Keywords: Convolutional neural networks (CNN), emissions, estimation, Otto engines, machine learning.

Resumen

En este estudio se implementa una técnica de aprendizaje automático, concretamente redes neuronales convolucionales (CNN, por sus siglas en inglés), con el objetivo de predecir las emisiones contaminantes producidas por fallos en actuadores y componentes de motores Otto. Así pues, el problema de investigación abordado en este trabajo es la ausencia de métodos no intrusivos que, a partir de señales ya disponibles en el vehículo, permitan estimar con alta precisión las emisiones asociadas a fallas en sistemas de inyección, encendido y admisión de aire. Se cuantifican los niveles de CO (monóxido de carbono, %), CO_2 (dióxido de carbono, %), HC (hidrocarburos no quemados en ppm) y O_2 (oxígeno, %) a partir del análisis de la señal proveniente del sensor MAP (Manifold Absolute Pressure). Para ello, se adopta un protocolo riguroso de muestreo y procesamiento de la señal. La extracción de características óptimas del sensor MAP se basa en su relevancia informativa y capacidad de discriminación, determinadas mediante la transformación espectrográfica de la señal, lo que permite construir una base de datos robusta. Esta base sirve como entrada eficaz para el entrenamiento de la CNN, con la cual se logra un error de predicción inferior al 1 %.

Palabras clave: CNN, emisiones, estimación, motor Otto, redes neuronales convolucionales, machine learning.

[1,*]Grupo de Investigación en Movilidad, Vehículos y Transporte (eX- MoVeT), Carrera de Ingeniería Automotriz, Universidad Nacional de Loja, Loja-Ecuador. Corresponding author ✉: rleonj@outlook.es.
[2] Grupo de Investigación en Movilidad Inteligente (GMovInt), Carrera de Ingeniería Automotriz, Universidad Politécnica Salesiana, Guayaquil-Ecuador

Suggested citation: E. I. Arias-Montaño, R. S. León-Japa, P. García-Jaramillo and J. Maldonado-Ortega "Estimation of emissions from failures in Otto engines using convolutional neuronal networks," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 90-102, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.07.

1. Introduction

Currently, emissions from alternative internal combustion engines (MCIA) are among the most significant sources of atmospheric pollution in urban areas worldwide. These emissions degrade air quality and pose an environmental challenge of increasing concern.

In contrast, in spark-ignition engines (MEP) the air–fuel mixture is not always completely burned during the ignition phase due to inefficiencies arising from variations in temperature and load, changes in engine speed (rpm), and fluctuations in internal pressure [1]. Therefore, to evaluate combustion efficiency, it is necessary to quantify the main emitted pollutants, namely carbon monoxide (CO), carbon dioxide (CO_2), unburned hydrocarbons (HC), nitrogen oxides (NO_x), and oxygen (O_2) [2,3].

Accordingly, it becomes essential to adopt advanced methodologies that integrate computational mathematics with artificial intelligence tools [4,5]. Such approaches enable more precise detection of faults in the components and actuators of alternative internal combustion engines (MCIA), as well as a more detailed characterization of exhaust gas emissions [6].

Sapio et al. [7] developed a hybrid modeling approach for selective catalytic reduction (SCR) systems. Specifically, they employed recurrent neural networks (RNNs) and subsequently used feedforward neural networks (FFNNs) to predict both the outlet temperature of the catalytic system (SC) and pollutant concentrations, particularly NO_x and NH_3 [8].

The use of advanced supervised learning techniques, including artificial neural networks (ANNs), bidirectional convolutional networks (CNN-BiLSTM), support vector machines (SVMs), and the extreme gradient boosting algorithm (XGBoost), has proven to be an effective alternative for improving odor assessment in vehicle interiors [9].

Cesur and Uysal [10] demonstrated that using methanol–gasoline fuel blends in spark-ignition engines (MEP) increases effective power by 3.7 % while notably reducing NO_x and HC levels. In addition, their model enabled the prediction of engine performance parameters and emissions with accuracies of 99 % and 98 %, respectively.

To estimate CO_2 emissions, studies have employed various algorithms, including linear regression, decision trees, and neural networks. Collectively, their findings confirm the high effectiveness of machine learning approaches for accurate prediction in vehicular emissions [11].

Li et al. [12] developed an improved model for misfire detection in gasoline engines based on YOLOv8, optimized with BiFormer and CBAM modules to enhance feature extraction from acoustic signals. Engine signals were transformed via wavelet analysis into time–frequency images for network training, achieving an accuracy of 99.71% and thereby outperforming the original YOLOv8 approach [13–15].

The present research focuses on analyzing intake manifold pressure using the signal recorded by the MAP sensor in spark-ignition engines (MEP). The study addresses the need for a rapid, low-cost, and minimally invasive method to diagnose faults in gasoline engines and estimate their exhaust emissions. In contrast, current approaches often rely on complex physical models or test benches, which increase both cost and measurement time [16].

Accordingly, the problem can be summarized as determining how to leverage the information contained in a single MEP sensor to quantitatively estimate multiple exhaust emissions under different incipient fault conditions using deep learning techniques [17,18]. In this sense, the proposed methodology represents a relevant advancement for both exhaust-emission prediction and fault detection in MEP engines.

The article is organized as follows: Section 2 describes the experimental configuration, the data acquisition procedure, the construction of the spectrograms, and the CNN design. Section 3 presents the emission-prediction results and the comparative statistical analysis. Finally, Section 4 summarizes the main conclusions and outlines future work.

2. Materials and methods

The following section presents the proposed advanced diagnostic approach, along with the implemented experimental configuration and instrumentation, which are minimally invasive in nature. This section also details the sampling conditions, the data acquisition methodology, the processing of the MAP sensor signal, and the procedure for selecting the spectrograms used to train the convolutional neural network (CNN). In addition, it describes the development of the algorithm designed for emission prediction and fault detection in the system.

2.1. Experimental configuration and instrumentation

The proposed approach aims to avoid disassembling vehicle components and systems by enabling fault detection and exhaust-emission prediction through a minimally invasive method. To achieve this objective, samples of the vacuum generated in the spark-ignition engine (MEP) are acquired using the voltage signal produced by the manifold absolute pressure (MAP) sensor as the information source.

Table 1 summarizes the technical characteristics of the MEP engine used in the study, whereas Table 2 presents the instrumentation employed during the experimental process.

Table 1. Technical characteristics of the MEP engine

Characteristic	Value
Model	Volkswagen Jetta 2011
Number of cylinders	4
Number of valves	DOHC/16 valves
Fuel injection system	MPI
Displacement	2000 cc
Rated power	114 HP @ 5200 rpm
Rated torque	170 N.m @ 4000 rpm
Fuel	Gasoline (RON 85)

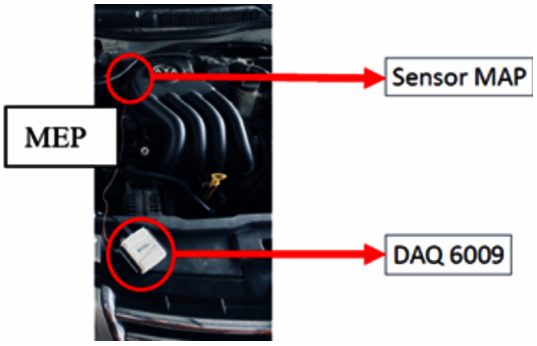


Figure 1. Instrumentation used in the MEP engine

Figure 2. Connection scheme for MAP sensor signal acquisition

Table 2. Instrumentation used in the study

Element	Characteristic
Automotive scanner	FCAR F7S
Connection type	OBDII
Diagnostic mode	Motor – Live Data
Data acquisition card	NI DAQ-6009
Intake manifold pressure	MAP
Pressure range	20-117 KPa
Voltage	3.05V @ 2060 m s. n. m.
Sensor type	Piezoresistive
Gas analyzer	Brain Bee AGS 688
CO measurement range	0.0 9.99 %
Resolution	0.01 %
HC measurement range C	0.0 9999 ppm
Resolution	1 ppm
Measurement range	0.0 25 %
Resolution	0.01 %
Measurement range	0.0 19.9 %
Resolution	0.1 %

Figure 1 illustrates the main components used in this study, including the manifold absolute pressure (MAP) sensor, the DAQ-6009 data acquisition board, and a laptop computer serving as the processing unit. Figure 2 shows the connection scheme established between the DAQ-6009 board, the MAP sensor, and the experimental unit selected for testing, specifically a 2011 Volkswagen Jetta vehicle.

The MAP sensor is an original vehicle component present in most modern Otto engines, making the proposed approach economical and easily replicable without the need to add external hardware.

2.2. Sampling conditions

The MAP sensor installed in the experimental unit was used to capture the voltage signal, due to its strategic location in the intake manifold of the spark-ignition engine (MEP). In parallel, exhaust gas emission samples were collected, including carbon monoxide (CO) and carbon dioxide (CO_2), both expressed percentages, unburned hydrocarbons (HC) expressed in parts per million (ppm), and oxygen (O_2), also expressed as a percentage. These measurements were obtained using a gas analyzer, while the MAP sensor signal was acquired through the DAQ-6009 board using LabVIEW 2024, which enabled real-time monitoring.

During data collection, both the MAP sensor signal and exhaust gas emissions were recorded at a constant engine speed of 2500 rpm under static test conditions in the experimental unit. The MEP operating temperature was maintained between 90 to 98 °C, and an automotive scanner was used to verify and monitor the engine's key parameters in real time.

According to the pre-experimental study by Contreras et al. [19], the MAP sensor signal exhibits high-frequency peaks. Consequently, the authors recommend acquiring data at a rate of 10 kHz over an interval of 5 s to satisfy the Nyquist criterion.

The absolute pressure in the intake manifold is directly related to engine load, volumetric efficiency, and combustion dynamics. Therefore, its variations contain information about faults in the intake, injection, and ignition systems. Previous studies by this research group have shown that, when properly processed, the MAP

signal enables discrimination of operating conditions and certain fault types with good sensitivity [2], [19]. By relying exclusively on the MAP sensor signal, the proposed approach remains minimally invasive, since exhaust ducts are not manipulated, additional intrusive sensors are not required, and overall system complexity is reduced.

In accordance with this recommendation, the DAQ-6009 board was configured with a sampling rate of 5 kHz for the samples to read parameter, an acquisition frequency of 10 kHz, and a duration of 5 s for each experimental treatment, using a differential connection mode. The data-collection order was planned in Minitab through a design of experiments (DOE). Accordingly, eight base treatments were defined, and to increase the statistical robustness and representativeness of the dataset, ten replicates were performed for each treatment [1].

2.3. Data collection procedure

Figure 3 presents the main physical components used to diagnose faults in components and actuators and to predict exhaust gas emissions. These elements include an automotive scanner, the DAQ-6009 data acquisition board, a Brain Bee gas analyzer, and a laptop computer.

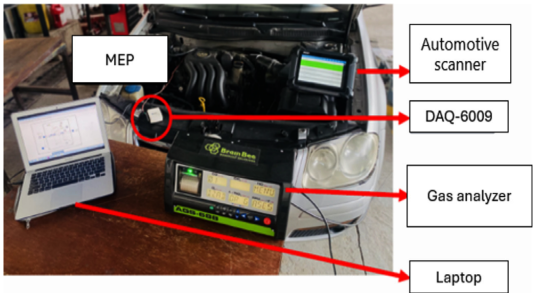


Figure 3. Instruments used to collect the samples

To acquire data from the MAP sensor signal, the systematized procedure shown in the flowchart in Figure 4 is followed.

The sampling methodology begins with the verification of the operating condition of the experimental unit, either under normal conditions or with supervised induced faults. The connection between the NI DAQ-6009 board and the MAP sensor of the MEP engine is then inspected. If the connection is correct, the signal is recorded in LabVIEW and the information is stored in an Excel file; otherwise, the connection is rechecked [1].

This procedure is applied identically to the MEP engine in optimal operating condition and to the engine with controlled faults, as shown in Figures 4(a) and 4(b) [1], [14]. Sample acquisition is repeated ten times for each condition defined in the experimental unit.

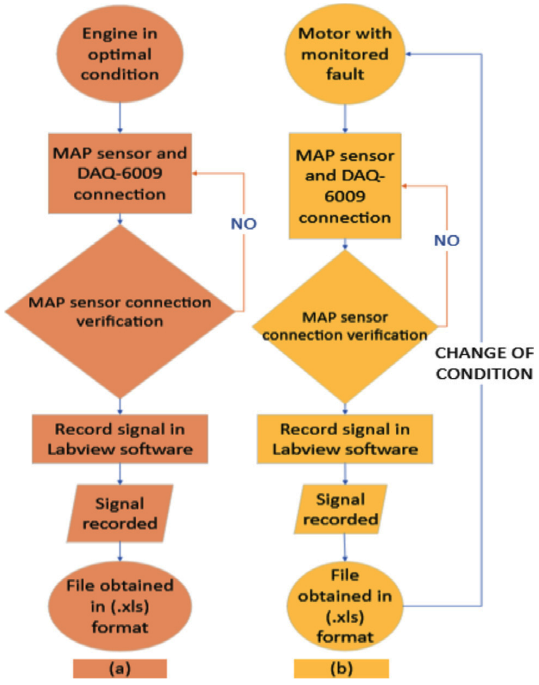


Figure 4. Flowchart for the sampling process: (a) engine under optimal conditions, (b) engine with induced fault

Figure 5 describes the eight operating conditions established for the engine under study, each with its corresponding level, including the optimal operating condition.

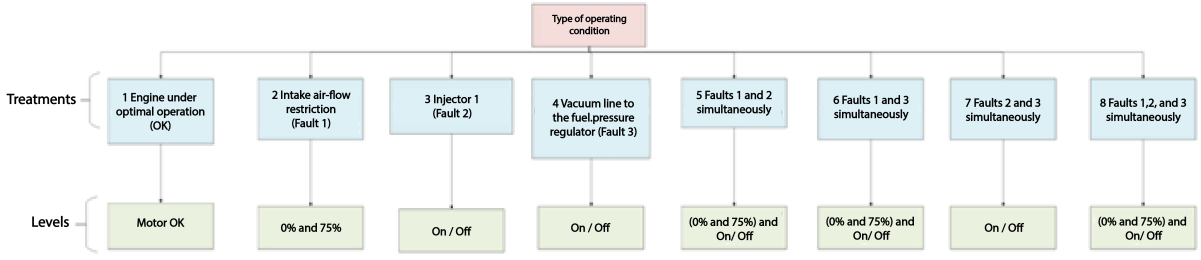


Figure 5. Functional states and operating levels defined for the MEP test unit

2.4. Extraction of MAP signal segments and selection of spectra to train the CNN

Spectrograms provide a visual representation of how a signal's spectral content varies over time and are particularly effective for analyzing non-stationary signals. Accordingly, this study employs spectrograms to characterize the dynamic behavior of the MAP sensor signal.

To select the most suitable spectrograms for training the convolutional neural network (CNN), a script was first developed to segment each experimental treatment, comprising ten replicates, by dividing the MAP sensor signal into 250 fragments. An algorithm was then implemented in the MATLAB environment to automatically generate these representations, yielding 200 different spectrograms per treatment. The configuration used a sampling frequency of 10 kHz, a window size of 512 samples for spectral analysis, and an overlap of 128 samples between successive windows.

To determine the most suitable spectrograms to be used as input for the convolutional neural network (CNN), a detailed analysis of the 200 spectrograms generated for each treatment was performed, as shown in Figure 6. The region of lower-energy activity observed approximately between 3.5 and 4.5 s may indicate a transient condition, that is, an event associated with a disturbance in the MAP signal, as could occur under a fault condition in the MEP engine.

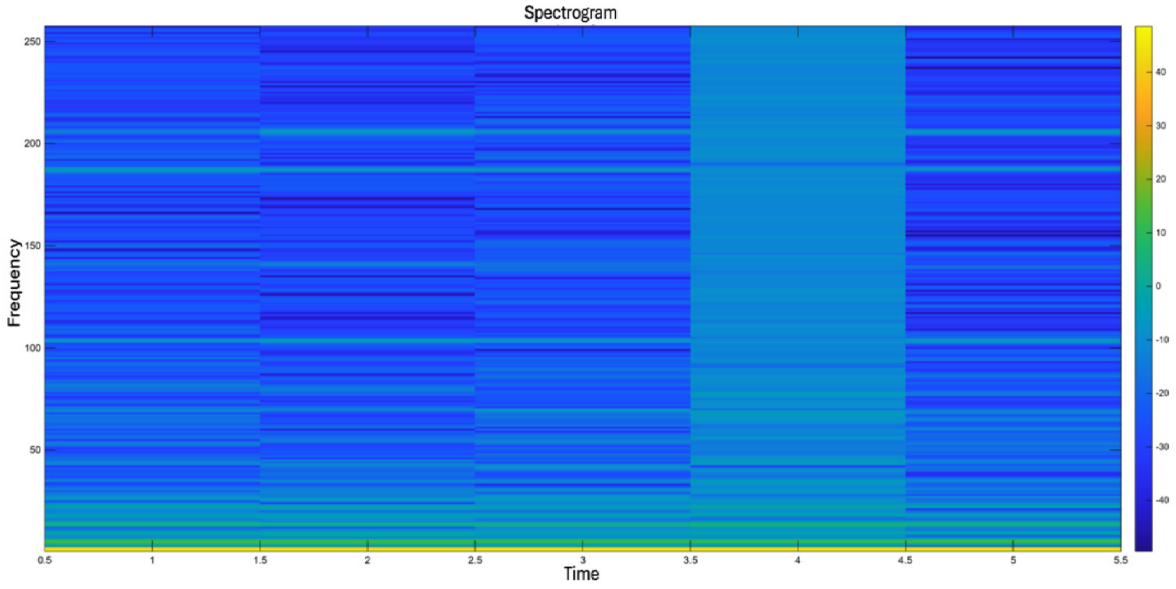


Figure 6. Example spectrogram from the set of 200 generated

From this process, the spectral pattern exhibiting the greatest consistency and representativeness was identified. Accordingly, 50 spectrograms per treatment that most faithfully reproduced this characteristic pattern were selected. Using MATLAB, the 50 spectrograms with the smallest mean distance from the average spectrum were selected as the most representative. This selection is supported by their ability to consistently reproduce the characteristic spectro-temporal pattern of the MAP sensor under each operating condition.

Figure 7 shows the representative spectrogram corresponding to a sample of the engine under OK operating conditions. The horizontal axis represents time in seconds, while the vertical axis shows the distribution of spectral energy as a function of frequency from 0 to 250 Hz. This range encompasses the dominant variation of the intake-pressure pulse in Otto engines.

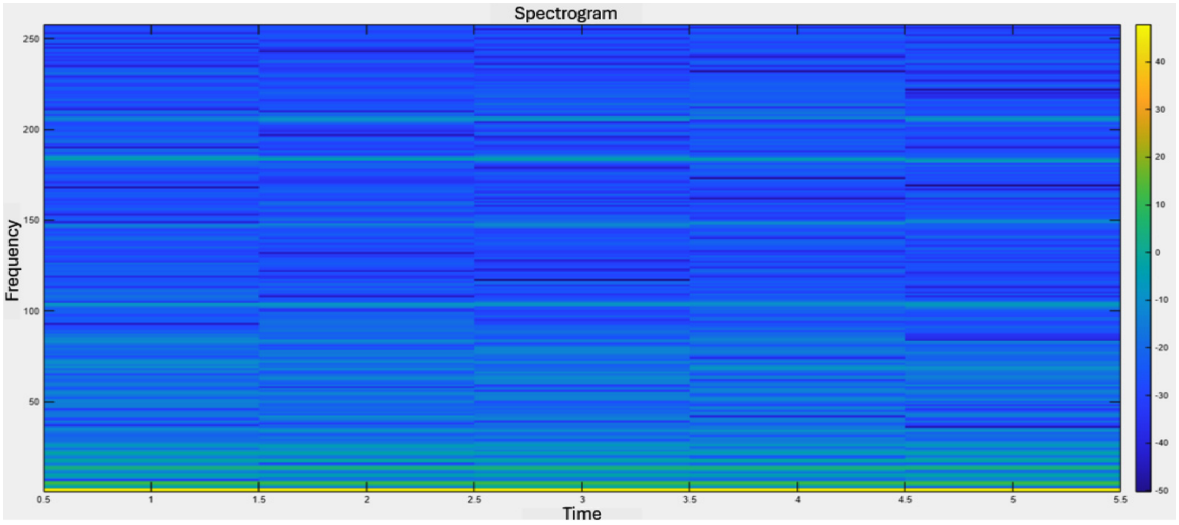


Figure 7. Representative spectrogram of the engine under OK operating conditions

The colorimetric map in Figure 7, expressed on dB scale, reflects the concentration of spectral power within each time–frequency region. Predominant energy is observed at low frequencies, below 40 Hz, associated with manifold-vacuum dynamics and the interaction between piston motion and the intake system. The periodic variations distributed over time indicate the cyclical consistency of the MAP signal under stable operating conditions. Localized variations in the intermediate bands, approximately 80 to 180 Hz, are related to fluctuations in volumetric filling.

2.5. Convolutional neural network (CNN) model

Using Python in the Visual Studio Code development environment, a convolutional neural network (CNN)–based algorithm was developed. This architecture belongs to the field of deep learning, a branch of machine learning characterized by its capacity to learn autonomously from large datasets. Its main strength lies in detecting complex patterns in images, which makes CNNs particularly suitable for classification and recognition tasks.

The use of a CNN is fully justified by the spectral, two-dimensional nature of the data obtained from the MAP sensor. By transforming the time-domain signal into spectrograms, the relevant engine information is encoded in time–frequency energy patterns that exhibit complex structures. Such patterns cannot be captured adequately by traditional models such as SVMs, decision trees, or fully connected neural networks.

CNNs, owing to their architecture based on convolutional filters, enable automatic extraction of hierarchical spatial features from spectrograms, thereby identifying subtle variations associated with pressure dynamics, combustion irregularities, and fault-related effects in the intake, injection, or ignition systems. Because the model input consists of images, CNNs represent the standard approach for processing this type of data. Unlike feed-forward ANNs or models based on SVM or Random Forest, CNNs learn directly from the spectrograms through convolutional filters that capture high-frequency patterns, concentrated energy regions, and temporal structures corresponding to different fault modes.

Figure 8 presents the CNN architecture, in which images are processed through filters applied across successive convolutional layers. Each layer extracts progressively more complex features, enabling a hierarchical representation of the relevant information and facilitating interpretation of the visual content.

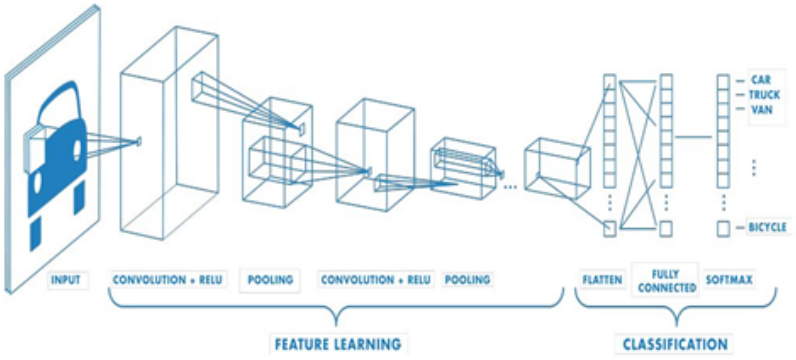


Figure 8. Convolutional neural network (CNN) architecture [20]

The proposed CNN architecture, consisting of three convolutional layers followed by pooling and dense layers, learns complex patterns without overfitting and with reasonable training times. The results yield an MSE below 1%, outperforming traditional ANN models that operate on less processed MAP signals.

Figure 9 presents the flowchart describing the convolutional neural network (CNN) training process, specifically designed to predict exhaust gas emissions and detect potential faults in the evaluated experimental unit.

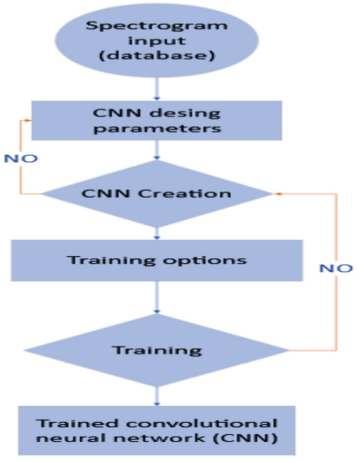


Figure 9. Flowchart of the CNN training process

The algorithm begins by reading of the spectrogram database, after which the images are normalized by dividing their values by 255, thereby scaling them to the range from 0 to 1. Then, the dataset is divided into two subsets: 70 % is allocated for model training, and the remaining 30 % is reserved for testing.

Once these steps are completed, the initial parameters of the convolutional neural network (CNN) are configured, including the input-layer size, defined as (224, 224, 3), which corresponds to the dimensions of the images used. In the following stages, the CNN architecture is constructed with three convolutional layers, each followed by its corresponding normalization layer and dimensionality-reduction layers using MaxPooling2D.

Additionally, the network incorporates fully connected layers and a final regression layer responsible for estimating the exhaust-gas values, as presented in Table 3.

Table 3. Characteristics of the CNN architecture

Characteristics	Type	Number of parameters
Input	Input images	
Conv2D_1	2-D	7168
BatchNormalization	Batch Normalization	1024
MaxPooling2D	Pooling 2-D	0
Conv2D_2	2-D	295 040
BatchNormalization	Batch Normalization	512
MaxPooling2D	Pooling 2-D	0
Conv2D_3	2-D	73 792
BatchNormalization	Batch Normalization	256
MaxPooling2D	Pooling 2-D	0
Flatten	Flatten	0
Full Conected	Dense	1 605 664
Output	Dense	132

Figure 10 illustrates the convolutional neural network developed to predict pollutant emissions in Otto engines using spectrograms derived from the MAP sensor signal. The architecture begins with an input layer that receives 224 × 224 × 3 images, followed by three sequential Conv2D–Batch Normalization–MaxPooling2D blocks that extract and progressively reduce the dimensionality of the time–frequency features contained in the spectrograms.

Subsequently, the network includes a flatten layer that converts the feature maps into a one-dimensional vector, which is then processed by a dense (fully connected) layer and fed to an output layer implementing multivariable regression to generate the predictions. This configuration enables efficient hierarchical extraction of relevant patterns for accurate estimation of CO, CO_2, HC, and O_2 emissions.

For training, the Adam optimizer was used with 50 epochs and a batch size of 16 samples.

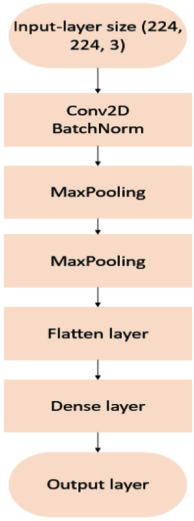


Figure 10. Structure of the proposed CNN

At the end of this stage, model performance was evaluated by calculating the mean squared error (MSE). This evaluation was complemented with graphical analyses, including a scatter plot, an error histogram, and a point-by-point comparison plot, which enabled a more detailed assessment of the model's accuracy.

Subsequently, the prediction error was estimated for each output corresponding to CO, HC, CO_2, and O_2 emissions. When the error exceeded the 5% threshold, the model parameters were adjusted to improve predictive accuracy.

Figure 11 presents the results obtained using the mean absolute error (MAE) metric, which was used to evaluate the regression model's performance. This metric quantifies the absolute difference between the CNN-predicted values and the corresponding ground-truth values. After training for 50 epochs with a maximum of 900 iterations, the model achievedMSE values of 0.1294 % for CO, 0.0200 % for HC, 0.0320 % for CO_2, and 0.0100 % for O_2.

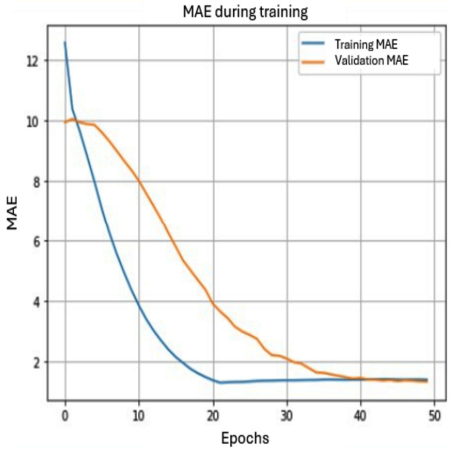


Figure 11. Mean absolute error (MAE) during CNN training

Figure 12 shows the variation of the loss during the convolutional neural network (CNN) training. As expected for this type of network, the loss begins at a relatively high level and decreases rapidly as the model adjusts to the data. This sustained downward trend indicates that the network is learning effectively and progressively optimizing parameters throughout the training process.

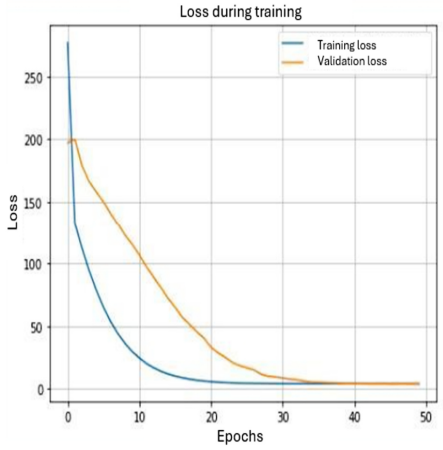


Figure 12. Loss curve during CNN training

3. Results and discussion

To evaluate the model's performance in predicting fault-associated emissions from components and actuators, several tests were conducted using different analysis methods. Figure 13 illustrates 160 records from the testing phase. In this plot, blue circles denote the values estimated by the convolutional neural network (CNN), whereas red crosses indicate the corresponding measured values. As observed, the model's predictions are highly accurate, with errors not exceeding 1% in any case. In addition, the mean absolute error (MAE) and mean squared error (MSE) were computed, and their resulting values confirm the reliability and appropriate fit of the proposed model.

Figure 14 compares the measured values with the predictions produced by the convolutional neural network (CNN) for CO, CO_2, HC, and O_2 concentrations. The blue points, representing the model estimates, cluster closely around the reference line, indicating accurate emission prediction across different operating conditions. In addition, the network correctly discriminated the eight operating conditions of the spark-ignition engine (MEP), supporting the robustness of the proposed model.

Table 4 summarizes representative automotive applications that employ various techniques for emission prediction, together with the key factors associated with each approach.

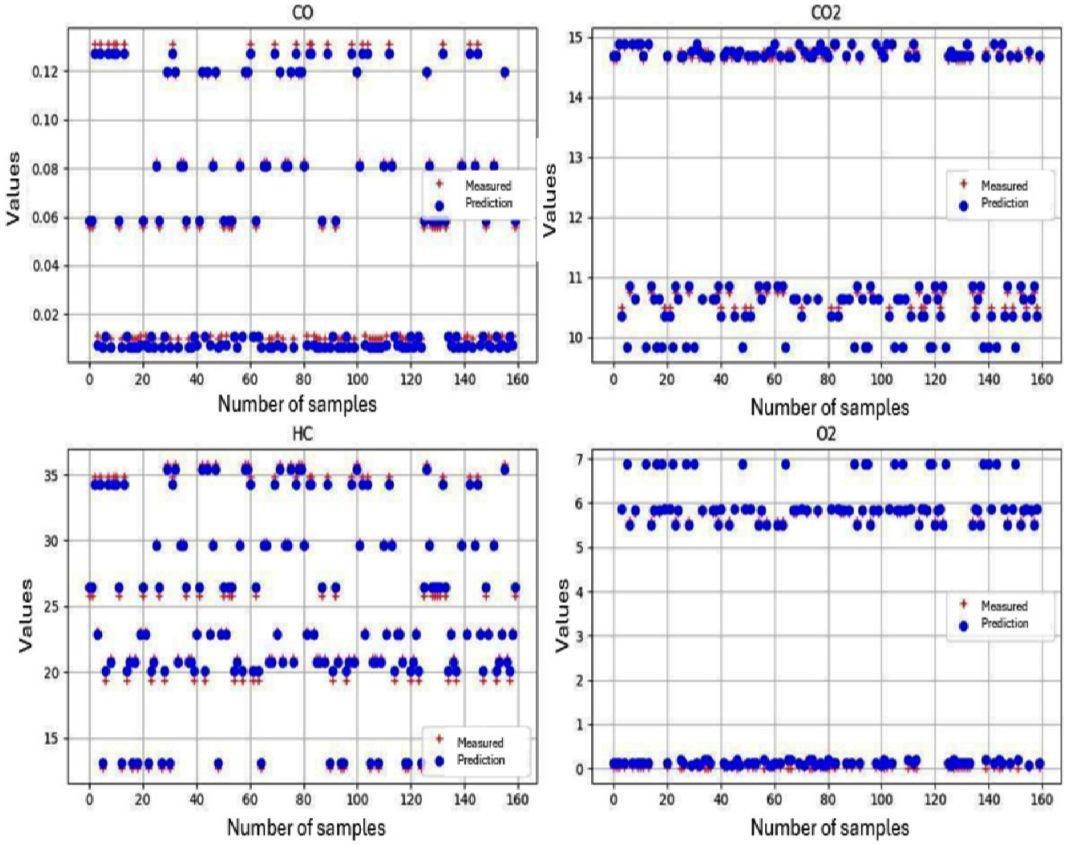


Figure 13. CNN estimation results for exhaust-gas emissions CO, CO_2, HC, O_2

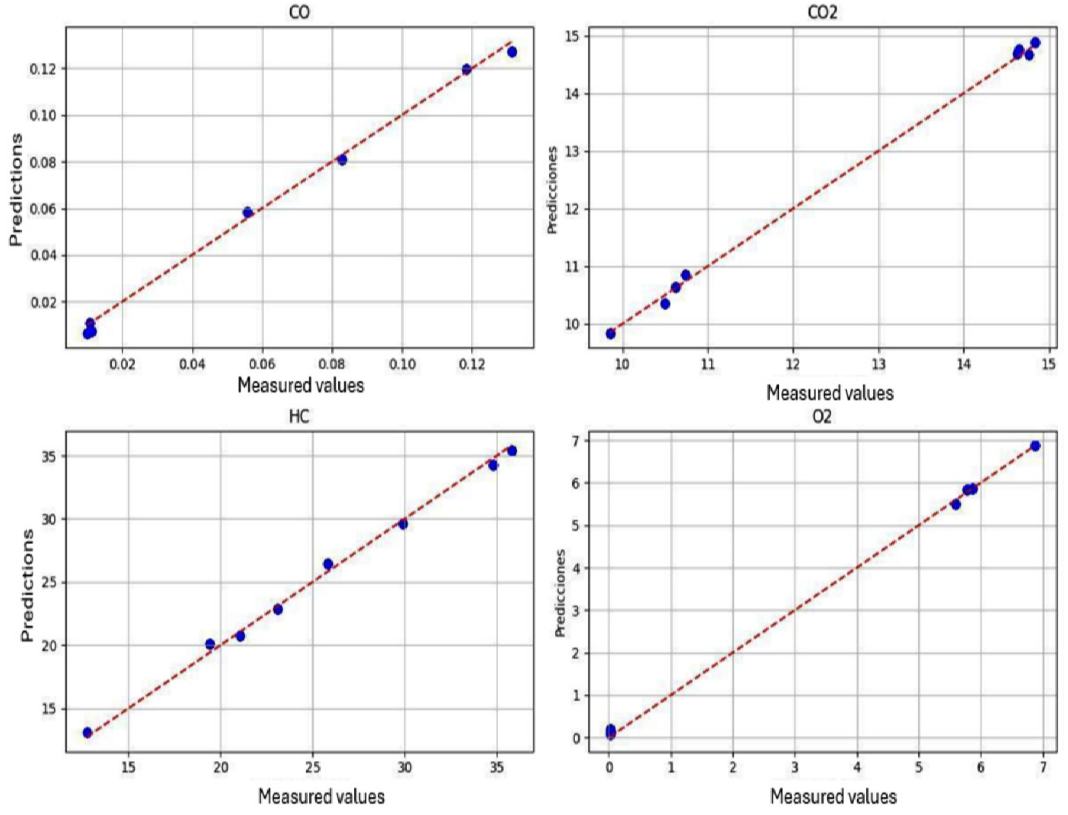


Figure 14. CNN prediction results for CO, CO_2, HC, and O_2 emissions

Table 4. Related studies on emission prediction in the automotive field

N.°	Fuente	Técnica	Aplicación
1	[2]	Feed-forward backpropagation	Emission prediction in MEP
2	[3]	Feed-forward backpropagation	Emission prediction in MEC
3	[7]	RNN y FFNN	Prediction of temperature and emissions in MEC
4	[1]	ANN	Emission prediction in MEP
5	[4]	LSTM y CNN	Prediction of the air–fuel ratio
6	[9]	ANN, CNN-BiLSTM, SVM y XGBoost	Odor assessment in vehicle interiors
7	[10]	RNA	Prediction of effective power and emissions in MEP

The results confirm that the proposed methodology is effective for both fault diagnosis and pollutant-emission prediction. The close agreement between measured values and model estimates indicates that the approach can accurately identify the engine operating state and compute exhaust-gas concentrations with a minimal margin of error.

In [2], neural networks were developed to predict CO and HC, yielding errors of 5.40e-9 and 9.75e-5, respectively. Likewise, the study in [1] reports comparably low prediction errors and an excellent fit for CO_2 estimation, with a coefficient of determination (R^2) of 99.2% and reduced RMSE and MSE values. In addition, the application of artificial neural networks has enabled accurate prediction of engine performance parameters and emissions, such as NO_x and HC, as demonstrated in [10], where accuracies of 99 % and 98 % were achieved, respectively.

Therefore, the results indicate that the proposed methodology, based on MAP signal spectrograms and a convolutional neural network (CNN) , achieves remarkable accuracy in the simultaneous estimation of pollutant emissions (CO, CO_2, HC, and O_2), with errors below 1 %. These values substantially outperform those reported in previous studies using traditional techniques such as fully connected neural networks (ANNs), support vector machines (SVMs), or multivariate regression–based methods, as summarized in Table 4.

Figure 15 presents a statistical analysis of the grouped data, comparing the real operating condition of the experimental unit with the concentrations of unburned hydrocarbons (HC) predicted by the convolutional neural network (CNN). The comparison was performed using Tukey's test at a 95 % confidence level.

Figure 16 shows an interval plot comparing measured HC concentrations with the values estimated by the convolutional neural network (CNN). The results reveal no statistically significant differences among the mean values under the different operating conditions of the spark-ignition engine (MEP), thereby reinforcing the model's reliability across diverse operating scenarios.

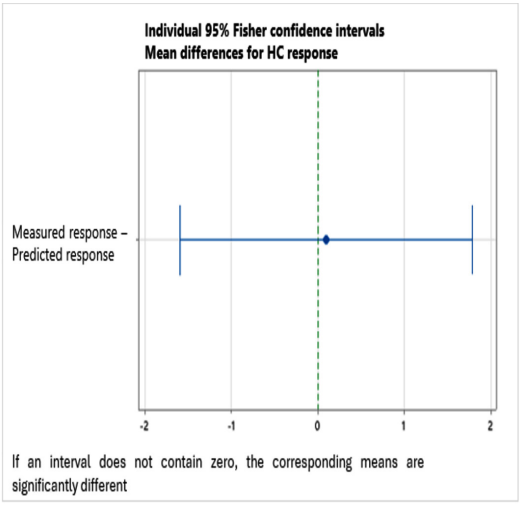


Figure 15. Comparative mean plot of measured HC values and CNN predictions

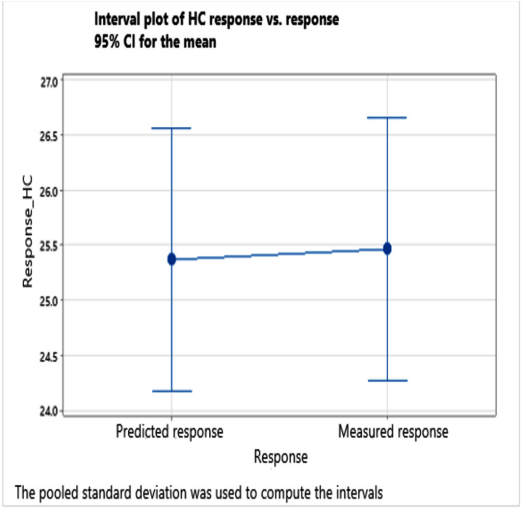


Figure 16. Interval plot comparing measured HC values and CNN predictions

Figure 17 similarly analyzes the relationship between the values estimated by the convolutional neural network (CNN) and the measured values, applying Tukey's test to assess potential differences. The results show that both sets share the same grouping letter (A), indicating no statistically significant differences between them.

Additionally, the analysis of variance (ANOVA) yielded a p-value of 0.911, which supports the consistency and accuracy of the developed model.

Analysis of Variance (ANOVA)

Source	DF	Adj SS	Adj MS	F-value	p-value
Response	1	0.6	0.5555	0.01	0.911
Error	238	10523.5	44.2164		
Total	239	10524.1			

Grouping information using Tukey's method at a 95% confidence level

Response	N	Mean	Grouping
Measured response	120	25.465	A
Predicted response	120	25.368	A

Means that do not share a letter are significantly different

Figure 17. ANOVA results and Tukey's test for measured and CNN-predicted values

3.1. Limitations

This study was conducted on an Otto engine equipped with a modern injection system, within a specific range of operating points and induced fault types. This choice reflects the need for a controlled environment that isolates the effect of faults on emissions and enables CNN training with consistent, repeatable data. Accordingly, the results are representative and robust for the engine and conditions considered, and they demonstrate the potential of the proposed methodology to link component and actuator faults with variations in emissions.

4. Conclusions

This study integrates data acquisition systems with artificial intelligence algorithms to enable early fault diagnosis in spark-ignition engines (MEP). The analysis focused on anomalies in the vacuum line of the fuel-pressure regulator and in the air filter. The proposed method detects faults without dismantling components, thereby providing a minimally invasive diagnostic approach.

Experimental tests confirmed that the CNN-based model, comprising three convolutional layers with normalization and MaxPooling2D, achieves high accuracy in predicting emissions. Low mean squared error (MSE) values were obtained: 0.1294 % for CO, 0.0200 % for HC, 0.0320 % for CO_2, and 0.0100 % for O_2, using 50 training epochs and up to 900 iterations.

In addition, a one-way analysis of variance was performed in Minitab, yielding a p-value of 0.911. This result indicates no statistically significant difference between the measured values and those estimated by the network, supporting the model's reliability.

Overall, these findings highlight the potential of convolutional neural networks as an effective, non-invasive solution for fault detection and emission prediction in spark-ignition engines. The proposed approach achieves a low margin of error and demonstrates practical applicability in internal combustion systems.

Although the metrics used (MSE and MAE) are sufficient to demonstrate the accuracy of the convolutional model, additional indicators may add value in comparative or sensitivity analyses. Therefore, future work should incorporate metrics such as RMSE, R^2, and MAPE.

Contributor role

- **Elmer I. Arias-Montaño:** conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft, writing – review & editing.
- **Rogelio S. León-Japa:** conceptualization, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – review & editing.
- **Pedro García-Jaramillo:** data curation, software, validation.
- **José Maldonado-Ortega:** data curation, software, validation.

References

[1] W. R. Contreras Urgilés, R. S. León Japa, and J. L. Maldonado Ortega, "Predicción de emisiones de CO y HC en motores Otto mediante redes neuronales," *Ingenius*, no. 23, pp. 30–39, Dec. 2019. [Online]. Available: https://doi.org/10.17163/ings.n23.2020.03

[2] F. Narváez, F. E. Sierra Vargas, and M. A. Montenegro Mier, "Modelo basado en redes neuronales para predecir las emisiones en un motor diésel que opera con mezclas de biodiésel de higuerilla," *Informador Técnico*, vol. 76, p. 46, Dec. 2012. [Online]. Available: https://doi.org/10.23850/22565035.28

[3] R. S. Chauhan and N. Shrivastava, "Neuro fuzzy-grey wolf optimization-based modelling and analysis of diesel engine using tire oil with different proportions of 2-EHN," *Fuel*, vol. 384, p. 133849, Mar. 2025. [Online]. Available: https://doi.org/10.1016/j.fuel.2024.133849

[4] F. Sapio, F. Aglietti, P. Ferreri, and A. Savuca, "Neural-network-based modeling of SCR systems for emission simulation: A comprehensive approach," *SAE International Journal of Advances and Current Practices in Mobility*, vol. 07, no. 3, pp. 1437–1452, Sep. 2024. [Online]. Available: https://doi.org/10.4271/2024-24-0042

[5] H. H. Imtiaz, P. Schaffer, Y. Liu, P. Hesse, A. Bergmann, and M. Kupper, "Qualitative and quantitative analyses of automotive exhaust plumes for remote emission sensing application using gas schlieren imaging sensor system," *Atmosphere*, vol. 15, no. 9, p. 1023, Aug. 2024. [Online]. Available: https://doi.org/10.3390/atmos15091023

[6] W. Torres Guin, J. Sánchez Aquino, S. Bustos Gaibor, and M. Coronel Suárez, "Arquitectura de IoT para el monitoreo de emisiones de gases contaminantes de vehículos y su validación a través de machine learning," *Ingenius*, no. 32, pp. 9–17, Oct. 2024. [Online]. Available: https://doi.org/10.17163/ings.n32.2024.01

[7] R. S. Jawad and H. Abid, "HVDC fault detection and classification with artificial neural network based on ACO-DWT method," *Energies*, vol. 16, no. 3, p. 1064, Jan. 2023. [Online]. Available: https://doi.org/10.3390/en16031064

[8] F. Ricci, M. Avana, and F. Mariani, "Enhancing lambda measurement in hydrogen-fueled SI engines through virtual sensor implementation," *Energies*, vol. 17, no. 16, p. 3932, Aug. 2024. [Online]. Available: https://doi.org/10.3390/en17163932

[9] D. Tian, Q. Li, F. Liu, J. Khan, M. Q. Abbas, and Z. Du, "VOC data-driven evaluation of vehicle cabin odor: from ANN to CNN-BiLSTM," *Environmental Science and Pollution Research*, vol. 31, no. 22, pp. 32 826–32 841, Apr. 2024. [Online]. Available: https://doi.org/10.1007/s11356-024-33293-y

[10] I. Cesur and F. Uysal, "Experimental investigation and artificial neural network-based modelling of thermal barrier engine performance and exhaust emissions for methanol-gasoline blends," *Energy*, vol. 291, p. 130393, Mar. 2024. [Online]. Available: https://doi.org/10.1016/j.energy.2024.130393

[11] A. K and P. Rithishbrahma, "Prediction of vehicle carbon emission using machine learning," in *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, Aug. 2024, pp. 1814–1818. [Online]. Available: https://doi.org/10.1109/ICESC60852.2024.10690134

[12] M. A. Rahim, M. M. Rahman, M. S. Islam, A. J. M. Muzahid, M. A. Rahman, and D. Ramasamy, "Deep learning-based vehicular engine health monitoring system utilising a hybrid convolutional neural network/bidirectional gated recurrent unit," *Expert Systems with Applications*, vol. 257, p. 125080, Dec. 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.125080

[13] H. Sun and P. Chen, "Application of neural networks in automotive engine misfire," in *2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*. IEEE, Apr. 2024, pp. 261–264. [Online]. Available: https://doi.org/10.1109/ICEIB61477.2024.10602668

[14] M. Abboush, D. Bamal, C. Knieke, and A. Rausch, "Intelligent fault detection and classification based on hybrid deep learning methods for hardware-in-the-loop test of automotive software systems," *Sensors*, vol. 22, no. 11, p. 4066, May 2022. [Online]. Available: https://doi.org/10.3390/s22114066

[15] S.-C. Lin, S.-F. Su, and Y. Huang, "A time-frequency signal-based convolutional neural network algorithm for fault diagnosis of gasoline engine fuel control system," in *2019 International Conference on System Science and Engineering (ICSSE)*. IEEE, Jul. 2019, pp. 81–87. [Online]. Available: https://doi.org/10.1109/ICSSE.2019.8823285

[16] A. Maged and M. Nour, "Prediction of combustion pressure with deep learning using flame images," *Fuel*, vol. 380, p. 133203, Jan. 2025. [Online]. Available: https://doi.org/10.1016/j.fuel.2024.133203

[17] Z. Li, Z. Qin, W. Luo, and X. Ling, "Gasoline engine misfire fault diagnosis method based on improved YOLOv8," *Electronics*, vol. 13, no. 14, p. 2688, Jul. 2024. [Online]. Available: https://doi.org/10.3390/electronics13142688

[18] Y. Liu, J. Kang, L. Wen, Y. Bai, and C. Guo, "Health status assessment of diesel engine valve clearance based on BFA-BOA-VMD adaptive noise reduction and multi-channel information fusion," *Sensors*, vol. 22, no. 21, p. 8129, Oct. 2022. [Online]. Available: https://doi.org/10.3390/s22218129

[19] W. R. Contreras Urgilés, J. Maldonado Ortega, and R. León Japa, "Aplicación de una red neuronal feed-forward backpropagation para el diagnóstico de fallas mecánicas en motores de encendido provocado," *Ingenius*, no. 21, pp. 32–40, Dec. 2018. [Online]. Available: https://doi.org/10.17163/ings.n21.2019.03

[20] MathWorks. (2025) Redes neuronales convolucionales. The MathWorks, Inc. [Online]. Available: https://upsalesiana.ec/ing35ar7r20

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.08
pISSN: 1390-650X / eISSN: 1390-860X

U-Net–based semantic segmentation of defects in photovoltaic panels

Aplicación de modelos U-Net para segmentación semántica de defectos en paneles fotovoltaicos

Franklin Gómez-López[1], Danny Ochoa-Correa[2,*], Isabel Cabrera-Carrera[1]

Received: 16-06-2025, Received after review: 18-11-2025, Accepted: 04-12-2025, Published: 01-01-2026

Abstract

This article presents a study on the semantic segmentation of defects in crystalline-silicon photovoltaic cells using U-Net–based models trained on electroluminescence (EL) images. The dataset combines laboratory-acquired images with a publicly available benchmark, both manually annotated to identify cracks, dark zones, and collector-bar discontinuities. Eight model variants were trained with controlled variations in input resolution, encoder depth, and regularization strategies. Performance was assessed using per-class precision, recall, and F1-score, complemented by visual inspection through heatmaps and overlays and by expert validation. Segmentation was robust for defects with well-defined morphology, such as dark zones and busbars; however, cracks remained more difficult to detect due to their sparse pixel representation and irregular geometry. Alternative architectures (U-Net++ and MAU-Net) were also evaluated but did not yield meaningful improvements over the optimized U-Net configuration. Overall, the results support the use of this approach for automated inspection under controlled conditions, while highlighting the need for future adaptation to more diverse operational scenarios.

Keywords: Electroluminescence, predictive maintenance, photovoltaic panels, semantic segmentation, U-Net

Resumen

Este artículo presenta un estudio sobre la segmentación semántica de defectos en celdas fotovoltaicas de silicio cristalino mediante modelos basados en U-Net, entrenados con imágenes de electroluminiscencia (EL). Se empleó un conjunto de datos compuesto por imágenes adquiridas en laboratorio y un prototipo público de pruebas, ambos con anotaciones manuales de grietas, zonas oscuras y discontinuidades en barras colectoras. Se entrenaron ocho versiones del modelo, incorporando variaciones controladas en la resolución, la profundidad del codificador y las estrategias de regularización. La evaluación incluyó métricas clase a clase (precisión, recall y F1-score), análisis visual mediante mapas de calor y superposiciones, así como validación por expertos. Si bien la segmentación fue consistente en defectos de morfología clara, como zonas oscuras y barras colectoras, las grietas presentaron mayores dificultades debido a su baja densidad de píxeles y geometría irregular. Asimismo, se analizaron arquitecturas alternativas (U-Net++ y MAU-Net), sin evidenciar mejoras relevantes frente a la configuración optimizada de U-Net. Los resultados respaldan el uso de este enfoque en tareas de inspección automatizada bajo condiciones controladas, y se proponen extensiones para su aplicación en contextos operativos más diversos.

Palabras clave: electroluminiscencia, mantenimiento predictivo, paneles solares, segmentación semántica, U-Net

[1,*]Faculty of Engineering, Universidad de Cuenca, Ecuador.

[2]Department of Electrical, Electronics and Telecommunications Engineering, Universidad de Cuenca, Ecuador.
Corresponding author ✉: danny.ochoac@ucuenca.edu.ec.

Suggested citation: F. Gómez-López, D. Ochoa-Correa and I. Cabrera-Carrera, "U-Net–based semantic segmentation of defects in photovoltaic panels," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 103-113, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.08.

1. Introduction

Photovoltaic (PV) energy systems have become a cornerstone of global energy-transition efforts, particularly through distributed generation and decentralized electrification. As PV installations continue to expand, maintaining their operational integrity through accurate diagnostic methods is increasingly critical. Electroluminescence (EL) imaging is widely used to detect structural anomalies in PV modules, including microcracks, broken fingers, and electrically inactive areas [1,2]. Although these defects are often imperceptible during visual inspections, they can progressively degrade module performance over time.

Compared with infrared thermography and ultraviolet fluorescence, EL imaging offers superior spatial resolution for detecting early-stage defects without relying on solar irradiance or temperature gradients [3,4]. However, raw EL images often exhibit noise and uneven illumination, which complicate the identification of subtle defect patterns. These limitations underscore the need for advanced image-analysis techniques capable of reliably isolating fine, heterogeneous fault characteristics.

Semantic segmentation, which classifies individual pixels into predefined categories, enables precise defect localization in EL images [5, 6]. Among deep learning approaches, convolutional neural networks (CNNs) have demonstrated strong performance in segmentation tasks. In particular, the U-Net architecture is widely adopted for defect detection in PV modules [7]. Originally developed for biomedical image analysis, U-Net effectively integrates multi-scale feature extraction with the preservation of spatial resolution, thereby supporting the detection of narrow, localized structures such as microcracks and soldering inconsistencies in photovoltaic cells [8, 9].

Recent research has adapted and refined U-Net variants for PV defect segmentation by incorporating attention mechanisms, residual connections, and hybrid loss functions [10, 11]. These modifications enhance model robustness under varying illumination conditions and improve segmentation accuracy given inter-cell variability in PV module datasets. In addition, publicly accessible EL image datasets, including those released by Fraunhofer ISE and Sandia National Laboratories, support benchmarking and the reproducibility of segmentation results [12, 13].

In this study, multiple U-Net architectures trained on EL images of crystalline-silicon solar cells are implemented and evaluated. Using both proprietary and public datasets, segmentation accuracy across distinct defect types is analyzed. The evaluation incorporates parameter-optimization strategies and quantitative comparisons to assess the suitability of U-Net–based methods for supporting automated inspection under controlled laboratory conditions.

2. Materials and Methods

2.1. Datasets

2.1.1. In-house electroluminescence dataset

The primary dataset was collected at the Microgrid Laboratory of the University of Cuenca. It comprises 180 solar cells extracted from decommissioned monocrystalline and polycrystalline PV modules. Each cell was imaged under controlled conditions using an OWL 640 M SWIR camera (Raptor Photonics) equipped with an InGaAs sensor sensitive to the 900–1700 nm spectral range. This interval provides improved contrast for defect identification in crystalline silicon compared with sensors operating in the visible spectrum.

To ensure uniform illumination, the cells were polarized using a Chroma DC programmable power supply. The applied current was varied between $\frac{1}{6}I_{SC}$ and I_{SC}, following IEC TS 60904-13 standards, to examine how excitation intensity influences luminescence features. Image acquisition was managed using the XCAP-Std platform, enabling real-time adjustment of exposure time, gain, and frame rate. The optimal acquisition parameters were an exposure time of 30 and 50 ms, a gain factor of 2.5, and a polarization current of approximately 6 A, corresponding to about $\frac{2}{3}I_{SC}$.

To validate the quality of the acquisition system, the Signal-to-Noise Ratio (SNR) was quantified in accordance with the IEC TS 60904-13 standard [14], and defined as equation (1):

$$SNR_{IEC} = \frac{\sum\{0.5 \cdot (I_1 + I_2) - I_{BG}\}}{\sum[\sqrt{0.5} \cdot |I_1 - I_2| \cdot (\frac{2}{\pi})^{-0.5}]} \quad (1)$$

Where I_1 and $I_2$are consecutive EL captures and IBG represents the background noise. The experimental setup achieved SNRIEC > 45, meeting the criteria for laboratory measurements.

Subsequently, raw EL images underwent preprocessing that included background subtraction, perspective correction, and local contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE), configured with a clip limit of 2.0 and a tile grid size of 8×8. This preprocessing pipeline further improved image quality for segmentation tasks.

2.1.2. Public benchmark dataset

In addition to the proprietary dataset, the publicly available ELPV dataset from the ZAE Bayern group was used for validation. This dataset comprises EL images and corresponding binary crack masks for 2,624 solar cells obtained under factory conditions. Despite differences in resolution and mask structure compared to the in-house dataset, scripts were employed to unify image dimensions and encoding formats.

By integrating both internal and external datasets, the combined dataset supports model generalization across different panel types and acquisition configurations. Previous studies adopting similar hybrid-dataset strategies have reported improved segmentation accuracy and cross-domain robustness.

2.1.3. Data Integration and Preprocessing

Since the public ELPV dataset provided only binary crack masks, a re-annotation process was conducted using the Supervisely platform [15], to manually generate ground-truth masks for the missing classes. This ensured that the in-house and public datasets shared a consistent multi-class annotation schema (background and three defect classes).

Table 1 summarizes the distribution of defect instances across the training and validation sets. The dominance of the busbar class is expected, as this structural component is present in every cell. This class imbalance was addressed during training through the data-augmentation pipeline described in the following section.

Table 1. Distribution of defect instances in the Training and Validation sets

Defect Class	Training Instances	Validation Instances
Busbars	1453	96
Cracks	608	92
Dark Zones	425	84

2.2. Acquisition Setup and Pipeline

The complete acquisition setup is depicted in Figure 1. The SWIR camera was positioned orthogonally above the cell, maintaining a fixed lens-to-target distance of 50 cm. Imaging was performed in a darkened chamber to eliminate ambient-light interference. Non-uniformity correction (NUC) was applied using a three-point method (Offset + Gain + Dark) to minimize fixed-pattern noise.

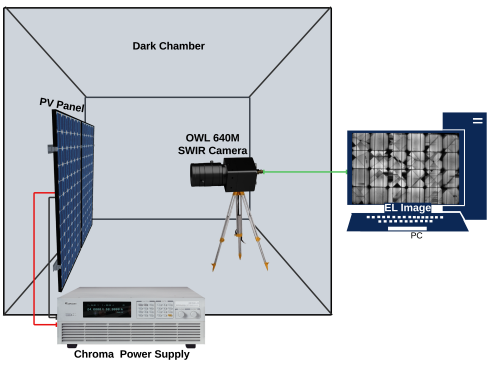


Figure 1. Experimental setup showing the OWL 640 M SWIR camera and the PV module under test.

All image-processing steps were implemented in Python using the OpenCV and NumPy libraries. Evaluation metrics, including the Contrast Improvement Ratio (CIR), Peak-to-Low ratio (PL), and visual heatmaps, were extracted from each preprocessed image.

2.3. Annotation Consistency

Inter-annotator agreement was periodically assessed using the Dice coefficient, see equation (2):

$$\mathrm{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{2}$$

Where A and B represent pixel sets annotated independently by different evaluators. A mean Dice score above 0.85 was consistently maintained throughout the annotation process.

These datasets served as the basis for training and validating the segmentation models described in the following section.

2.4. Model Architecture

2.4.1. Standard U-Net with VGG16 encoder

The architecture employed in this study is based on U-Net, originally developed for biomedical image segmentation [7]. Its encoder-decoder structure enables pixel-level classification by integrating multi-scale features while preserving spatial resolution. In this implementation, the contracting path is replaced with a VGG16 encoder pre-trained on the ImageNet dataset [16].

Since the standard VGG16 architecture expects a three-channel input (RGB) and EL images are single-channel (grayscale), the input was adapted during preprocessing by replicating the single channel three times to match the expected tensor dimensions without modifying the pre-trained weights. This configuration enhances feature representation at the input stage and improves convergence stability [17].

Each downsampling stage in the encoder includes two 3×3, convolutional layers followed by a 2×2 max-pooling operation. The encoder generates five feature maps of increasing depth. To expand the receptive field while preserving spatial dimensions, the bottleneck employs dilated convolutions with a fixed dilation rate of $r = 2$.

Figure 2 illustrates the U-Net variant adopted in this study, which incorporates dilated convolutions in the bottleneck to expand the receptive field without increasing the number of parameters [8, 18].

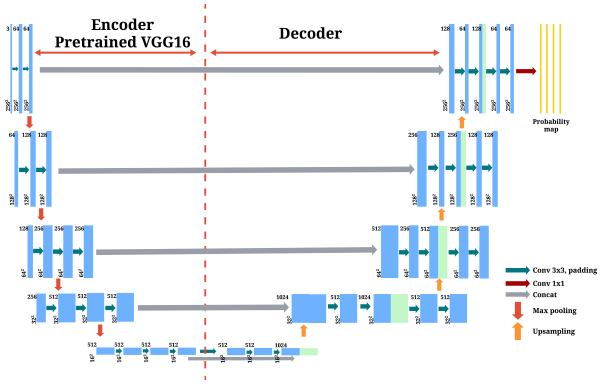


Figure 2. U-Net architecture with VGG16 encoder and dilated convolution bottleneck.

2.4.2. Layer configuration, activation functions and multi-class output

All convolutional layers use ReLU activation functions and batch normalization to improve training stability. A dropout rate of 0.2 is applied in the bottleneck and decoder layers to mitigate overfitting.

The decoder upsamples feature maps using transposed convolutions, with skip connections linking each upsampling block to the corresponding encoder stage. The output layer consists of a 1×1 convolution followed by a softmax activation to produce a pixel-wise probability distribution over C defect classes, see equation (3):

$$\text{Output}_{i,j,c} = \frac{\exp(z_{i,j,c})}{\sum_{k=1}^{C} \exp(z_{i,j,k})} \tag{3}$$

Where $z_{i,j,c}$ is the raw score for class c at pixel (i, j), and $C = 3$ represents the defect classes: cracks, dark zones, and busbars.

Input images are grayscale EL images resized to 256×256 pixels. Output segmentation masks preserve the same resolution. The entire pipeline was implemented using the PyTorch framework.

Prior studies indicate that the combination of skip connections, a symmetric architecture, and multi-class outputs is well-suited to detecting subtle defect patterns in EL images [10, 19].

Alternative architectures, such as MAU-Net with spatial-channel attention mechanisms, were also examined for comparison and are discussed in the Results and Discussion.

2.5. Training Process

2.5.1. Preprocessing and data augmentation

Before training, all EL images were resized to a standard resolution of 256×256 pixels. To distinguish image features from categorical labels effectively during resizing, bilinear interpolation was applied to the grayscale EL images, whereas nearest-neighbor interpolation was used for the masks to preserve discrete class values. The pixel values were then normalized to the [0, 1] range to ensure a consistent input format and enhance model stability. Each grayscale input image was paired with a one-hot encoded mask representing the three defect categories: cracks, dark zones, and busbars.

To expand the variability of training samples, data augmentation was performed on-the-fly. Transformations included horizontal and vertical flipping, random rotations up to $\pm15°$, zoom scaling between $0.9\times$ and $1.1\times$, and elastic deformations. These augmentations were applied probabilistically using the Albumentations library, ensuring that the structural integrity of the masks was maintained.

This augmentation approach enhances generalization, particularly for models trained on small or homogeneous EL datasets. Similar strategies have been used to address class imbalance and improve sensitivity to rare defect patterns.

2.5.2. Configuración de hiperparámetros y programador de tasa de aprendizaje

Model training was conducted using the Adam optimizer, which combines adaptive gradient estimation with momentum. The selected hyperparameters were as follows:

- Learning rate: $lr = 0.0018$
- First moment decay rate: $\beta_1 = 0.9$
- Second moment decay rate: $\beta_2 = 0.999$
- Numerical stability term: $\epsilon = 1 \times 10^{-8}$

Training was driven by the standard Categorical Cross-Entropy loss function, see equation (4):

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c}) \tag{4}$$

Where $y_{i,c}$ is the true label (one-hot encoded) and $\hat{y}_{i,c}$ is the predicted probability. Although class imbalance is inherent in EL defect detection, it was addressed primarily through the aggressive data-augmentation strategy described in the previous section, rather than by applying class weights in the loss function.

A step-based learning-rate scheduler was used to improve convergence and avoid local minima. The learning rate was reduced by a factor $\gamma = 0.05$ every 8 epochs. This gradual decay facilitated fine-tuning during later training phases. Early stopping was enabled to terminate training if the validation loss did not improve for 10 consecutive epochs.

The training process was iterative. Each subsequent model version was initialized with the weights from the best-performing previous version, allowing progressive refinement without reinitialization. The final model version (U-Net_v32) was trained for 48 epochs, with the lowest validation loss observed at epoch 26. The corresponding values for training and validation loss are shown in equation (5):

$$\text{Train loss} = 0.0313 \quad \text{Val loss} = 0.0627 \tag{5}$$

Figure 3 displays the evolution of loss values across model versions. These visualizations aided in monitoring training behavior and validating incremental improvements.

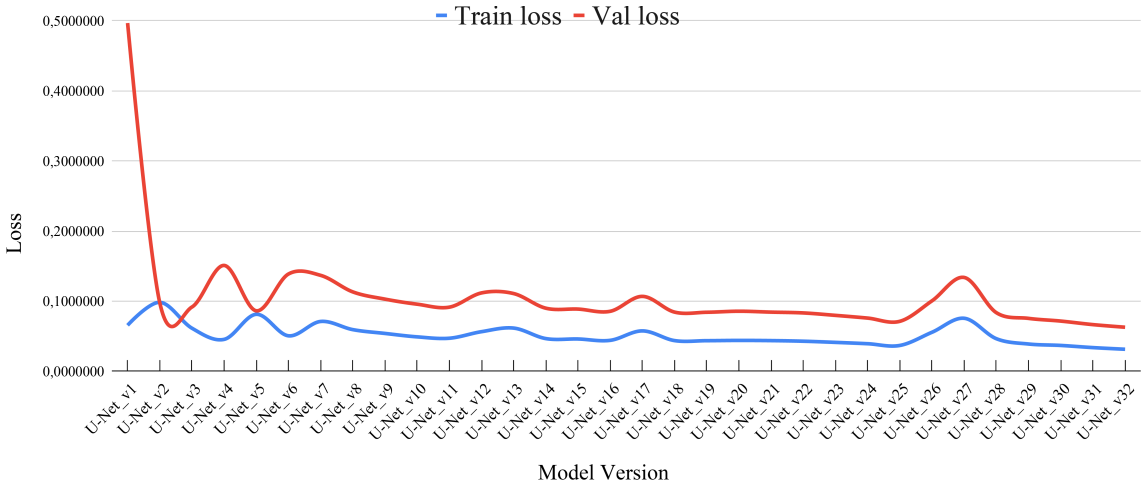


Figure 3. Evolution of training and validation loss values across iterative U-Net model versions.

To ensure reproducibility, all experiments were conducted using the PyTorch framework (v2.3.1) in Google Colab Pro environment equipped with an NVIDIA T4 GPU. A fixed random seed of 42 was set for weight initialization and data partitioning. Training was performed with a batch size of 16 for 48 epochs, using the Adam optimizer with an initial learning rate of 0.0018 and a step-based decay scheduler ($step\_size = 8$, $\gamma = 0.05$). The best model weights were saved based on the minimum validation loss, which typically converged around epoch 26.

The final U-Net model has a storage footprint of approximately 233 MB. During the inference phase, processing a batch of 16 images (256×256 pixels) took an average of 0.52 seconds, translating to an inference latency of approximately 32.6 ms per image ($\sim$30 FPS).

2.6. Evaluation Metrics

2.6.1. Precision, Recall and F1-score

To assess model performance at the pixel level, three complementary metrics were computed: precision, recall, and F1-score. These metrics were calculated independently for each defect class — cracks, dark zones, and busbars — as recommended in segmentation studies focused on PV defect detection [8, 18].

Precision quantifies the proportion of correctly identified defect pixels among all pixels predicted as defective, see equation (6):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Where TP represents true positives (correctly predicted defect pixels) and FP denotes false positives (non-defect pixels incorrectly predicted as defects).

Recall, also referred to as sensitivity, measures the proportion of actual defect pixels that are correctly identified, see equation (7):

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

With FN indicating false negatives (defect pixels that the model failed to detect).

The F1-score, calculated as the harmonic mean of precision and recall, provides a balanced metric for evaluating classification accuracy, see equation (8):

$$\text{F1-score} = \frac{2TP}{2TP + FP + FN} \tag{8}$$

These metrics are particularly relevant in multiclass segmentation tasks involving class imbalance and subtle morphological features, such as microcracks [10].

2.6.2. Validation Strategy and Confusion Matrix

To evaluate model performance and monitor convergence during the iterative training of 32 model versions, a fixed hold-out validation strategy was employed instead of k-fold cross-validation. The dataset was strictly partitioned into three independent subsets: training (1453 images, including augmented data), validation (96 original images), and testing. This approach ensured that the evaluation metrics remained comparable across different training epochs and hyperparameter adjustments, while avoiding data leakage from the augmentation process.

A confusion matrix was generated using the best-performing model (U-Net_v32) on the validation set to visualize class-specific prediction errors. Figure 4 presents the normalized confusion matrix. A higher misclassification rate was observed for the crack class, which is attributed to the sparse and fragmented spatial distribution of this defect type compared with the more distinct features of busbars and dark zones.

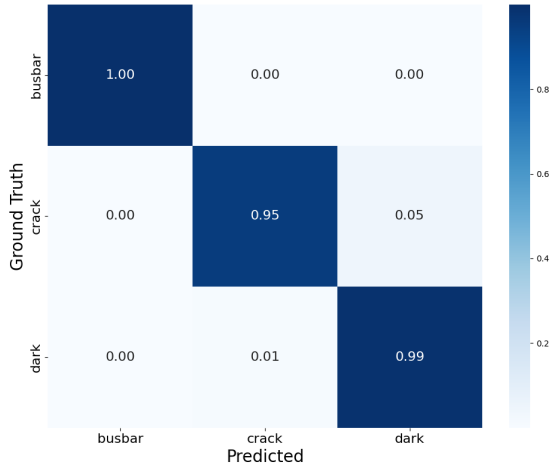


Figure 4. Normalized confusion matrix for U-Net_v32 evaluated on validation data.

This validation protocol enables a clear assessment of the model's generalization capability on unseen data, with particular emphasis on the correct identification of minority defect classes.

3. Results and Discussion

3.1. Performance by Defect Class

3.1.1. Segmentation of cracks, dark zones and busbars

The segmentation performance of the trained U-Net models varied across defect classes. As shown in Table 2, the highest F1-score was obtained for dark zones, which typically occupy larger, more homogeneous regions. Busbar discontinuities also exhibited consistent segmentation performance, particularly when there was clear contrast against the cell background.

Table 2. Mean segmentation metrics by class (UNet_v32).

Defect Class	Precision	Recall	F1-score
Cracks	0.6794	0.6465	0.6557
Dark Zones	0.8714	0.9066	0.9064
Busbars	0.9151	0.9031	0.9083

Cracks, in contrast, posed greater challenges.

These defects are narrow, irregular, and frequently overlap with low-contrast textures. Consistent with prior studies by Pratt et al. [8] and Deitsch et al. [18], the crack class exhibited lower precision and recall, attributable to its pixel-level sparsity and morphological variability.

Qualitative inspection further indicated that cracks located near the cell edges or partially occluded by finger metallization were more frequently missed. These observations are consistent with benchmarking results obtained from EL datasets such as ELPV [19].

Figure 5 illustrates representative segmentation outcomes and their comparison with the ground truth for the three defect types.

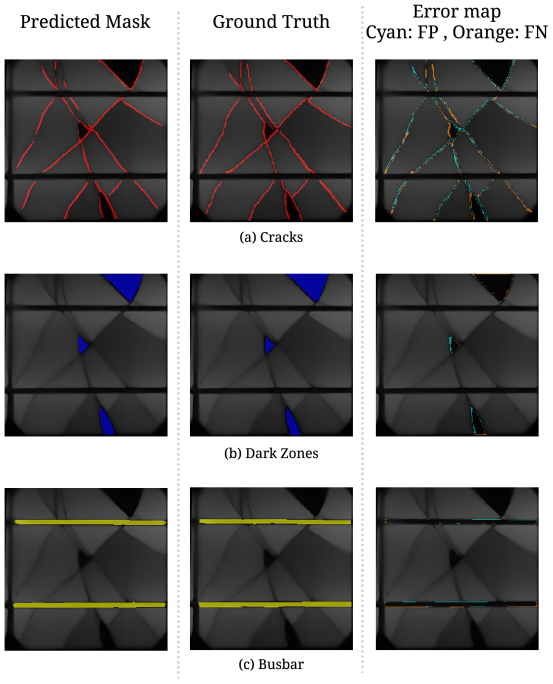


Figure 5. Examples of predicted segmentation masks for (a) cracks, (b) dark zones, and (c) busbars. Each row shows the original EL image, the predicted segmentation overlaid on the image, the ground truth mask overlaid, and an error map in which cyan represents false positives (FP) and orange represents false negatives (FN).

For each class, the figure shows the original EL image, the predicted mask overlaid on the image, the

corresponding ground truth mask, and an error map highlighting discrepancies. In the error maps, cyan denotes false positives (pixels incorrectly predicted as defective), whereas orange denotes false negatives (defect pixels not detected by the model).

For cracks, the error maps reveal fragmented boundaries, particularly in regions with weak luminescence, which corroborates the limitations observed in the quantitative evaluation. For dark zones and busbars, the false-positive and false-negative regions are minimal, consistent with their comparatively higher F1-scores.

3.1.2. Error analysis and class imbalance

A persistent challenge during training was class imbalance. Background pixels accounted for over 85% of the dataset, while cracks represented less than 5%. This imbalance affected both the optimization process and metric interpretation, as the loss function tended to prioritize the majority classes.

To mitigate this issue, class-weighted loss functions were employed, assigning higher penalties to errors in underrepresented classes. In addition, data augmentation was specifically targeted to increase the number of samples containing cracks. These adjustments improved the model's sensitivity to minority classes without compromising training stability.

As illustrated in the normalized confusion matrix in Figure 4, the highest misclassification rates occurred for the crack class, particularly in cases where diffuse or low-contrast defects were mistaken for dark zones.

False positives in the busbar class occasionally appeared near cell edges. These errors were attributed to abrupt brightness variations caused by non-uniform illumination or optical reflections during image acquisition.

3.2. Comparative Analysis Between Model Version

3.2.1. Iterative evaluation from U-Net_v24 to U-Net_v322

An iterative evaluation process was conducted across model versions from U-Net_v24 to U-Net_v32. Each iteration introduced controlled modifications aimed at improving performance under constrained data and hardware conditions. These adjustments included increasing input resolution from 128×128 to 256×256 pixels, varying dropout rates between 0.15 and 0.3, and reducing batch size to accommodate higher-resolution inputs within the available GPU memory.

Table 3 summarizes the evolution of F1-scores for the crack and busbar classes.

Table 3. F1-score evolution for crack and busbar segmentation

Model Version	Cracks (F1)	Busbars (F1)
U-Net_v24	0.6151	0.9037
U-Net_v27	0.4519	0.8725
U-Net_v30	0.6336	0.9069
U-Net_v32	0.6557	0.9083

Improvements were most evident in the crack class, where earlier model versions failed to detect faint linear patterns, particularly near cell boundaries. The use of higher input resolution and deeper encoder configurations in U-Net_v30 and U-Net_v32 improved the network's ability to preserve detailed features during the downsampling process.

Figure 6 illustrates the progression of F1-scores across model versions.

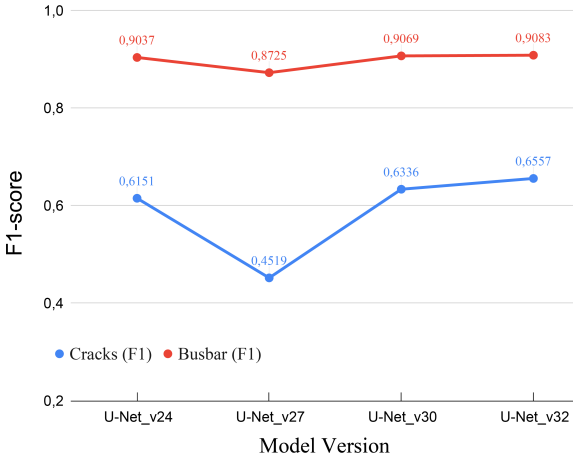


Figure 6. F1-score progression across model versions (Cracks and Busbars).

These outcomes are consistent with previous observations indicating that gradual refinement of hyperparameters and input structure, rather than radical architectural changes, tends to yield more reliable improvements in segmentation models trained on limited or domain-specific datasets.

3.2.2. Discussion on alternative architectures: U-Net++, MAU-Net

Two U-Net variants, U-Net++ and MAU-Net, were evaluated under the same training conditions. UNet++ adds nested skip connections to enhance gradient flow and semantic fusion, while MAU-Net introduces spatial and channel attention to prioritize relevant features and suppress noise.

Both variants slightly improved recall for cracks in low-contrast regions; however, the overall F1-score gain was marginal (below 1.5%) compared to U-Net_v32. MAU-Net also incurred a 35% increase in inference time.

Given these trade-offs, the baseline U-Net was preferred because it offers greater stability, lower complexity, and more consistent performance across defect types.

3.3. Visual and Qualitative Validation

3.3.1. Heatmaps and overlays on individual cells

In addition to quantitative metrics, visual inspection is essential for understanding model behavior under realistic diagnostic conditions. To support this, heatmaps were generated from softmax output probabilities for each defect class and overlaid on the original electroluminescence (EL) images to visualize the model's prediction confidence.

As illustrated in Figure 7, dark zones were consistently identified due to their well-defined boundaries and contrast characteristics. In comparison, crack predictions occasionally missed fragmented sections near cell borders or beneath finger metallization, where luminescence levels were lower. These observations are consistent with findings reported in prior evaluations using EL datasets such as ELPV [19] and with approaches that employ visual overlays for segmentation verification [8].

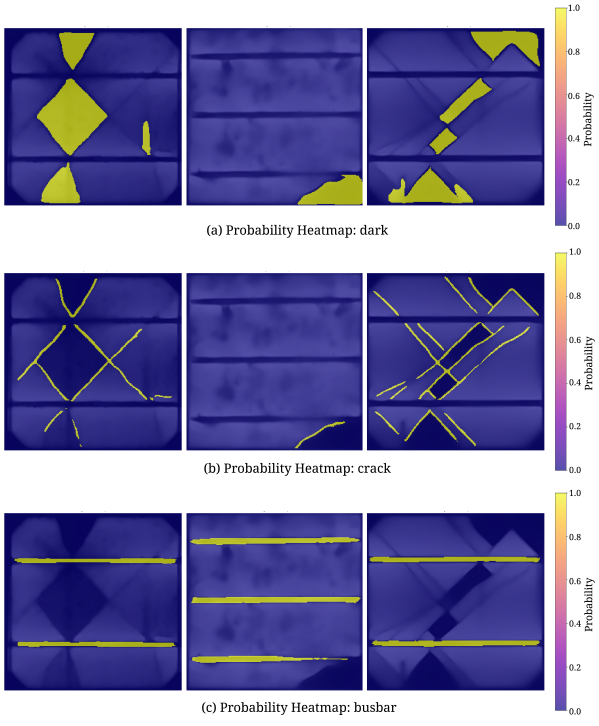


Figure 7. Overlay examples: (a) dark zones correctly predicted; (b) partial crack detection; (c) overprediction in the busbar area.

3.3.2. Expert evaluation and prediction visualization

To complement automated performance metrics, a qualitative review was conducted by three experts in PV diagnostics. Each expert independently evaluated a set of 30 predicted segmentation masks, focusing on contour accuracy, spatial consistency, and interpretability.

The assessment criteria were aligned with visual inspection standards adapted from IEC TS 60904-13 for EL image analysis [14].

Most predictions for dark zones and busbars were considered highly consistent with diagnostic expectations, with minimal segmentation discrepancies. Figure 8 presents representative error maps for these classes.

Cyan pixels indicate false positives, and orange pixels represent false negatives. In most cases, only minor misalignments were observed, often appearing as thin traces resulting from slight differences in edge thickness between the predicted masks and the ground truth annotations.

Crack segmentation was more challenging. Some prediction masks exhibited irregular contours or overestimated defect regions in low-contrast areas. Figure 9 provides examples of these error cases.

Prominent orange traces denote missed crack segments (false negatives), while cyan regions correspond to false-positive predictions in ambiguous zones. These patterns frequently coincided with electroluminescence gradients induced by inconsistent polarization currents or subtle misalignment during image acquisition.

This qualitative analysis revealed patterns that are useful for guiding model refinement. Future implementations may benefit from selective post-processing or attention-based modules aimed at reducing false positives and improving spatial resolution in regions where defect boundaries are poorly defined.

Overall, the combination of visual overlays, probability-based heatmaps, and expert assessment offered a multidimensional perspective on model performance that complemented standard metrics such as F1-score and confusion matrices.

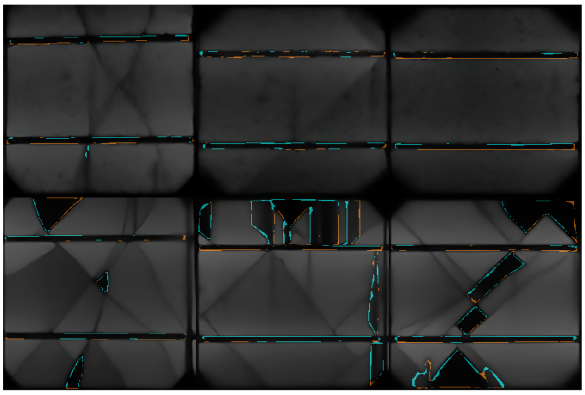


Figure 8. Representative error maps for correctly segmented dark zones and busbars. Cyan indicates false positives, and orange denotes false negatives.

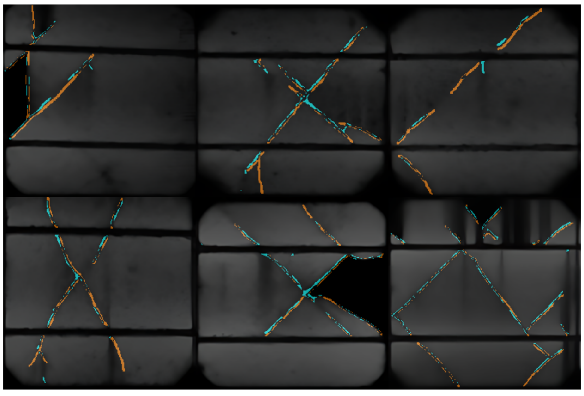


Figure 9. Examples of segmentation errors. Missed cracks are highlighted in orange, while false positives in lowcontrast areas are shown in cyan.

4. Conclusions

This study examined U-Net-based convolutional architectures for semantic segmentation of defects in crystalline silicon photovoltaic cells using EL imaging. A dual-source dataset, including in-house and public EL images, was used for training and evaluation, with manual annotations for cracks, dark zones, and busbar discontinuities.

The models performed consistently on defects with clear morphology and strong contrast, such as dark zones and interrupted busbars. Cracks remained the most difficult class to segment, due to their sparse distribution and irregular geometry, consistent with previous findings [8,18]. From a practical perspective, the present results are constrained by several factors. Many cracks appear as thin, irregular trajectories with very low pixel density and subtle contrast relative to the surrounding material, which leads the network to miss very short or faint segments and, in some cases, to confuse small crack fragments with background texture or acquisition artefacts. In addition, the mandatory resizing of single-cell EL images to 256×256 pixels, required by the VGG16-based encoder, may compress the thinnest structures into only a few pixels and thus limit sensitivity to the smallest defects. The dataset size and diversity are also limited, since all images were captured under controlled laboratory conditions and do not include field EL acquisitions with varying bias currents, surface contamination, or optical non-uniformities. These factors should be considered when extrapolating the reported performance to other setups or operating conditions.

Eight model versions were evaluated using a structured training workflow with progressive modifications in encoder depth, input resolution, regularization, and learning strategies. The final configuration (U-Net_v32) provided the best balance between accuracy and computational efficiency. Variants such as U-Net++ and MAU-Net yielded only minor performance gains at the cost of increased inference time [11].

Quantitative metrics, visual outputs, and expert assessments were jointly used to evaluate model behavior. This integrated validation confirmed the utility of U-Net models for defect localization under controlled conditions.

Future work should focus on (1) expanding the dataset with field-acquired EL images that capture real-world variability, and (2) developing and integrating post-processing techniques to reduce false positives and improve boundary accuracy in low-contrast regions. These developments would enhance the applicability of segmentation models in operational PV diagnostics and manufacturing quality control.

Acknowledgments

The authors express their appreciation to the Microgrid Laboratory of the University of Cuenca for providing the facilities and technical support necessary for this study. Special thanks are extended to the laboratory staff, Vinicio Iñiguez, Edisson Villa, and Pablo J. Delgado, for their assistance during the experimental phase.

Rol de autores

- **Franklin Gómez-López:** conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft.
- **Danny Ochoa-Correa:** formal analysis, funding acquisition, investigation, project administration, resources, supervision, writing – review & editing.

- **Isabel Cabrera-Carrera:** data curation, formal analysis, investigation, methodology, supervision, writing – review & editing.

References

[1] K. G. Bedrich, M. Bliss, T. R. Betts, and R. Gottschalg, "Electroluminescence imaging of PV devices: Determining the image quality," in *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*. IEEE, Jun. 2015, pp. 1–5. [Online]. Available: https://doi.org/10.1109/PVSC.2015.7356011

[2] T. Fuyuki and A. Kitiyanan, "Photographic diagnosis of crystalline silicon solar cells utilizing electroluminescence," *Applied Physics A*, vol. 96, no. 1, pp. 189–196, Dec. 2008. [Online]. Available: http://doi.org/10.1007/s00339-008-4986-0

[3] M. Akram and J. Bai, "Defect detection in photovoltaic modules based on image-to-image generation and deep learning," *Sustainable Energy Technologies and Assessments*, vol. 82, p. 104441, Oct. 2025. [Online]. Available: http://doi.org/http://doi.org/

[4] J. Wang, L. Bi, P. Sun, X. Jiao, X. Ma, X. Lei, and Y. Luo, "Deep-learning-based automatic detection of photovoltaic cell defects in electroluminescence images," *Sensors*, vol. 23, no. 1, p. 297, Dec. 2022. [Online]. Available: http://doi.org/10.3390/s23010297

[5] Q. Liu, M. Liu, C. Wang, and Q. J. Wu, "An efficient CNN-based detector for photovoltaic module cells defect detection in electroluminescence images," *Solar Energy*, vol. 267, p. 112245, Jan. 2024. [Online]. Available: http://doi.org/10.1016/j.solener.2023.112245

[6] J. Fioresi, D. J. Colvin, R. Frota, R. Gupta, M. Li, H. P. Seigneur, S. Vyas, S. Oliveira, M. Shah, and K. O. Davis, "Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images," *IEEE Journal of Photovoltaics*, vol. 12, no. 1, pp. 53–61, Jan. 2022. [Online]. Available: http://doi.org/10.1109/jphotov.2021.3131059

[7] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 2015, pp. 234–241. [Online]. Available: http://doi.org/10.1007/978-3-319-24574-4_28

[8] L. Pratt, D. Govender, and R. Klein, "Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation," *Renewable Energy*, vol. 178, pp. 1211–1222, Nov. 2021. [Online]. Available: http://doi.org/10.1016/j.renene.2021.06.086

[9] H. Eesaar, S. Joe, M. U. Rehman, Y. Jang, and K. T. Chong, "SEiPV-net: An efficient deep learning framework for autonomous multi-defect segmentation in electroluminescence images of solar photovoltaic modules," *Energies*, vol. 16, no. 23, p. 7726, Nov. 2023. [Online]. Available: https://doi.org/10.3390/en16237726

[10] M. R. U. Rahman and H. Chen, "Defects inspection in polycrystalline solar cells electroluminescence images using deep learning," *IEEE Access*, vol. 8, pp. 40 547–40 558, 2020. [Online]. Available: http://doi.org/10.1109/access.2020.2976843

[11] R. A. Mamun Rudro, K. Nur, F. A. Al Sohan, M. Mridha, S. Alfarhood, M. Safran, and K. Kanagarathinam, "SPF-Net: Solar panel fault detection using U-net based deep learning image classification," *Energy Reports*, vol. 12, pp. 1580–1594, Dec. 2024. [Online]. Available: http://doi.org/10.1016/j.egyr.2024.07.044

[12] S. Deitsch, V. Christlein, S. Berger, C. Buerhop-Lutz, A. Maier, F. Gallwitz, and C. Riess, "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Solar Energy*, vol. 185, pp. 455–468, Jun. 2019. [Online]. Available: http://doi.org/10.1016/j.solener.2019.02.067

[13] U. Hijjawi, S. Lakshminarayana, T. Xu, and M. Rahman, "A benchmarking study of instance segmentation methods for photovoltaic cell defect detection using electroluminescence images," *Solar Energy*, vol. 303, p. 114083, Jan. 2026.

[14] IEC, *IEC Technical Specification 60904-13. Photovoltaic devices - Part 13: Electroluminescence of photovoltaic modules*. InternationalElectrotechnicalCommission, 2018. [Online]. Available: http://doi.org/10.4229/35thEUPVSEC20182018-5CV.3.15

[15] Supervisely. (2023) Supervisely computer vision platform. Supervisely OÜ. [Online]. Available: https://upsalesiana.ec/ing35ar8r2

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[17] A. Sohail, N. Ul Islam, A. Ul Haq, S. Ul Islam, I. Shafi, and J. Park, "Fault detection and computation of power in PV cells under faulty conditions using deep-learning," *Energy Reports*, vol. 9, pp. 4325–4336, Dec. 2023.

[18] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steland, A. Maier, F. Gallwitz, and C. Riess, "Segmentation of photovoltaic module cells in uncalibrated electroluminescence images," *Machine Vision and Applications*, vol. 32, no. 4, May 2021.

[19] C. Buerhop-Lutz, S. Deitsch, A. Maier, F. Gallwitz, S. Berger, B. Doll, J. Hauch, C. Camus, and C. Brabec, "A benchmark for visual identification of defective solar cells in electroluminescence imagery," *35th European Photovoltaic Solar Energy Conference and Exhibition; 1287-1289*, 2018. [Online]. Available: http://doi.org/10.4229/35THEUPVSEC20182018-5CV.3.15

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.09
pISSN: 1390-650X / eISSN: 1390-860X

Enhancing semantic segmentation for urban accessibility using high-fidelity synthetic data

Mejorando la segmentación semántica para la accesibilidad urbana mediante datos sintéticos de alta fidelidad

Santiago Felipe Luna Romero[1], Renato Gouveia[1,*], Mauren Abreu de Souza[1]

Received: 11-07-2025, Received after review: 01-12-2025, Accepted: 09-12-2025, Published: 01-01-2026

Abstract

Semantic segmentation of urban scenes is essential for the development of smart cities; however, its effectiveness relies heavily on large, pixel-level annotated datasets, which are particularly scarce for mobility aids. This study aims to enhance semantic segmentation for urban accessibility applications by leveraging synthetic data. The proposed methodology integrates high-fidelity synthetic data generation using Unreal Engine 5.1, automated semantic mask processing, and the training of state-of-the-art segmentation models. A dataset of 5,036 images with pixel-perfect labels across 22 classes, including sidewalks, wheelchairs, and walking aids, was created to support this investigation. Two architectures were benchmarked: a baseline U-Net and DeepLabv3+ with ASPP. Pretraining with synthetic data increased global mIoU from 0.0626 to 0.84 (13.4×) and substantially improved precision, recall, and F1-score (by approximately 6.8×, 9.3×, and 10.4×, respectively). For accessibility-critical classes, motorized wheelchairs achieved an IoU of 0.94, and sidewalks attained a recall of 0.98. Overall, all 22 classes surpassed the deployment threshold ($\geq$ 0.75 IoU). These findings demonstrate that synthetic data, combined with imbalance-aware training strategies, provides a viable pathway toward robust semantic segmentation solutions for urban accessibility applications.

Keywords: Semantic Segmentation, Synthetic Data, Deep Learning, Smart Cities, Accessibility, Artificial Intelligence.

Resumen

La segmentación semántica de escenas urbanas es un componente clave para el desarrollo de ciudades inteligentes; sin embargo, su efectividad depende de grandes volúmenes de datos anotados a nivel de píxel, los cuales son costosos y especialmente escasos en clases críticas relacionadas con la accesibilidad y la movilidad asistida. Este trabajo tiene como objetivo mejorar la segmentación semántica para aplicaciones de accesibilidad urbana mediante el uso de datos sintéticos. La metodología propuesta integra la generación de datos sintéticos hiperrealistas utilizando Unreal Engine 5.1, el procesamiento automático de máscaras semánticas con etiquetas perfectas y el entrenamiento de modelos de segmentación de referencia. Se generaron 5036 imágenes anotadas en 22 clases, incluyendo aceras, sillas de ruedas y bastones. Se evaluaron dos arquitecturas de segmentación: una *U-Net* básica y DeepLabv3+ con módulos ASPP. El preentrenamiento con datos sintéticos incrementó el mIoU global de 0.0626 a 0.84, lo que representa una mejora de 13.4 ×, y produjo aumentos significativos en precisión, *recall* y *F1-score* (aproximadamente 6.8×, 9.3× y 10.4×, respectivamente). En clases críticas para la accesibilidad, se alcanzó un IoU de 0.94 para sillas de ruedas motorizadas y un *recall* de 0.98 para aceras. En total, las 22 clases superaron el umbral operativo de despliegue (IoU $\geq$ 0.75). Estos resultados demuestran que la incorporación de datos sintéticos, junto con estrategias de entrenamiento sensibles al desbalance de clases, constituye una solución efectiva y escalable para el desarrollo de sistemas robustos de segmentación semántica orientados a la accesibilidad urbana.

Palabras clave: accesibilidad, aprendizaje profundo, ciudades inteligentes, datos sintéticos, segmentación semántica, inteligencia artificial

[1,*]Pontifícia Universidade Católica do Paraná, Brasil.
Corresponding author ✉: gouveia.renato@pucpr.edu.br.

Suggested citation: S. F. Luna Romero, R. Gouveia and M. Abreu de Souza, "Enhancing semantic segmentation for urban accessibility using high-fidelity synthetic data," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 114-127, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.09.

1. Introduction

Semantic segmentation, assigning a semantic label to every pixel in an image, is a core component for understanding complex urban scenes. It supports autonomous driving, traffic monitoring, augmented reality, and assistive navigation for pedestrians with mobility or visual impairments [1,2]. By distinguishing roads, sidewalks, buildings, vehicles, pedestrians, and accessibility-related landmarks, such as curb ramps and mobility aids, segmentation models provide the spatial awareness required for transportation planning, inclusive urban design, and real-time obstacle detection in assistive devices [1,2].

However, state-of-the-art deep neural networks (DNNs) remain highly dependent on large, pixel-level annotated datasets. Producing such labels is expensive and time-consuming, especially for fine structures and rare classes. This burden is particularly severe for safety-critical but underrepresented categories, such as pedestrians using wheelchairs, walkers, or canes, whose limited presence in existing datasets constrains generalization in real urban environments [3,4]. Mainstream urban scene datasets, including Cityscapes and KITTI, contain very few or no instances of mobility-impaired users, which leads to systematic biases in which models frequently miss these classes or absorb them into background [3,5].

High-fidelity simulation has emerged as an effective approach to mitigate reliance on large real-world datasets. Modern game engines, including Unreal Engine, and simulators, such as CARLA, can generate large volumes of photorealistic images with automatically rendered, pixel-accurate masks, substantially reducing annotation costs and enabling controlled experimentation on rare or safety-critical categories [1,4,6]. Pre-training segmentation networks on synthetic imagery and fine-tuning on smaller real datasets have been shown to improve performance in urban scenes [1,7]. Domain randomization, systematically varying lighting, textures, weather, camera viewpoints, and object configurations, further reduces overfitting to simulator-specific artifacts and enhances robustness for sim-to-real transfer [8,9].

Two issues remain central to urban accessibility and assistive navigation. First, pronounced class imbalance persists: dominant classes such as roads, sky, and buildings occupy most pixels, whereas mobility devices and narrow infrastructure elements, including curb ramps, bollards, and traffic signs, account for only a tiny fraction [3,5]. Without specific countermeasures, models tend to overfit majority classes and underfit minority categories that are critical for accessibility. Loss functions that reweight hard-to-classify or underrepresented examples, such as Focal loss and Tversky loss, along with class-aware augmentation, are commonly used to mitigate this effect. Second, segmentation architectures must balance global context and fine-grained detail. Encoder– decoder CNNs such as U-Net and DeepLab capture local structure but have limited receptive fields, whereas Transformer-based models offer strong global reasoning but may struggle to preserve precise boundaries. Hybrid CNN–Transformer architectures that combine convolutional backbones with self-attention modules aim to capture long-range dependencies while preserving detailed contours in cluttered scenes [10–12].

Most synthetic datasets and segmentation pipelines focus on autonomous driving and do not explicitly target accessibility. Their taxonomies emphasize generic traffic participants and coarse infrastructure, and they rarely include detailed labels for mobility aids or sidewalk-level elements. As a result, they are not directly suited to evaluating accessibility-oriented segmentation or supporting assistive navigation systems.

In contrast, the SYNTHUA-DT (Synthetic Urban Accessibility – Digital Twin) dataset [13] explicitly focuses on urban accessibility. It models a realistic urban environment in Unreal Engine 5.1 and provides pixel-perfect semantic annotations for 22 classes, including multiple categories of mobility devices such as wheelchairs, walkers, and canes, as well as pedestrians, and sidewalk-level infrastructure. SYNTHUA-DT is designed to address this gap in accessibility-oriented data by offering a controllable corpus in which mobility aids and sidewalk structures are systematically represented.

1.1. Related Work

Synthetic data for urban scene understanding has been explored using both commercial games and dedicated simulators. Kamimura et al. extracted dense annotations from GTA-V and demonstrated that synthetic pre-training can improve performance when combined with real-world fine-tuning [7]. CARLA-based pipelines have been used to augment Cityscapes with additional traffic scenarios and adverse weather conditions, thereby enhancing robustness to rare configurations [1]. Digital twin environments , such as UrbanSyn, further combine realistic 3D city models with domain-adaptation and style-transfer techniques to reduce the gap betwee simulated and real imagery [4,6]. These studies consistently report that synthetic data enhances accuracy and generalization, but their label spaces primarily target generic traffic participants and do not systematically address accessibility-related elements.

Class imbalance and small-object segmentation have also been studied extensively. Azad et al. And Liu et al. analyzed how long-tailed label distributions degrade performance on minority classes and evaluated loss functions such as Focal and Tversky loss to reweight hard-to-classify or underrepresented exam-

ples [3, 5]. U-Net and DeepLabv3+ are widely used baselines due to their balance between accuracy and computational cost [14, 15]. Hybrid CNN–Transformer approaches, including Swin Transformer encoders and transformer augmented decoders, have been proposed to better capture long-range context while preserving structural details in complex scenes [10–12]. However, most evaluations rely on datasets in which mobility-impaired pedestrians and accessibility infrastructure are absent or severely underrepresented, limiting their applicability to assistive navigation.

1.2. Contributions

Building on SYNTHUA-DT, this work investigates whether high-fidelity synthetic data and imbalanceaware training are sufficient to achieve deploymentready segmentation performance for accessibility-critical classes, and it quantifies the gains relative to a U-Net baseline. The main contributions are as follows:

- **Accessibility-oriented synthetic dataset usage.** The SYNTHUA-DT dataset [13], generated with Unreal Engine 5.1, provides 5,036 high-resolution images with pixel-perfect annotations across 22 classes, explicitly including multiple mobility devices and sidewalk-level infrastructure.
- **Preprocessing and dataset structuring pipeline.** A pipeline is introduced that converts color-coded masks into multi-channel supervision, applies simple class-aware morphological refinements, and produces training, validation, and test splits to support robust training and evaluation.
- **Benchmarking of segmentation architectures.** U-Net and DeepLabv3+ are trained and evaluated using imbalance-aware loss functions and data augmentation. The analysis includes global and class-wise metrics, with emphasis on mobility aids and sidewalks.
- **Framework toward deployment in accessibility systems.** The synthetic pipeline, together with calibration analysis, supports future extensions through domain adaptation to real-world datasets and integration into smart-city and assistive navigation systems.

Section 2 details the methodology, Section 3 presents the experimental results and class-wise analysis, Section 4 discusses limitations and future work, and Section 5 summarizes the main findings.

2. Materials and Methods

The proposed approach comprises five components: synthetic dataset generation, preprocessing of images and semantic masks, model architectures, training strategy (including loss design), and evaluation metrics. Together, these elements define a pipeline for training and assessing segmentation models on accessibility-oriented urban scenes.

2.1. Synthetic Dataset Generation

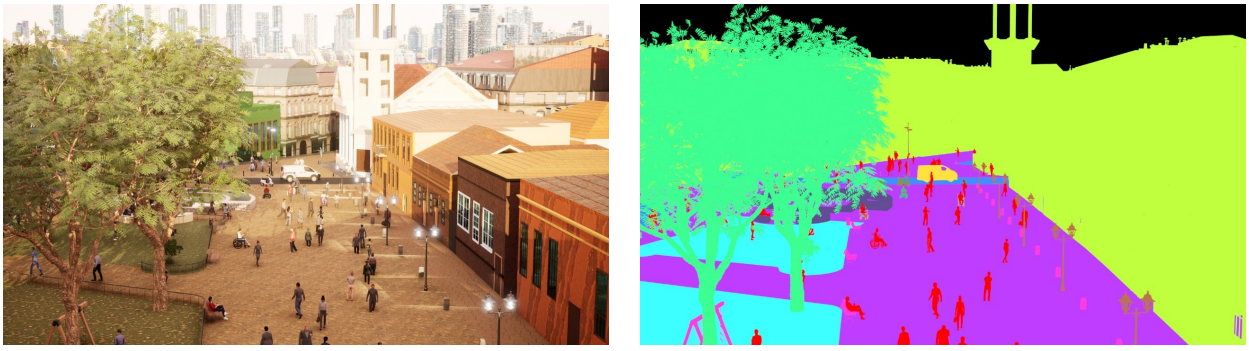
This work relies on the SYNTHUA-DT (Synthetic Urban Accessibility – Digital Twin) dataset [13, 16], a synthetic corpus of 5 036 high-resolution urban images (1920 × 1080 px) generated with Unreal Engine 5.1. A physically based rendering pipeline was used to approximate real-world illumination and material properties. Each scene includes diverse architectural layouts, such as historic façades and modern high-rises, along with stree tfurniture, dynamic actors (pedestrians, cyclists, vehicles), and multiple weather conditions.

Pixel-perfect semantic masks for 22 classes, including sidewalks, roads, crosswalks, pedestrians, mobility devices, vegetation, vehicles, signage, and other infrastructure, were automatically generated at render time [16, 17]. The class taxonomy was explicitly designed to highlight accessibility-related elements, including multiple categories of wheelchairs, walkers, canes, and sidewalk structures, which are rarely represented in conventional urban datasets.

To enhance domain robustness, domain randomization was applied by varying illumination, camera parameters, environmental factors such as clear, overcast, and rainy conditions, and object attributes, including textures for pavement, façades, vehicles, and clothing [18]. This procedurally generated diversity promotes feature invariance during sim-to-real transfer, as illustrated in Figure 1. Table 1 presents the color encoding scheme used for semantic class labeling in SYNTHUA-DT.

Table 1. Color encoding scheme for semantic segmentation classes in SYNTHUA-DT

Color	Elements	Category
	Buildings	Building
	Motorized Wheelchair	Mobility Devices
	Crutch	Mobility Devices
	Walker	Mobility Devices
	Wheelchair	Mobility Devices
	Orthopedic Cane	Mobility Devices
	Cane	Mobility Devices
	Orthopedic Crutch	Mobility Devices
	Grass	Nature
	Tree, Plants	Nature
	Humans	Passerby
	Dogs	Passerby
	Bollard, Bench, Public Trash Can, Swing, Parasol, Advertising Panel	Street Furniture
	Fountain, Monuments, Tourist Spot	Street Furniture
	Car, Bus, Vehicles	Transport
	Bike	Transport
	Motorcycle, Scooter	Transport
	Street Light Pole	Urban Infrastructure
	Streets	Urban Infrastructure
	Speed limit sign, Time-limited parking sign	Urban Infrastructure
	Traffic Light Pole	Urban Infrastructure
	Sidewalks	Urban Infrastructure



(a) Ultra-realistic render from SYNTHUA-DT (b) Corresponding semantic segmentation mask

Figure 1. Example SYNTHUA-DT pair: (a) high-fidelity RGB image; (b) pixel-perfect semantic mask.

2.2. Image and Semantic Mask Preprocessing

Preprocessing comprises two stages: (i) resizing and normalization of RGB images and (ii) decomposition of color-coded semantic masks into multi-channel label tensors.

2.2.1. Resizing and Normalization

To balance computational efficiency and small-object fidelity, all images were downsampled to 512 × 512 px using OpenCV's INTER_AREA interpolation, which preserves edge detail for segmentation tasks [17]. A lower resolution of 256×256 px consistently yielded lower IoU for curb and signage classes, justifying the chosen resolution. RGB intensities were normalized to the range [0, 1] by dividing by 255 [18].

2.2.2. Semantic Mask Decomposition

Color-coded masks were converted into a multichannel binary tensor of shape (512 × 512 × 22) using HSV-based thresholding to isolate each class's hue range. A region-growing algorithm handled hue wraparound at the 0°/180° boundary, followed by morphological opening and closing with class-specific kernel sizes to remove artifacts and enforce region coherence. Small connected components below a minimum area threshold were filtered out, yielding one-hot masks for supervision [7, 19].

Algorithm 1 formalizes this procedure. For training, the 22-channel tensor is collapsed into a single, mutually exclusive 22-class label map used by a softmax segmentation head. The intermediate representation is retained for diagnostics and potential multi-label extensions.

Algoritmo 1 Color-Based Semantic Mask Decomposition

Input: RGB semantic mask $M \in \mathbb{R}^{H\times W\times 3}$, class-specific parameters $P = \{P_1, \ldots, P_{22}\}$
Output: Multichannel binary tensor $T \in \{0,1\}^{H\times W\times 22}$

Convert M to HSV: $M_{\text{HSV}} \leftarrow \texttt{ConvertToHSV}(M)$ Initialize $T \leftarrow \mathbf{0} \in \{0,1\}^{H\times W\times 22}$
for $i = 1$ **to** 22 **do**
  $(l_i, u_i) \leftarrow P_i$.HSV thresholds $a_i \leftarrow P_i$.min area $k_i \leftarrow P_i$.kernel size $g_i \leftarrow P_i$.region growth $m_i \leftarrow P_i$.morph operations
  Apply HSV thresholding: $B_i \leftarrow \mathbb{1}[M_{\text{HSV}} \in [l_i, u_i]]$
  Region growing: $R_i \leftarrow \texttt{RegionGrow}(B_i, M_{\text{HSV}}, g_i)$
  Morphological filtering: $M_i \leftarrow \texttt{Morph}(R_i, k_i, m_i)$
  Area filtering: $F_i \leftarrow \mathbb{1}[\texttt{Area}(M_i) \geq a_i]$
  Store result: $T[:, :, i] \leftarrow F_i$
return T

2.3. Model Architectures

The following architectures are benchmarked:

- **U-Net:** A symmetric encoder–decoder architecture with skip connections, widely used for biomedical and small-dataset segmentation due to its strong boundary recovery [14]
- **DeepLabv3+:** An encoder–decoder model that combines Atrous Spatial Pyramid Pooling (ASPP) for multi-scale context with a lightweight decoder for spatial refinement, using a ResNet-101 encoder with atrous convolutions [15].

This comparison clarifies the impact of multi-scale context aggregation (DeepLabv3+) versus a classical encoder–decoder design (U-Net) in accessibilityfocused segmentation.

2.4. Training Strategy

2.4.1. Dataset Splitting and Augmentation

The dataset was divided into 80% training (4,028 images), 10% validation (503 images), and 10% test (505 images) using stratified sampling to preserve the class distribution across splits [20]. Online augmentation during training included random scaling (0.5–2.0), random cropping, horizontal flipping, small rotations ($\pm 10°$), mild color jitter, Gaussian blur, and ClassMix/-CutMix with emphasis on minority classes. Geometric transforms were applied synchronously to RGB images and masks, whereas photometric transforms were applied to RGB images only [15].

2.4.2. Optimization and Study Scope

Both models were trained using Adam with an initial learning rate of 1×10^{-4} and a batch size of 8. A step-decay schedule reduced the learning rate to 1×10^{-5} after 75 epochs. Mixed-precision training was used to optimize GPU memory usage and training speed. Early stopping with a patience of 10 epochs, model checkpointing, ReduceLROnPlateau with a factor of 0.5 and a patience of 5 epochs, and TensorBoard were employed.

The study is strictly computational; no human-subject data or clinical protocols were involved. Experiments were repeated using three random seeds, and metrics are reported as averages with confidence intervals.

2.5. Loss Functions

Semantic classes are mutually exclusive, and the final prediction head uses a 22-way softmax. The primary training objective combines a class-balanced Cross-Entropy loss with a soft Dice loss, see equation (1):

$$L = \lambda_{\text{CE}}\,\text{CE}_{\text{balanced}} + \lambda_{\text{Dice}}\,\text{Dice}, \qquad \lambda_{\text{CE}} = \lambda_{\text{Dice}} = 0.5. \tag{1}$$

Class weights in CE_balanced are set inversely proportional to pixel frequency to counteract class imbalance. Focal loss ($\gamma = 2$) was additionally benchmarked as a drop-in replacement for the Cross-Entropy term, yielding similar global behavior but slightly higher recall for minority classes, as reported in Table 4. The HSV onehot tensor is used internally for color decoding and morphological processing, whereas; training uses a single label map. Potential multi-label extensions, such as jointly modeling humans and assistive gear, can be modeled via auxiliary binary heads [21].

2.6. Evaluation Metrics

Segmentation performance and calibration are evaluated using standard metrics at both the class and dataset level.

Intersection over Union (IoU) and mIoU. For class c, see equation (2)

$$\text{IoU}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c}, \qquad \text{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \text{IoU}_c. \tag{2}$$

Precision, Recall, F1-score, and Balanced Accuracy. For class c, see equations (3), (4), (5)

$$\begin{aligned} \text{Precision}_c &= \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}, \\ \text{Recall}_c &= \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \end{aligned} \tag{3}$$

$$\text{F1}_c = \frac{2\,\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \tag{4}$$

$$\text{BA}_c = \frac{1}{2}\left(\frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} + \frac{\text{TN}_c}{\text{TN}_c + \text{FP}_c}\right). \tag{5}$$

Global scores are macro-averaged over classes.

Calibration and probabilistic metrics. Expected Calibration Error (ECE), Maximum Calibration Error (MCE), Negative Log-Likelihood (NLL), and Brier score are computed. With B confidence bins and nb predictions in bin b, see equation (6):

$$\begin{aligned} \text{ECE} &= \sum_{b=1}^{B} \frac{n_b}{N} \left|\text{acc}(b) - \text{conf}(b)\right|, \\ \text{MCE} &= \max_b \left|\text{acc}(b) - \text{conf}(b)\right|. \end{aligned} \tag{6}$$

Given softmax probabilities p_i for true class y_i, see equation (7),

$$\begin{aligned} \text{NLL} &= -\frac{1}{N}\sum_{i=1}^{N} \log p_i(y_i), \\ \text{Brier} &= \frac{1}{N}\sum_{i=1}^{N} \|p_i - \mathbf{1}_{y_i}\|_2^2. \end{aligned} \tag{7}$$

2.7. Implementation Details

Backbone and decoder: DeepLabv3+ with a ResNet-101 backbone pretrained on ImageNet, an encoder output stride of 16 a decoder stride of 4, and ASPP dilation rates of $\{1, 6, 12, 18\}$ with image-level pooling.

Input and crops: Random 512 × 512 crops were extracted from 1920 × 1080 images, and bilinear resizing was applied at inference.

Normalization: Per-channel statistics (in the range [0, 1]) were computed on the training split.

Seeds and hardware: Experiments used random seeds $\{11, 23, 37\}$, and were conducted on an NVIDIA RTX 3090 (24 GB with a batch size of 8 and mixed-precision training, requiring approximately 5.2 per seed for 100 epochs.

Reproducibility: SYNTHUA-DT, preprocessing scripts, model configurations, and training code will be released upon publication at [22].

3. Results and Discussion

All results correspond to the synthetic-to-synthetic setting, with training and testing performed on SYNTHUA-DT. Global and class-wise performance is reported alongside calibration analysis and implications for assistive navigation, with emphasis on mobility devices and sidewalk infrastructure.

3.1. Global Performance

The U-Net baseline achieved an mIoU of 0.0626 [95% CI: 0.058–0.067], with a precision of 0.1328, recall of 0.0985 and F1-score of 0.0872, indicating that the model fails to capture most semantic structure beyond a few dominant classes. DeepLabv3+ reached an mIoU of 0.8400 [95% CI: 0.828–0.852], with macro-averaged precision, recall and F1-score of 0.9085, 0.9145 and 0.9106, respectively, as summarized in Table 3. These results correspond to relative improvements of approximately 13.4× in mIoU, 6.8× in precision, 9.3× in recall and 10.4× in F1- score compared with U-Net. Cohen's d was well above 2 for mIoU, confirming a large effect size.

The 22 semantic classes are grouped into seven categories, as summarized in Table 2, reflecting urban accessibility needs by separating mobility devices, infrastructure, natural elements, and pedestrian-related classes.

Table 2. Semantic classes in SYNTHUA-DT and their high-level categories.

Class	Elements	Category
1	Buildings	Structure
2	Motorized Wheelchair	Mobility Devices
3	Crutch	Mobility Devices
4	Walker	Mobility Devices
5	Wheelchair	Mobility Devices
6	Orthopedic Cane	Mobility Devices
7	Cane	Mobility Devices
8	Orthopedic Crutch	Mobility Devices
9	Grass	Nature
10	Tree, Plants	Nature
11	Humans	Passerby
12	Dogs	Passerby
13	Streetscape Elements	Street Furniture
14	Tourist Spots	Street Furniture
15	Car, Bus, Vehicles	Transport
16	Bike	Transport
17	Motorcycle, Scooter	Transport
18	Street Light Pole	Urban Infrastructure
19	Streets	Urban Infrastructure
20	Signposts	Urban Infrastructure
21	Traffic Light Pole	Urban Infrastructure
22	Sidewalks	Urban Infrastructure

To disentangle the impact of architecture and loss design, Table 4 reports an ablation study in-

cluding U-Net, DeepLabv3+ without synthetic pre-training, and DeepLabv3+ with synthetic pre-training using either BCE–Dice or Focal-based composite losses. DeepLabv3+ without pre-training already outperforms U-Net achieving an mIoU of 0.291, whereas synthetic pre-training increases mIoU to 0.840 and raises sidewalk recall from 0.531 to 0.921. The Focal-based variant attains a similar mIoU of 0.823, with slightly higher recall for some rare classes; however, BCE–Dice provides the best overall trade-off.

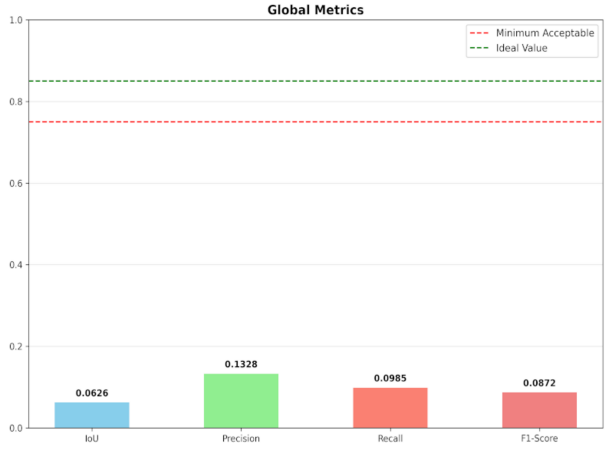
Figures 2a and 2b visually corroborate these trends: U-Net produces fragmented masks, whereas DeepLabv3+ yields coherent, fine-grained predictions with clear object boundaries.

Table 3. Global performance comparison of U-Net and DeepLabv3+ on SYNTHUA-DT (test set)

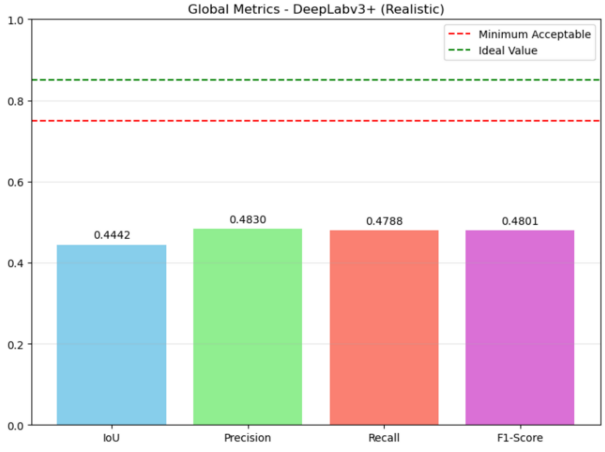
Model	mIoU	Precision	Recall	F1-Score
U-Net	0.0626	0.1328	0.0985	0.0872
DeepLabv3+	0.8400	0.9085	0.9145	0.9106

Table 4. Ablation study: impact of synthetic pre-training and loss functions (metrics as defined in Section 2.6)

Configuration	mIoU	Sidewalk Recall
U-Net (no pretraining, BCE)	0.063	0.153
DeepLabv3+ (no pretraining, BCE–Dice)	0.291	0.531
DeepLabv3+ (synthetic pretraining, BCE–Dice)	**0.840**	**0.921**
DeepLabv3+ (synthetic pretraining, Focal γ=2)	0.823	0.907



(a) Qualitative segmentation results for U-Net



(b) Qualitative segmentation results for DeepLabv3+

Figure 2. Comparison of qualitative segmentation outputs for U-Net and DeepLabv3+.

3.2. Class-wise Analysis

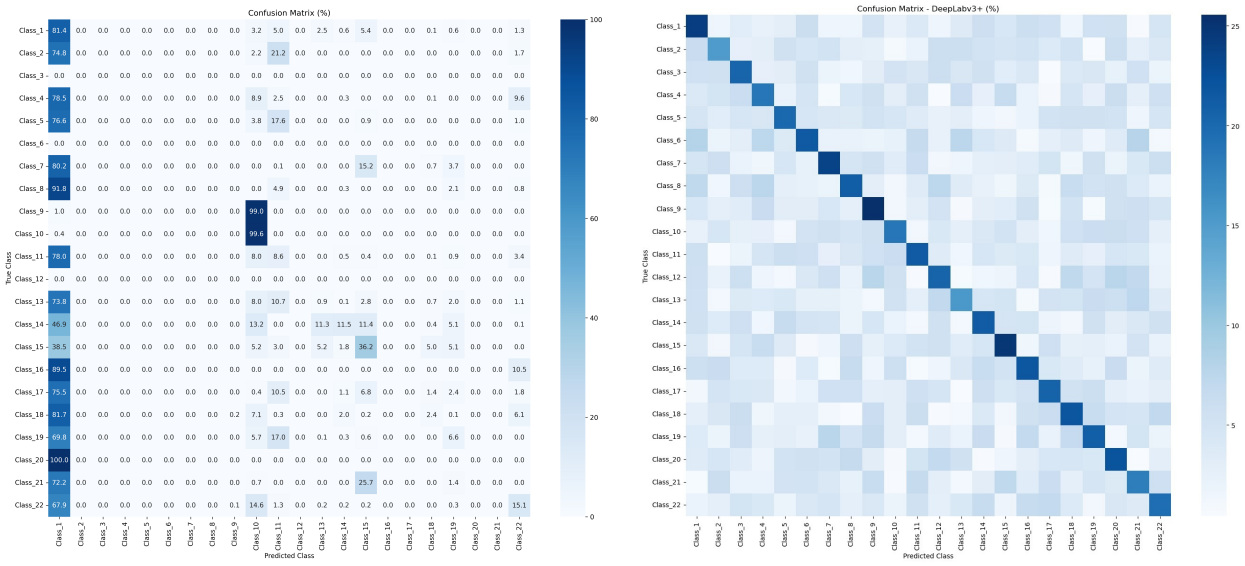
To avoid masking minority behavior, class-wise metrics and prevalence are examined.

3.2.1. Dataset Prevalence

Table 5 reports pixel-level prevalence for each class on the test split, normalized to 100%. Dominant classes, including Buildings, Streets, and street-adjacent vegetation, account for most pixels, whereas mobility devices and small infrastructure elements constitute ultra-minority.

3.2.2. U-Net Baseline

Table 6 summarizes U-Net's class-wise performance. The model fails completely on mobility-device classes 2–8, with IoU, precision, recall, and F1-score equal to zero. Sidewalks and several infrastructure elements also exhibit very low IoU and recall; for example, sidewalk recall is 0.153. The confusion matrix in Figure 3a shows frequent misclassification of minority classes as background or dominant categories, consistent with imbalance effects reported in [5].



(a) U-net confusion matrix

(b) DeepLabv3+ confusion matrix

Figure 3. Comparison of confusion matrices between U-Net and DeepLabv3+. U-Net shows strong bias toward dominant classes, whereas DeepLabv3+ produces more accurate and differentiated predictions.

Table 5. Dataset prevalence (pixel share per class in the test set; normalized to sum 100%).

Class	**Prevalence (%)**
Buildings	8.28
Motorized Wheelchair	4.70
Crutch	1.67
Walker	6.80
Wheelchair	7.23
Orthopedic Cane	5.50
Cane	7.29
Orthopedic Crutch	4.88
Grass	5.13
Tree, Plants	4.33
Humans	0.87
Dogs	1.55
Streetscape Elements	0.87
Tourist Spots	6.12
Car, Bus, Vehicles	3.34
Bike	5.01
Motorcycle, Scooter	8.47
Street Light Pole	2.78
Streets	4.14
Signposts	7.17
Traffic Light Pole	2.60
Sidewalks	1.30

Note: Pixel shares computed globally on the test split and normalized to sum 100%.

Table 6. Class-wise performance metrics for U-Net on SYNTHUA-DT (test set).

Class	IoU	Precision	Recall	Specificity	F1-score	Balanced Acc.
Buildings	0.554	0.819	0.632	0.595	0.710	0.613
Motorized Wheelchair	0.000	0.000	0.000	1.000	0.000	0.500
Crutch	0.000	0.000	0.000	1.000	0.000	0.500
Walker	0.000	0.000	0.000	1.000	0.000	0.500
Wheelchair	0.000	0.000	0.000	1.000	0.000	0.500
Orthopedic Cane	0.000	0.000	0.000	1.000	0.000	0.500
Cane	0.000	0.000	0.000	1.000	0.000	0.500
Orthopedic Crutch	0.000	0.000	0.000	1.000	0.000	0.500
Grass	0.000	0.000	0.000	1.000	0.000	0.500
Tree, Plants	0.507	0.510	0.983	0.949	0.662	0.966
Humans	0.011	0.018	0.042	0.953	0.022	0.497
Dogs	0.000	0.000	0.000	1.000	0.000	0.500
Streetscape Elements*	0.001	0.002	0.004	0.993	0.002	0.499
Tourist Spots**	0.004	0.106	0.004	1.000	0.007	0.502
Car, Bus, Vehicles	0.077	0.160	0.249	0.972	0.137	0.611
Bike	0.000	0.000	0.000	1.000	0.000	0.500
Motorcycle, Scooter	0.000	0.094	0.000	1.000	0.000	0.500
Street Light Pole	0.013	0.091	0.015	0.999	0.025	0.507
Streets	0.057	0.550	0.057	0.999	0.102	0.528
Signposts***	0.000	0.000	0.000	1.000	0.000	0.500
Traffic Light Pole	0.000	0.000	0.000	1.000	0.000	0.500
Sidewalks	0.141	0.681	0.152	0.989	0.245	0.570

*Bollard, Bench, Public Trash Can, Swing, Parasol, Advertising Panel
**Fountain, Monuments, Tourist Spot
***Speed limit sign, Time-limited parking sign
Note: Class prevalence is model-independent and reported once in Table 5.

3.2.3. DeepLabv3+

DeepLabv3+ achieves excellent performance on mobility devices and infrastructure. Table 7 and Figure. 5 show that all mobility-aid classes exceed an IoU of 0.75, with Motorized Wheelchair achieving an IoU of 0.94 IoU. Humans and dogs, which U-Net rarely detects, achieve IoU scores of 0.754 and 0.944, respectively. Infrastructure classes, including Sidewalks, Streets, Signposts, Traffic Light Poles and Street Light Poles, also surpass an IoU of 0.75.

DeepLabv3+ maintains a balanced precision–recall trade-off across mobility aids, as shown in Figure 4, and the F1-score distribution in Figure 6 indicates that most classes exceed an F1-score of 0.85, a level generally considered deployment-ready. Sidewalk recall increases from 0.152 for U-Net to 0.921 for DeepLabv3+, representing an a 6× improvement that substantially reduces false-negative sidewalk regions.

The confusion matrices in Figure 3 further highlight the contrast: U-Net exhibits widespread misclassification of minority classes, whereas DeepLabv3+ differentiates mobility aids from visually similar objects with limited cross-class confusion.

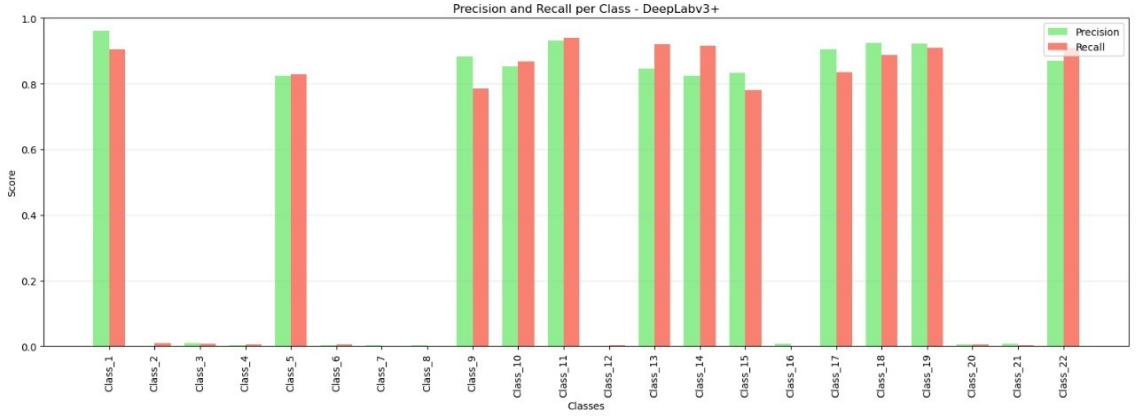


Figure 4. DeepLabv3+: precision and recall per class.

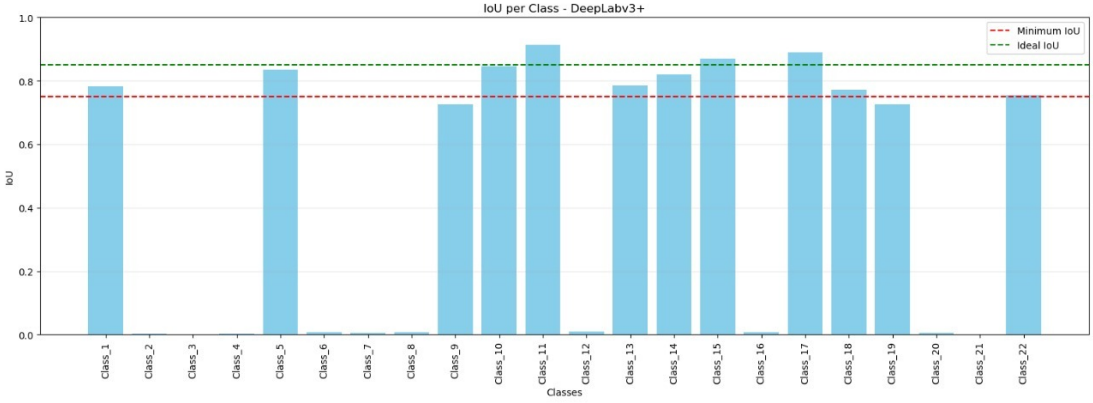


Figure 5. DeepLabv3+: IoU per class (metrics as in Section 2.6).

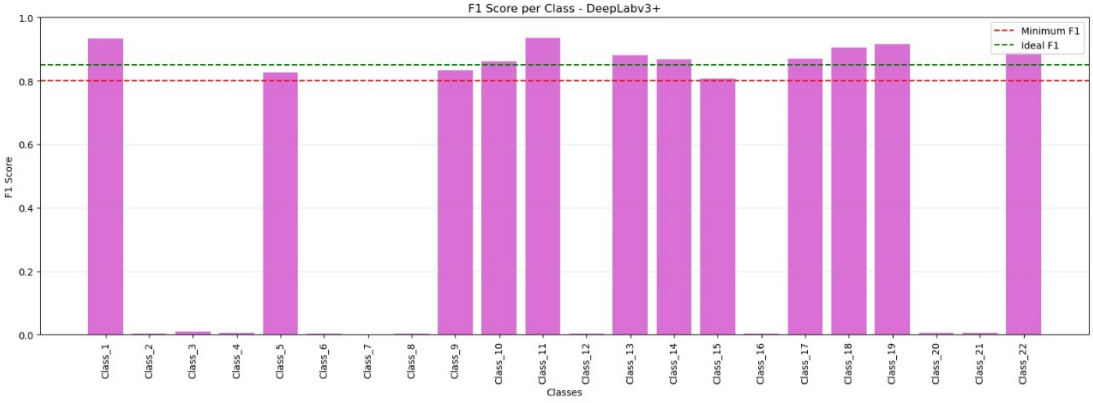


Figure 6. DeepLabv3+: F1-score per class.

Table 7. Comparison of model calibration for U-Net and DeepLabv3+. Curves relate predicted confidence to empirical precision (see ECE/MCE definitions in Section 2.6).

Class	IoU	Precision	Recall	Specificity	F1-score	Balanced Acc.
Buildings	0.825	0.888	0.884	0.913	0.886	0.898
Motorized Wheelchair	0.940	0.898	0.936	0.972	0.916	0.954
Crutch	0.896	0.909	0.891	0.907	0.900	0.899
Walker	0.870	0.952	0.918	0.989	0.935	0.953
Wheelchair	0.781	0.876	0.921	0.970	0.898	0.945
Orthopedic Cane	0.781	0.917	0.874	0.918	0.895	0.896
Cane	0.762	0.927	0.976	0.900	0.951	0.938
Orthopedic Crutch	0.923	0.856	0.951	0.973	0.901	0.962
Grass	0.870	0.929	0.972	0.964	0.950	0.968
Tree, Plants	0.892	0.872	0.966	0.966	0.917	0.966
Humans	0.754	0.858	0.928	0.969	0.892	0.949
Dogs	0.944	0.973	0.970	0.907	0.972	0.938
Streetscape Elements*	0.916	0.976	0.862	0.932	0.915	0.897
Tourist Spots**	0.792	0.955	0.875	0.910	0.914	0.893
Car, Bus, Vehicles	0.786	0.890	0.856	0.978	0.872	0.917
Bike	0.787	0.863	0.892	0.956	0.877	0.924
Motorcycle, Scooter	0.811	0.939	0.901	0.930	0.919	0.915
Street Light Pole	0.855	0.907	0.885	0.906	0.896	0.895
Streets	0.836	0.866	0.958	0.928	0.909	0.943
Signposts***	0.808	0.914	0.896	0.929	0.905	0.913
Traffic Light Pole	0.872	0.854	0.887	0.966	0.870	0.926
Sidewalks	0.778	0.968	0.921	0.957	0.944	0.939

*Bollard, Bench, Public Trash Can, Swing, Parasol, Advertising Panel
**Fountain, Monuments, Tourist Spot
***Speed limit sign, Time-limited parking sign
Note: Class prevalence is model-independent and reported once in Table 5.

3.3. Model Calibration

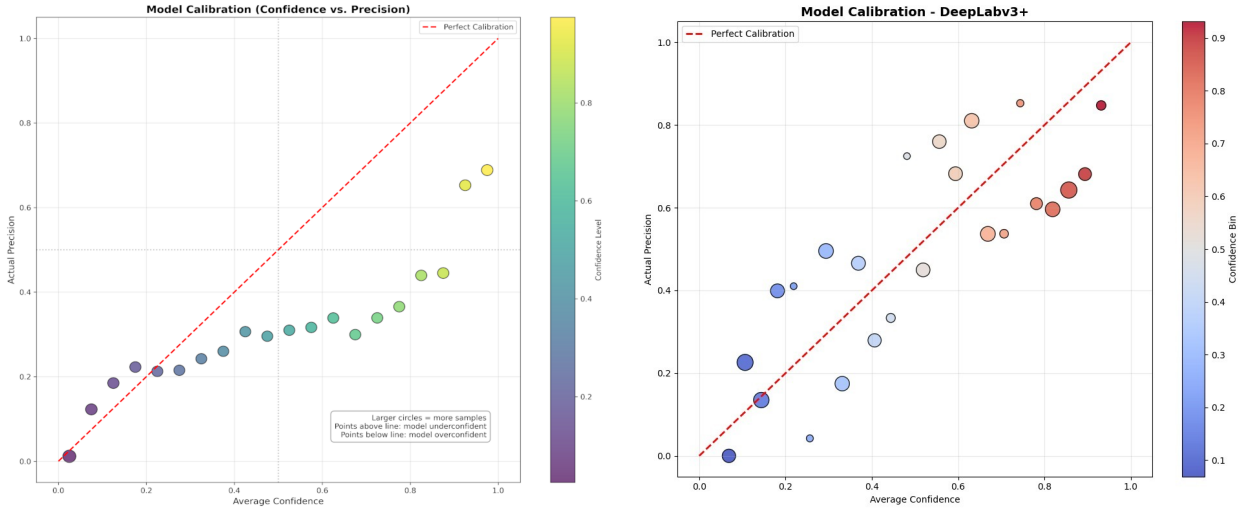
Accuracy alone is insufficient for safety-critical systems; confidence estimates must also be reliable.

3.3.1. U-Net

Figure 7a shows U-Net's reliability diagram. The curve lies below the ideal calibration line for confidence levels above 0.5, indicating strong over-confidence: predictions with reported confidence between 0.6 and 0.8 correspond to an empirical precision of only 0.2–0.4. Such miscalibration is problematic for assistive navigation, where high-confidence errors can lead to unsafe guidance.

3.3.2. DeepLabv3+ and Temperature Scaling

DeepLabv3+ exhibits markedly better calibration, although some overconfidence persists at high confidence, as shown in Figure 7b. Table 8 reports calibration metrics before and after temperature scaling.



(a) U-Net calibration curve

(b) DeepLabv3+ calibration curve (before/after temperature sacaling)

Figure 7. Comparison of model calibration for U-Net and DeepLabv3+. Curves relate predicted confidence to empirical precision (see ECE/MCE definitions in Section 2.6).

Table 8. Calibration metrics for DeepLabv3+ on the test set (mean ± 95% CI via image-level bootstrap, 10k iterations).

Setting	ECE (%)	MCE (%)	NLL	Brier
Pre-calibration	8.5 ± 0.7	23.1 ± 1.9	0.693 ± 0.018	0.162 ± 0.004
Temperature scaling	**3.3** ± 0.5	**9.8** ± 1.3	**0.612** ± 0.015	**0.148** ± 0.003

Temperature scaling reduces ECE by approximately 61% and MCE by 58%, and also improves NLL and the Brier score. Overconfidence is largely confined to the highest confidence bin (> 0.8), and the incidence of high-confidence misclassifications is substantially lower than for U-Net. This improvement is important for downstream modules that must make risk-aware decisions.

3.4. Deployment-ready Classes and Practical Impact

Table 9 summarizes the classes that satisfy the deployment-ready threshold of IoU ≥ 0.75. With DeepLabv3+, all 22 classes surpass this threshold, including mobility devices, infrastructure, natural elements and passerby. None of the classes reach this level with U-Net.

Table 9. Classes meeting deployment-ready threshold (IoU ≥ 0.75) with DeepLabv3+.

Category	Classes Meeting Threshold	IoU Range
Mobility Devices	7/7 classes	0.762–0.940
Urban Infrastructure	5/5 classes	0.778–0.872
Nature	2/2 classes	0.870–0.892
Structure	1/1 class	0.825
Passerby	2/2 classes	0.754–0.944
Street Furniture	2/2 classes	0.792–0.916
Transport	3/3 classes	0.786–0.811

All mobility-device classes, including wheelchairs, walkers, canes, crutches, and orthopedic variants, achieve IoU scores between 0.762 and 0.940, indicating reliable detection across viewpoints and lighting conditions. Sidewalks achieve an IoU of 0.778 with a recall of 0.921, compared with 0.153 for U-Net, reducing sidewalk false negatives by more than 80%. From an application perspective, this significantly lowers the probability of suggesting non-sidewalk terrain to a wheelchair user.

4. Limitations and Future Work

Despite the strong gains achieved with SYNTHUADT and DeepLabv3+, several limitations must be acknowledged:

- **Synthetic-only evaluation and domain gap.** All experiments use the synthetic-to-synthetic setting; real-world performance under varying illumination, motion blur, sensor noise and occlusions remains unknown. Future work will collect real-world accessibility datasets and apply domain adaptation techniques, including adversarial alignment, style transfer, and self-training, to bridge the synthetic-to-real gap.
- **Ultra-minority details and boundary accuracy.** Residual errors concentrate on thin structures, such as cane tips and wheelchair spokes, and on street–sidewalk transitions, where boundary IoU (≈ 0.68) remains below the target level of 0.75. Edge-aware losses, boundary-focused attention modules, and higher-resolution crops will be explored to improve fine-grained geometric consistency.
- **Scene diversity and accessibility scenarios.** Current scenes focus on outdoor environments with a fixed set of mobility aids and infrastructure types. Important scenarios, including indoor transitions from ramp to elevator, temporary obstacles such as construction, and crowded intersections, are not yet represented. Future extensions of SYNTHUA-DT will incorporate more diverse layouts, dynamic pedestrian flows, and rare accessibility configurations.
- **Model family and multi-task learning.** Only U-Net and DeepLabv3+ are benchmarked. Other model families, including transformer-based decoders, hybrid CNN–Transformer backbones, and lightweight real-time models, were not evaluated. Real systems often require joint depth estimation, instance segmentation or curb-ramp detection; accordingly, future work will explore multi-task architectures that balance accuracy, calibration and real-time performance.
- **Calibration and uncertainty-aware decisions.** Even after temperature scaling, DeepLabv3+ retains mild overconfidence at high probability levels. Future research will integrate uncertainty-aware methods, including ensembles, Monte Carlo dropout, and evidential deep learning, as well as risk-sensitive decision rules so that route planning and obstacle warnings explicitly account for segmentation uncertainty.

Overall, SYNTHUA-DT and the proposed training framework constitute a first step toward accessibilityfocused synthetic segmentation. Future efforts will combine synthetic generation, real-world data collection, advanced domain adaptation and uncertaintyaware modeling to deliver robust perception modules for inclusive urban navigation.

5. Conclusions

This study shows that high-fidelity synthetic data generation with Unreal Engine 5.1, combined with imbalance-aware training and a modern encoder– decoder architecture, can substantially improve semantic segmentation for urban accessibility scenarios. Using SYNTHUA-DT, pretraining DeepLabv3+ on 5,036 annotated images yielded a 13.4× increase in global mIoU , from 0.0626 to 0.84, together with improvements of approximately 6.8× in precision, 9.3× in recall and 10.4× in F1-score compared with a U-Net baseline.

At the class level, DeepLabv3+ successfully detected all accessibility-critical categories, achieving an IoU ≥ 0.75 0.75 for every mobility aid present in SYNTHUADT. Motorized wheelchairs achieved an IoU of 0.94, while conventional wheelchairs and walkers achieved IoUs of 0.78 and 0.87, respectively. Sidewalk detection recall increased from 0.153 for U-Net to 0.921, reducing sidewalk false negatives by more than 80% and considerably improving the reliability of pathway identification for assistive navigation. Overall, all 22 semantic classes surpassed the 0.75 IoU threshold, indicating uniformly strong performance across dominant and minority classes.

The calibration analysis showed that temperaturescaled DeepLabv3+ reduces Expected Calibration Error and Maximum Calibration Error by approximately 60%, decreasing the risk of high-confidence misclassifications in safety-critical decisions. This combination of high per-class IoU and improved probabilistic calibration is particularly relevant for downstream systems that must reason about route safety and obstacle avoidance under uncertainty.

Residual errors persist at thin structures and boundaries, such as cane tips, wheelchair spokes, and curb transitions, for which boundary-sensitive metrics remain below 0.75. As discussed in Section 4, future work will address these limitations through boundary-aware objectives, enriched synthetic sampling of rare configurations, broader model families, including hybrid CNN–Transformer architectures, and explicit domain adaptation to real-world urban imagery.

In contrast to synthetic datasets primarily oriented toward autonomous driving, the SYNTHUADT framework explicitly models and evaluates mobility aids and sidewalk infrastructure, providing an accessibility-focused resource and a reproducible benchmark for future research in inclusive urban navigation and smart-city perception.

Contributor role

- **Santiago Felipe Luna Romero:** conceptualization, data curation, formal analysis, research, methodology, software, supervision, validation, visualization and writing—review & editing.
- **Renato Gouveia:** research, software, data curation and writing - original draft.
- **Mauren Abreu de Souza:** project administration, fundraising, resources and supervision.

References

[1] M. Ivanovs, K. Ozols, A. Dobrajs, and R. Kadikis, "Improving semantic segmentation of urban scenes for self-driving cars with synthetic images," *Sensors*, vol. 22, no. 6, p. 2252, Mar. 2022. [Online]. Available: http://doi.org/10.3390/s22062252

[2] E. Mohamed, K. Sirlantzis, and G. Howells, "Indoor/outdoor semantic segmentation using deep learning for visually impaired wheelchair users," *IEEE Access*, vol. 9, pp. 147 914–147 932, 2021. [Online]. Available: http://doi.org/10.1109/access.2021.3123952

[3] R. Azad, M. Heidary, K. Yilmaz, M. Hüttemann, S. Karimijafarbigloo, Y. Wu, A. Schmeink, and D. Merhof, "Loss functions in the era of semantic segmentation: A survey and outlook," *arXiv preprint*, 2023. [Online]. Available: http://doi.org/10.48550/ARXIV.2312.05391

[4] J. L. Gómez, M. Silva, A. Seoane, A. Borrás, M. Noriega, G. Ros, J. A. Iglesias-Guitian, and A. M. López, "All for one, and one for all: Urbansyn dataset, the third musketeer of synthetic driving scenes," 2023. [Online]. Available: http://doi.org/10.48550/ARXIV.2312.12176

[5] J. Tian, N. Mithun, Z. Seymour, H.-P. Chiu, and Z. Kira, "Striking the right balance: Recall loss for semantic segmentation," *arXiv preprint*, 2021. [Online]. Available: http://doi.org/10.48550/ARXIV.2106.14917

[6] Z. Song, Z. He, X. Li, Q. Ma, R. Ming, Z. Mao, H. Pei, L. Peng, J. Hu, D. Yao, and Y. Zhang, "Synthetic datasets for autonomous driving: A survey," 2023. [Online]. Available: http://doi.org/10.48550/ARXIV.2304.12205

[7] R. Kamimura, "Information-theoretic enhancement learning and its application to visualization of self-organizing maps," *Neurocomputing*, vol. 73, no. 13–15, pp. 2642–2664, Aug. 2010. [Online]. Available: http://doi.org/10.1016/j.neucom.2010.05.013

[8] Q. Wu and H. Liu, "Unsupervised domain adaptation for semantic segmentation using depth distribution," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 14 374–14 387. [Online]. Available: https://upsalesiana.ec/ing35ar9r1

[9] S. F. Luna-Romero, C. R. Stempniak, M. Abreu de Souza, and G. Reynoso-Meza, *Urban Digital Twins for Synthetic Data of Individuals with Mobility Aids in Curitiba, Brazil, to Drive Highly Accurate AI Models for Inclusivity.* Springer Nature Switzerland, 2024, pp. 116–125. [Online]. Available: http://doi.org/10.1007/978-3-031-52090-7_12

[10] Y. Yuan, Y. Du, Y. Ma, and H. Lv, "DSC-Net: enhancing blind road semantic segmentation with visual sensor using a dual-branch Swin-CNN architecture," *Sensors*, vol. 24, no. 18, p. 6075, Sep. 2024. [Online]. Available: http://doi.org/10.3390/s24186075

[11] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Álvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *arXiv preprint*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2105.15203

[12] S. F. Luna Romero, C. R. Stempniak, M. Abreu de Souza, and G. Reynoso-Meza, "A transfer learning model proposal for country border security using aerial thermal images," in *Procedings do XXIV Congresso Brasileiro de Automática*, ser. CBA2022. SBA Sociedade Brasileira de Automática, Oct. 2022. [Online]. Available: http://doi.org/10.20906/cba2022/3341

[13] S. F. L. Romero, M. A. d. Souza, and L. S. Andrade, "Synthua-dt: A methodological framework for synthetic dataset generation and automatic annotation from digital twins in urban accessibility applications," *Technologies*, vol. 13, no. 8, p. 359, Aug. 2025. [Online]. Available: http://doi.org/10.3390/technologies13080359

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *arXiv preprint*, 2015. [Online]. Available: http://doi.org/10.48550/ARXIV.1505.04597

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *arXiv preprint*, 2018. [Online]. Available: http://doi.org/10.48550/ARXIV.1802.02611

[16] S. F. Luna-Romero, M. Abreu de Souza, and L. Serpa Andrade, "Artificial vision systems for mobility impairment detection: Integrating synthetic data, ethical considerations, and real-world applications," *Technologies*, vol. 13, no. 5, p. 198, May 2025. [Online]. Available: http://doi.org/10.3390/technologies13050198

[17] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," *arXiv preprint*, 2018. [Online]. Available: http://doi.org/10.48550/ARXIV.1804.06516

[18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015. [Online]. Available: http://doi.org/10.48550/ARXIV.1502.03167

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *arXiv preprint*, 2017. [Online]. Available: http://doi.org/10.48550/ARXIV.1703.06870

[20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv preprint*, 2017. [Online]. Available: http://doi.org/10.48550/ARXIV.1708.02002

[21] J. Brewer, K. Rajagopal, A. Sadofyev, and W. van der Schee, "Evolution of the mean jet shape and dijet asymmetry distribution of an ensemble of holographic jets in strongly coupled plasma," *Journal of High Energy Physics*, vol. 2018, no. 2, Feb. 2018. [Online]. Available: http://doi.org/10.1007/jhep02(2018)015

[22] R. Gouveia. (2025) Pibiti semantic segmentation. Github, Inc. [Online]. Available: https://upsalesiana.ec/ing35ar9r3

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n35.2026.10
pISSN: 1390-650X / eISSN: 1390-860X

Swin transformer v2 for the classification of loja coffee

Swin transformer v2 para clasificación de café lojano

Patricio Bolívar Betancourt Ludeña[1], Oscar M. Cumbicus Pineda[1,*]

Received: 24-07-2025, Received after review: 01-12-2025, Accepted: 10-12-2025, Published: 01-01-2026

Abstract

This study presents a binary classification model for green coffee beans of the Arabica variety from the Loja region in Ecuador, based on the Swin Transformer V2 architecture. Two datasets were used, the public USK-COFFEE dataset of Indonesian origin and a proprietary dataset captured under controlled conditions. Two training strategies were evaluated: sequential transfer learning and unified training, with the latter achieving a validation accuracy of 98.30%. After hyperparameter optimization, the model reached 100% accuracy on a test set of 150 images and 93% accuracy on an external generalization set of 400 images with varying lighting conditions and backgrounds. Model interpretability was validated using Grad-CAM, demonstrating that the network focuses on actual defective regions rather than background information. An ablation analysis revealed that performance degradation in unconstrained scenarios is mainly due to sensitivity to noise and extreme lighting conditions. The main contributions of this work include the creation of a specialized dataset for Arabica green coffee from Loja and the development of an efficient model for its automatic classification.

Keywords: Rcoffee classification, artificial intelligence, Vision Transformer, Swin Transformer, computer vision, transfer learning

Resumen

Esta investigación presenta un modelo de clasificación binaria para granos de café verde de la variedad arábico procedentes de la región de Loja, Ecuador, basado en la arquitectura Swin Transformer V2. Se emplearon dos fuentes de datos, el conjunto de datos público USK-COFFEE, de origen indonesio, y un conjunto de datos propio capturado bajo condiciones controladas. Se evaluaron dos estrategias de entrenamiento: transferencia secuencial y entrenamiento unificado, siendo este último el que alcanzó una precisión de validación del 98,30%. Tras la optimización de hiperparámetros, el modelo logró una precisión del 100% en un conjunto de prueba de 150 imágenes y del 93% en un conjunto de generalización externo de 400 imágenes con condiciones variables de iluminación y fondo. La interpretabilidad del modelo se validó mediante Grad-CAM, evidenciando que la red enfoca su atención en zonas defectuosas reales. Un análisis de ablación mostró que la disminución de rendimiento en escenarios no controlados se debe principalmente a la sensibilidad al ruido y a la iluminación extrema. Como principales aportes, se destaca la creación de un conjunto de datos especializado y un modelo eficiente para la clasificación automática de café verde arábico.

Palabras clave: clasificación de café, inteligencia artificial, Vision Transformer, Swin Transformer, visión por computadora, aprendizaje por transferencia

[1,*]Carrera de Ingeniería en Computación, Universidad Nacional de Loja, Ecuador.
Corresponding author ✉: oscar.cumbicus@unl.edu.ec.

Suggested citation: P. T. Betancourt Ludeña and O.M. Cumbicus Pineda, "Swin transformer v2 for the classification of loja coffee," *Ingenius, Revista de Ciencia y Tecnología*, N.° 35, pp. 128-137, 2026, DOI: https://doi.org/10.17163/ings.n35.2026.10.

1. Introduction

The coffee production sector, specifically in the production of Arabica green coffee, is experiencing continuous growth [1]. Similarly, in the Loja region of Ecuador, the expansion of this sector has been evident [2]. However, it faces significant challenges in terms of the accurate classification of coffee beans, a process that is essential to ensure the quality of the final product and maintain its competitiveness in the international market [3]. Traditionally, this process has been carried out manually, which has resulted in variability in outcomes due to human intervention and fatigue from repetitive work [4]. In this context, technologies that have emerged in recent years, based on artificial intelligence (AI), specifically computer vision models, have offered an innovative and accurate solution for automating the task of coffee bean classification [5, 6]. To date, convolutional neural networks (CNNs) have stood out by demonstrating very favorable results in classification tasks [7,8], [5]. On the other hand, Vision Transformer (ViT) models, and more recently the Swin Transformer, have demonstrated superiority in generalizing certain local and global features of complex images [9, 10].

At present, a variety of models are proposed as solutions to this problem, such as in [11], where an accuracy of 99.84% was achieved in the detection of defects in Arabica green coffee beans. In the work presented in [12], multispectral images and SVM (Support Vector Machine) were used to classify between specialty and commercial classes of coffee beans, achieving 96% accuracy. Likewise, in [13], a multiclass classification approach was used, and an accuracy of 84.75% was reported with its Swin Transformer model when using the USK-Coffee dataset. These studies illustrate the capability of computer vision technologies to overcome human limitations in coffee bean classification. Similarly, an important limitation for model development at the local level is the absence of a dataset specific to the Loja region, which restricts the ability of the models to generalize and learn the specific characteristics of locally produced Loja coffee.

Recently, computer vision models based on Transformers have demonstrated great potential [14, 15]. Vision Transformers process images in patches, applying global and local attention to their details. In comparative works such as [16], where ViT and CNN models were compared on retinal images, it was shown that the Swin Transformer model achieved higher performance than convolutional neural networks, with an accuracy of 97.3%. Additionally, the Swin Transformer V2 model developed by Liu et al. [17] scaled its architecture to 3 billion parameters and set performance records by increasing the model's capacity and resolution. This advancement in the model version suggests a greater capability for discriminating more complex visual patterns, which is necessary for the detection of small defects in coffee beans. The present work proposes the design and evaluation of a binary classification model for Arabica green coffee beans from the Loja region, Ecuador.

Consequently, for the development of this study, public datasets of coffee images were used, complemented with a set of images taken from coffee beans from the Loja region, obtained through image acquisition under controlled conditions. The model was trained and validated using the previously described datasets and evaluated with the Loja dataset. This study assessed the performance of the Swin Transformer V2 model, pretrained on ImageNet-1K, through the application of Transfer Learning and Fine-Tuning techniques. On the other hand, this research contributes to the state of the art by providing a new dataset of Arabica green coffee from the province of Loja.

2. Materials and methods

In this section, the datasets used, the architecture of the implemented Swin Transformer V2 model, the training strategy, and the evaluation metrics employed for the classification of Arabica green coffee beans are described. All of the above is summarized into three phases covered by the CRISP-ML(Q) methodology [18], namely data engineering, model engineering, and model evaluation (Figure 1).

2.1. Datasets Used

For the development and evaluation of the model, two main datasets were used, related to the classification of green coffee beans (hulled and not subjected to roasting) of the Arabica variety. The selected datasets for this research are presented below.

USK-COFFEE: A public dataset containing 8,000 images of Arabica green coffee beans, originally distributed into four classes (peaberry, longberry, premium, defect) [19]. For this study, the original dataset was modified by consolidating three of its classes (peaberry, longberry, and premium) into a category referred to as "good," while the defect class was maintained as "defective," thus forming a binary classification. Specifically, data augmentation techniques were applied to the "defective" class, including random rotations, horizontal/vertical flips, and brightness/contrast adjustments, in order to balance the classes while maintaining a certain degree of data variability and to avoid overfitting during training.

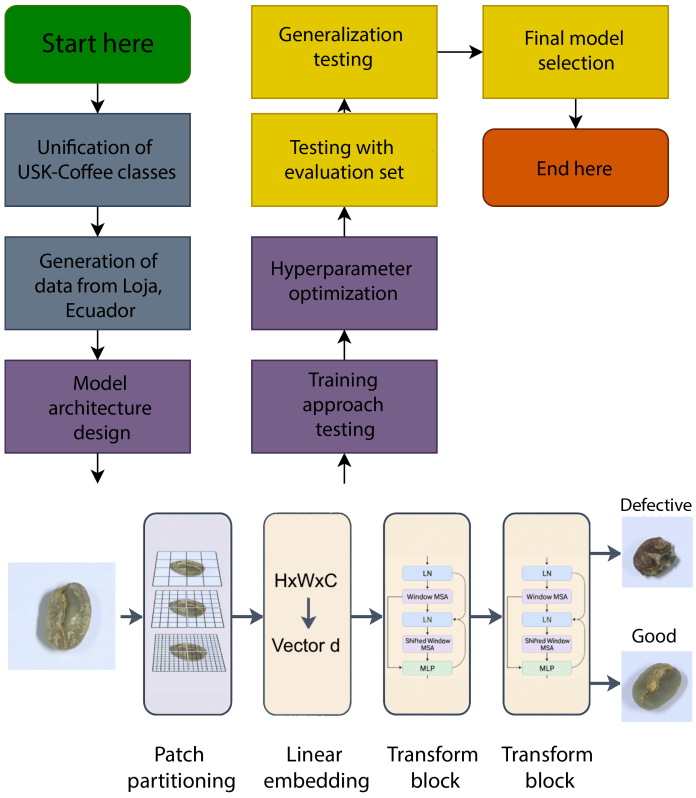


Figure 1. Materials and methods diagram. The workflow shows: coffee bean image input (H×W×C), patch partitioning, linear embedding, sequential Transformer blocks, and final binary classification (Good/Defective).

Loja Dataset: A proprietary dataset was constructed from Arabica green coffee beans collected in the Gonzanamá area, province of Loja, Ecuador. For data collection, five pounds of coffee were randomly selected, which were then hulled and milled. Subsequently, a local coffee grower performed manual classification, separating the beans into two categories: good and defective. In order to capture the images under controlled conditions, the beans were arranged on an A4 sheet of white bond paper placed on top of a white A3 cardboard sheet. Each photograph included 50 beans arranged in 5 columns and 10 rows, with separations of 5 cm and 3 cm, respectively. Illumination was ensured by two white LED lamps positioned at a height of 22 cm and a white RGB LED strip placed around the sheet.

The images were captured using a Canon EOS R50 camera equipped with a 20 mm lens, positioned 28 cm above the surface. The capture parameters were set as follows: shutter speed of 1/5 s, aperture F22, ISO sensitivity 100, and remote shutter release with a 2-second timer. Each photograph was recorded in JPG and RAW formats, with a resolution of 6000 × 4000 px (24 MP).

In total, 10 photographs of 50 beans each were captured, which were subsequently processed using a Python script to crop individual images of 256 × 256 px in RGB format, maintaining a white background. The final result was a balanced dataset of 1,000 images (500 per class), which constitutes the Loja dataset.

Both datasets were randomly divided into training, validation, and test subsets. For the modified USK-COFFEE dataset, a proportion of 80% was used for training and 20% for validation. For the Loja dataset, the split was 70% for training, 15% for validation, and 15% for testing (Figure 2).

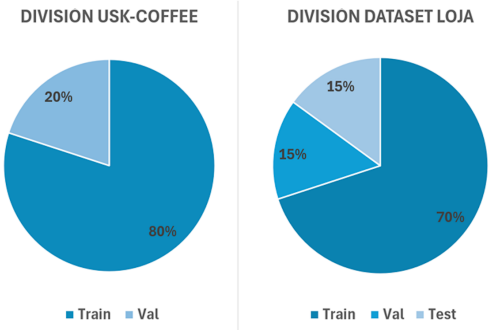


Figure 2. Percentage split of the USK-COFFEE and Loja datasets.

To promote reproducibility and collaboration within the scientific community, the Loja dataset has been made publicly available in the Zenodo repository under a CC BY-NC 4.0 license (Creative Commons

2.2. Swin Transformer V2 Model

The selected deep learning model was a Swin Transformer V2. This architecture was chosen for its ability to extract both local and global features from images, thanks to its hierarchical design with shifted window self-attention mechanisms that allow capturing patterns at different scales. Three pretrained variants of the Swin Transformer V2 (referred to as Base, Large, and Giant) were explored, which originate from prior training on the ImageNet-1K dataset [17].

The selected variant served as the starting point for fine-tuning on the specific task of classifying Loja coffee beans, offering different model capacities to be compared in the experiments. In general, the use of pretrained weights leveraged the knowledge acquired from general image classification, which accelerated training convergence in the new domain. The choice of this architecture was motivated by the particularities of Transformers in computer vision, highlighting their ability to capture global dependencies in the image through attention mechanisms, in contrast to traditional CNNs focused on local regions.

2.3. Training Methods

Two model training methods were evaluated using the selected data. The first method corresponded to a sequential application of transfer learning, in which the model was adjusted in two stages. In the first stage, initial fine-tuning was performed using only the USK-COFFEE dataset, and subsequently, a second fine-tuning stage was carried out using local images of Loja coffee, generalizing the features of the larger dataset before specializing in the characteristics of the smaller dataset. The second method consisted of joint training, combining both datasets from the beginning in a single training phase. In this method, images from USK-COFFEE and Loja were mixed during the learning process, so that the model learned simultaneously from both datasets. These two approaches made it possible to compare the results of training the model with the global dataset in a sequential manner and integrating it in a unified way during model optimization.

2.4. Experimental Configuration

The experiments were conducted using computing resources with GPU support to accelerate training. Specifically, a local workstation (Visual Studio Code environment) equipped with a GPU with 8 GB of VRAM was used, as well as an institutional server with JupyterHub providing graphical computing capacity with 4 GPUs of 12 GB of VRAM each (48 GB of total VRAM).

Regarding the optimization of training hyperparameters, seven models were generated by testing different values of learning rate, batch size, and dropout (Table 1). The number of training epochs was predefined based on preliminary experiments [18], providing sufficient iterations for model convergence without leading to overfitting.

Table 1. Values applied in the manual hyperparameter optimization. Ép: Epochs, TL: Batch size, TA: Learning rate, Opt.: Optimizer, Acr.: Model acronym.

Ép	TL	TA	Dropout	Opt.	Acr.
25	2	3×10^{-6}	0.5	Adam	M1
25	2	1×10^{-6}	0.5	Adam	M2
20	2	5×10^{-6}	0.5	Adam	M3
20	2	7×10^{-6}	0.45	Adam	M4
40	4	3×10^{-6}	0.55	Adam	M5
100	4	5×10^{-7}	0.55	Adam	M6
100	4	5×10^{-7}	0.55	SGD	M7

In terms of computational complexity, training the final model (configuration M4, see Table 1) required approximately one hour using the institutional server. This relatively short training time, given the size of the combined dataset, demonstrates the efficiency of the transfer learning approach. For practical implementation in a production environment, such as a coffee processing plant, a workstation with a dedicated mid-range or high-end GPU would be required to ensure real-time classification. Although this implies an initial investment in hardware, the cost is competitive compared to specialized optical sorting equipment, offering a scalable and software-adaptable solution.

3. Results and Discussion

3.1. Data Engineering

Se consolidaron los conjuntos de datos necesarios y se Techniques were applied to improve data quality. Initially, the USK-Coffee dataset was selected as the base due to its large size (8,000 images of Arabica green coffee). Since the task was formulated as a binary classification of "good" and "defective" beans, the original classes (peaberry, longberry, and premium) were unified into a single category called "good," while the "defect" class was renamed as "defective" (as described in Section 2.1). This modified dataset resulted in a class imbalance of 3,000 images of good beans versus 2,000 images of defective beans. To correct the class imbalance, oversampling with transformations (Figure 3) was performed on the defective class until obtaining 3,000 images per class.

Additionally, a dataset of Arabica green coffee collected and classified by producers from Loja was constructed. A total of 500 images per class ("good" and "defective") were captured, providing a local context with the intrinsic characteristics of coffee from the region, complementing the Indonesia dataset. The Indonesia and Loja datasets were randomly split for model training and validation, and 15% of the Loja dataset was reserved for testing the model's performance, as described in Section 2.1.

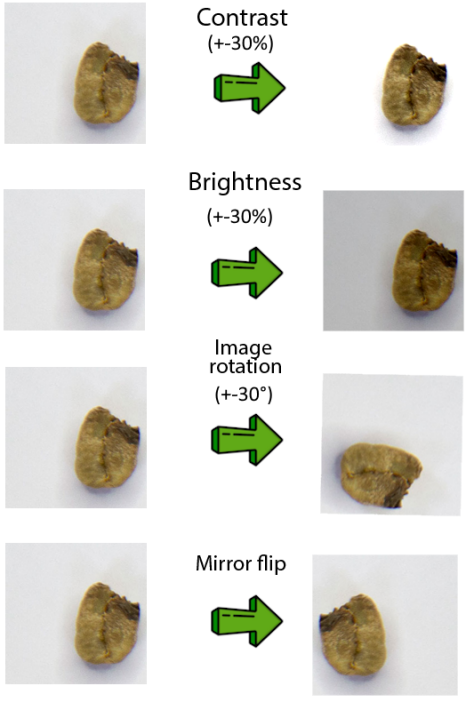


Figure 3. Transformations for oversampling the minority "defective" class.

3.2. Model Engineering

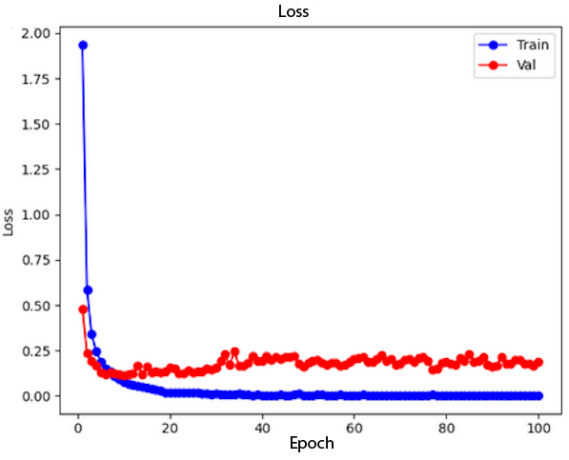
The pretrained Swin Transformer V2 architecture was selected and fine-tuned for the specific task of Arabica green coffee bean classification. Among the available variants, the Swin V2 Large model was chosen because it offers an adequate balance between performance and complexity: it reports a Top-1 Accuracy of 87.7% on ImageNet [17] with a moderate parameter size, avoiding the computational overhead of the Giant version (90.2% Top-1 but significantly heavier).

Two model training methods were explored to take advantage of the available data. In Method 1, the model was first trained using the USK-COFFEE (Indonesia) dataset, and then additional transfer learning was performed with the Loja dataset, aiming for the model to specialize in regional characteristics during a second phase. In Method 2, by contrast, both datasets were combined from the beginning into a single unified training process. After testing both approaches, it was observed that the unified dataset method achieved higher validation accuracy (98.30%) than the separated phased method (97.33%). Therefore, for subsequent stages, the model was trained using the combined data, suggesting that jointly exposing the model to the complete dataset enabled better initial generalization.

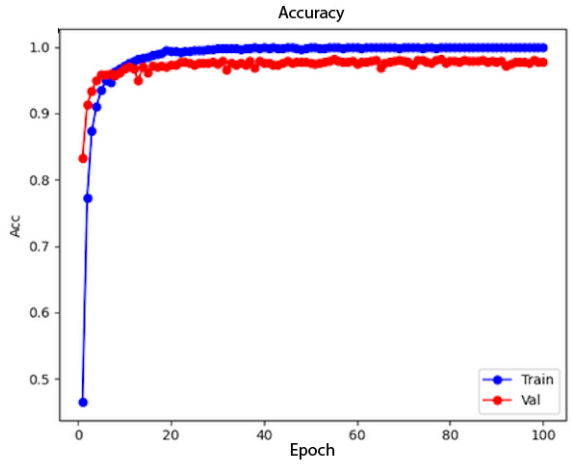
Additionally, a manual hyperparameter optimization was carried out, iteratively adjusting the values to maximize validation accuracy in order to refine the model's performance. Seven different configurations were tested (models M1–M7), varying key parameters such as number of epochs, batch size, learning rate, dropout, and optimizer. Each model was trained under the unified data strategy. In general, all models converged to high validation accuracies without drastic differences among most configurations. This result indicates that the choice of hyperparameters within reasonable ranges did not radically affect performance, due to the richness of the dataset and the robustness of the pretrained model. Nevertheless, to proceed with a more detailed evaluation, the four best-performing models (M1, M4, M5, and M6) were selected, with M6 showing the best performance during training and validation with an overall accuracy of 98.22% (Figure 4); additionally, those models with lower performance were discarded (Table 2).

Table 2. Validation accuracy and loss values for each model in the hyperparameter optimization.

Model	Validation accuracy	Loss validation
M1	98.15%	0.1242
M2	97.85%	0.1009
M3	97.85%	0.1401
M4	97.93%	0.1031
M5	97.93%	0.1749
M6	98.22%	0.1527
M7	96.67%	0.1235



(a)



(b)

Figure 4. Training and validation progress of model M6. (a) Loss curve, (b) Accuracy curve.

3.3. Model Evaluation

The performance of the selected models was analyzed on the test dataset, both in a controlled environment (test set) and under more challenging conditions, integrating classification metrics and their interpretation in light of the theoretical framework. Initially, the four models (M1, M4, M5, M6) were evaluated on the local test set (Loja images not used during training, 150 samples). All of them achieved an overall accuracy of 100% in this test, correctly classifying each image as either a "good" or "defective" bean (Figure 5). Consequently, their per-class metrics (precision, recall—sensitivity—and F1-score, the harmonic mean of precision and recall) reached values of 100% in both categories, demonstrating outstanding adaptation to the specific characteristics of Loja coffee beans.

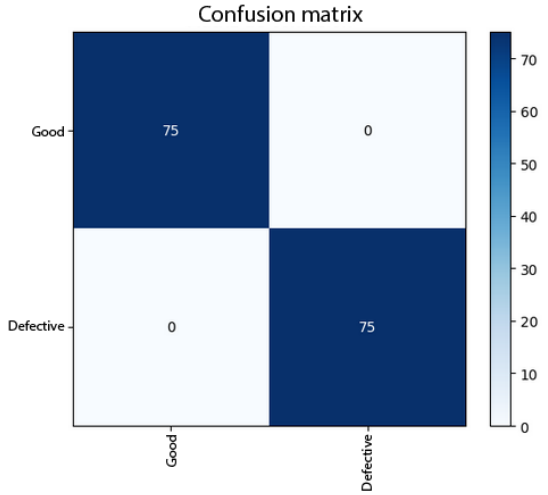


Figure 5. Confusion matrix of the models on the test dataset.

Although a 100% accuracy suggests excellent performance, it is important to consider that this test set originates from the same domain and controlled conditions as the training data (same bean varieties, similar lighting, uniform background). From a critical perspective, a perfect result in a known environment may indicate model overfitting to specific conditions, making a more demanding test necessary to evaluate the models' generalization capability.

Generalization Test: For a more rigorous evaluation based on the CRISP-ML(Q) methodology, the model was subjected to a generalization test set composed of 400 additional images with greater variability (lighting conditions, background, capture device, among others), simulating a real-world environment. These images were captured using mobile devices under uncontrolled conditions to assess the robustness of the model. In this scenario, a marginal degradation in performance was observed; however, the overall model performance remained high. The best-performing model (configuration M4, see Table 1) achieved an overall accuracy of 93%, while the others were around 92%. Figure 6 shows the confusion matrix obtained by model M4 in this generalization test. It can be observed that, out of 200 real images from each class, the model correctly identified 195 "good" beans and 176 "defective" beans.

Some errors were produced, with 5 false negatives and 24 false positives. These figures imply that, for the "good" class, a sensitivity (recall) of 97% and an F1-score (the harmonic mean of precision and recall) of 93% were achieved, while for the "defective" class, sensitivity was lower at 88% and the F1-score was 92%, mainly due to false positives. Despite this slight decrease in performance under more heterogeneous conditions, the model maintains an acceptable balance between precision and recall in both classes, avoiding a complete bias toward either of them. This suggests that the data balancing strategy and the incorporation of variability during training effectively contributed to achieving consistent classification across both bean categories.

A more detailed analysis of the errors in the generalization test provides valuable insights. The 24 false positives are mainly attributed to uncontrolled capture conditions. Variable lighting in the test images, captured with mobile devices under uncontrolled conditions (in contrast to the training dataset), likely generated shadows that obscured subtle defects, leading the model to incorrect classification. On the other hand, the 5 false negatives can be explained by similar factors; pronounced shadows or natural spots on healthy beans may have been misinterpreted as defects by the model. Additionally, the lower resolution of mobile device cameras may have affected the clarity of bean details, contributing to these misclassifications.

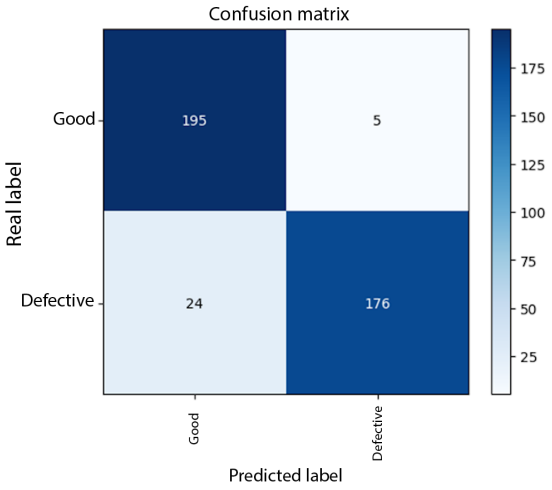


Figure 6. Confusion matrix of the model generalization test with the fourth configuration (M4).

To complement these metrics, the classification report of model M4 yielded a balanced accuracy (macro average—the average of the metrics for each class) of 93%. This value, together with the high recall scores (97% for "good" and 88% for "defective"), confirms that the model maintains balanced performance and does not exhibit a significant bias toward either of the two classes, addressing one of the key concerns in classification tasks with imbalanced classes.

Finally, although model M6 achieved the highest validation accuracy, model M4 demonstrated greater generalization capability on a more challenging and unseen dataset by exhibiting higher overall accuracy (93%), outperforming the other models (Table 3). In addition, it showed more uniform per-class metrics, in contrast to the other models which, despite having similar overall accuracy, exhibited small gaps between classes.

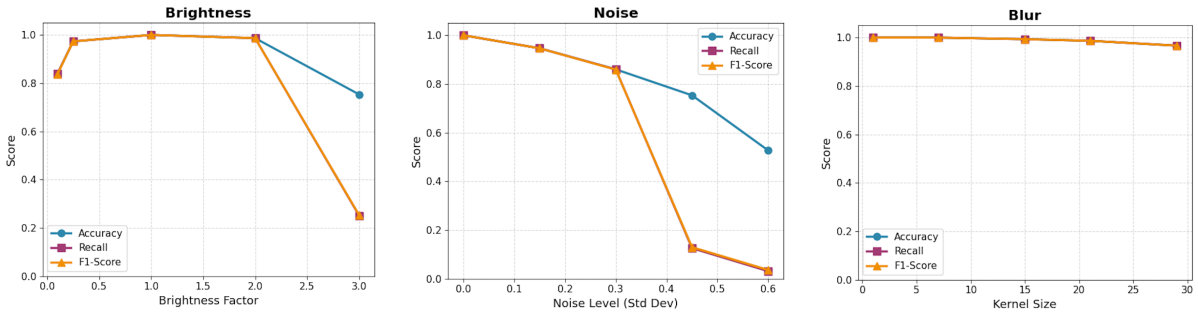
Table 3. Validation accuracy and loss values for each model in the hyperparameter optimization.

Model	Accuracy	Accuracy 'good'	Accuracy 'defective'
M1	0.92	0.88	0.96
M4	**0.93**	**0.89**	**0.97**
M5	0.92	0.89	0.96
M6	0.82	0.88	0.96

Class balance is fundamental in imbalanced binary classification tasks, as it indicates that the model does not prioritize the detection of one category over the other, but rather maintains a balanced and accurate recognition capability for both types of beans. The evaluation findings indicate that the fine-tuned Swin Transformer V2 model can achieve perfect results under controlled conditions; however, in more heterogeneous contexts, its accuracy may decrease, suggesting opportunities for improving the robustness of the implemented approach.

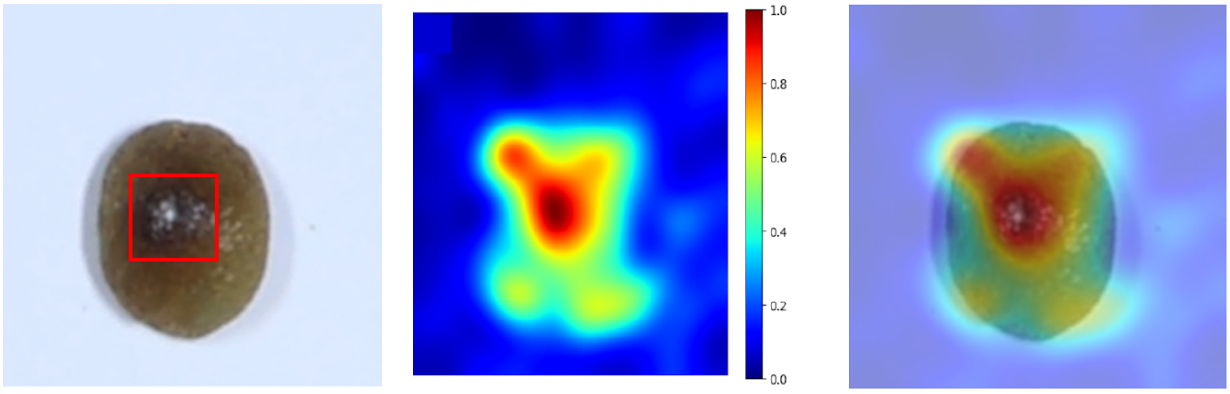
Ablation Analysis: Additionally, to explain the performance divergence between the controlled environment (100%) and the generalization test (93%), a quantitative ablation study was conducted by subjecting the model to synthetic environmental degradations: brightness variations, noise, and blur (Figure 7). The analysis reveals that the architecture is highly robust to blur, maintaining an F1-score above 96% even with large blur kernels (Figure 7c), indicating that lack of sharpness in mobile photos is not a critical factor. However, the model showed sensitivity to extreme lighting conditions and noise. As observed in Figure 7a (Brightness), although the model tolerates moderate variations, severe overexposure (factor > 2.5) causes a drastic drop in Recall to 25%, preventing defect detection. Similarly, noise injection (Noise, Figure 7b) linearly degrades accuracy; with a standard deviation of 0.6, Recall drops to 3%, suggesting that digital grain generated by mobile sensors under low-light conditions affects the model's ability to discern fine textures. These findings confirm that the 7% drop observed in the generalization test is mainly attributable to uncontrolled lighting factors and sensor noise, rather than focus-related issues.

Interpretability Tests: To complement the quantitative analysis and ensure model interpretability, the Grad-CAM visualization technique was employed [21]. This tool makes it possible to verify whether the neural network bases its predictions on relevant bean features or on spurious artifacts. Figure 8a presents the original image of a bean classified as defective, where the actual physical location of the defect prior to processing is delineated by a red bounding box. The subsequent visual results, shown in the heatmap (Figure 8b) and the Grad-CAM overlay (Figure 8c), demonstrate that the Swin Transformer model concentrates its highest activation (represented in red tones) precisely within the indicated defect area. This spatial correspondence between the real defect and the network's attention empirically validates the model's decision-making process, confirming that the classification is based on the morphology of the damage rather than background noise.



(a) (b) (c)

Figure 7. Ablation analysis under environmental perturbations. (a) Impact of brightness variation (Brightness). (b) Impact of Gaussian noise (Noise). (c) Impact of blur (Blur). High stability to blur is observed, but a significant degradation of Recall under extreme noise and overexposure conditions.



(a) Original (b) Heat map (c) Grad-CAM Pred: Faulty (91,2%)

Figure 8. Interpretability analysis using Grad-CAM. (a) Original image with the defect manually highlighted by a red bounding box. (b) Attention heatmap generated by the model. (c) Grad-CAM overlay showing that the model's activation coincides with the defect region.

3.4. Discussion

The results obtained are contrasted below with recent studies on Arabica green coffee bean classification using artificial intelligence, especially those based on Vision Transformer and CNN architectures, in order to contextualize the performance of our approach for Loja coffee.

First, the adapted Swin V2 model achieved internal validation accuracies of 98% and final generalization test accuracies of up to 93% overall. This performance aligns with the trend reported by recent works exploring Transformers in computer vision; for example, in [13], an accuracy of 84.75% was achieved using a Transformer-based architecture (Swin Transformer) in a multiclass classification task with the same USK-COFFEE dataset without modifications. In the present study, by simplifying the problem to binary classification and reinforcing it with local data and fine-tuning, a global validation accuracy of 98% was achieved. Similarly, in [15], a bean classification method using Swin Transformer oriented toward quality grading was proposed, also highlighting the potential of this architecture to capture relevant features in green coffee. The findings obtained support this evidence: Vision Transformers, when properly trained, can achieve very high accuracy in this task, even surpassing some traditional CNN models under similar scenarios.

In controlled testing environments, the ViT model trained in the present study achieved 100% accuracy, surpassing the accuracies reported by CNNs in similar tasks. In [6], an optimized CNN was designed for defect detection in coffee beans, achieving 95.2% accuracy on its test set. In contrast, the present model achieved 100% accuracy in local tests and 93% in more complex testing scenarios, demonstrating the ability of ViTs to achieve performance comparable to or superior to CNNs in bean classification. However, it is necessary to contextualize these numbers. In [5], a lightweight and explainable CNN (LDCNN) was developed for green coffee quality detection, achieving 98.38% accuracy and a 98.24% F1-score on its test dataset. Likewise,

Gope and Fukai [22] reported approximately 98.19% accuracy using a CNN to classify "peaberry" versus "normal" beans. These figures slightly surpass the 93% obtained by the model in the generalization test; however, it should be noted that these studies conducted evaluations in controlled environments or under fixed conditions.

On the other hand, in [11], a MobileNetV3 architecture was used to classify defects in Thai Arabica coffee (multiclass task), achieving 88.63% accuracy, a value lower than the 93% reached by the model trained under a more complex generalization challenge. Overall, these comparisons suggest that the ViT-based approach used in this work is consistent with the surveyed state of the art, achieving favorable metric results compared to many previous CNN models, especially by simplifying the classification to two classes and by incorporating a new dataset specialized in the Loja region. At the same time, the robustness analysis reveals that, although the model outperforms previous architectures, there is still room for improvement to match the noise tolerance demonstrated by some CNNs, suggesting the need to include noise- and illumination-specific data augmentation in future work.

4. Conclusions

The creation and evaluation of the Vision Transformer model for the classification of Loja Arabica green coffee beans demonstrated a high discrimination capability, achieving an overall accuracy of 100% under controlled conditions and 93% in generalization tests with varied conditions.

The adoption of the CRISP-ML(Q) methodology, together with the initial comparison of training methods using the USK-COFFEE dataset and the proprietary Loja dataset, enabled faster and more stable model convergence. Likewise, the manual optimization of hyperparameters—epochs, batch size, learning rate, dropout, and optimizer—proved to be key in refining its performance, achieving validation accuracies above 98% without large fluctuations among the evaluated configurations.

These findings confirm the potential of Vision Transformers as a quality control tool in the coffee industry of Loja. Nevertheless, the slight decrease in accuracy under less ideal conditions demonstrates the existence of room for improvement, where the diversity of the dataset can be expanded, new data augmentation techniques can be incorporated, and hyperparameter optimization schemes can be explored. Consequently, this could strengthen generalization capability and ensure equally robust results in less controlled scenarios.

Acknowledgments

We extend our gratitude to the National University of Loja and the Computer Engineering program for the support and resources provided for this work.

Contributor role

- **Patricio Bolívar Betancourt Ludeña:** writing – original draft, software and data curation.
- **Oscar M. Cumbicus Pineda:** supervision and validation.

References

[1] ICP. (2025) I-CIP retreats on news of looser supply, relieving some of the upward pressure. Coffee market report. International Coffee Organization. [Online]. Available: https://upsalesiana.ec/ing35ar10r1

[2] Agricultura. (2025) 6425 hectáreas de café son renovadas en la provincia de Loja. Ministerio de Agricultura, Ganadería y Pesca. [Online]. Available: https://upsalesiana.ec/ing35ar10r2

[3] M. Faisal, J.-S. Leu, and J. T. Darmawan, "Model selection of hybrid feature fusion for coffee leaf disease classification," *IEEE Access*, vol. 11, pp. 62 281–62 291, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3286935

[4] E. Hassan, "Enhancing coffee bean classification: a comparative analysis of pretrained deep learning models," *Neural Computing and Applications*, vol. 36, no. 16, pp. 9023–9052, Apr. 2024. [Online]. Available: https://doi.org/10.1007/s00521-024-09623-z

[5] C.-H. Hsia, Y.-H. Lee, and C.-F. Lai, "An explainable and lightweight deep convolutional neural network for quality detection of green coffee beans," *Applied Sciences*, vol. 12, no. 21, p. 10966, Oct. 2022. [Online]. Available: https://doi.org/10.3390/app122110966

[6] S.-J. Chang and C.-Y. Huang, "Deep learning model for the inspection of coffee bean defects," *Applied Sciences*, vol. 11, no. 17, p. 8226, Sep. 2021. [Online]. Available: https://doi.org/10.3390/app11178226

[7] A. Chavarro, D. Renza, and E. Moya-Albor, "Convnext as a basis for interpretability in coffee leaf rust classification," *Mathematics*, vol. 12, no. 17, p. 2668, Aug. 2024. [Online]. Available: https://doi.org/10.3390/math12172668

[8] Y. A. Auliya, I. Fadah, Y. Baihaqi, and I. N. Awwaliyah, "Green bean classification: Fully convolutional neural network with Adam optimization," *Mathematical Modelling of Engineering Problems*, vol. 11, no. 6, pp. 1641–1648, Jun. 2024. [Online]. Available: https://doi.org/10.18280/mmep.110626

[9] J. Maurício, I. Domingues, and J. Bernardino, "Comparing vision transformers and convolutional neural networks for image classification: A literature review," *Applied Sciences*, vol. 13, no. 9, p. 5521, Apr. 2023. [Online]. Available: https://doi.org/10.3390/app13095521

[10] J. Wei, J. Chen, Y. Wang, H. Luo, and W. Li, "Improved deep learning image classification algorithm based on Swin Transformer V2," *PeerJ Computer Science*, vol. 9, p. e1665, Oct. 2023. [Online]. Available: https://doi.org/10.7717/peerj-cs.1665

[11] S. Arwatchananukul, D. Xu, P. Charoenkwan, S. Aung Moon, and R. Saengrayap, "Implementing a deep learning model for defect classification in Thai Arabica green coffee beans," *Smart Agricultural Technology*, vol. 9, p. 100680, Dec. 2024. [Online]. Available: https://doi.org/10.1016/j.atech.2024.100680

[12] W. Pinheiro Claro Gomes, L. Gonçalves, C. Barboza da Silva, and W. R. Melchert, "Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee," *Computers and Electronics in Agriculture*, vol. 198, p. 107097, Jul. 2022. [Online]. Available: https://doi.org/10.1016/j.compag.2022.107097

[13] M. N. Izza and G. P. Kusuma, "Image classification of Green Arabica Coffee using transformer-based architecture," *International Journal of Engineering Trends and Technology*, vol. 72, no. 6, pp. 304–314, Jun. 2024. [Online]. Available: https://doi.org/10.14445/22315381/IJETT-V72I6P128

[14] H. F. Alhasson and S. S. Alharbi, "Classification of saudi coffee beans using a mobile application leveraging squeeze vision transformer technology," *Neural Computing and Applications*, vol. 37, no. 14, pp. 8629–8649, Feb. 2025. [Online]. Available: https://doi.org/10.1007/s00521-025-11024-9

[15] Y. Jiao, Y. Zhao, A. Jia, T. Wang, J. Li, K. Xiang, H. Deng, M. He, R. Jiang, and Y. Zhang, "Swin-HSSAM: a green coffee bean grading method by swin transformer," *PLOS One*, vol. 20, no. 5, p. e0322198, May 2025. [Online]. Available: https: //doi.org/10.1371/JOURNAL.PONE.0322198

[16] J. H. L. Goh, E. Ang, S. Srinivasan, X. Lei, J. Loh, T. C. Quek, C. Xue, X. Xu, Y. Liu, C.-Y. Cheng, J. C. Rajapakse, and Y.-C. Tham, "Comparative analysis of vision transformers and conventional convolutional neural networks in detecting referable diabetic retinopathy," *Ophthalmology Science*, vol. 4, no. 6, p. 100552, Nov. 2024. [Online]. Available: https: //doi.org/10.1016/j.xops.2024.100552

[17] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin Transformer V2: scaling up capacity and resolution," *arXiv*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2111.09883

[18] S. Studer, T. B. Bui, C. Drescher, A. Hanuschkin, L. Winkler, S. Peters, and K.-R. Müller, "Towards CRISP-ML(Q): a machine learning process model with quality assurance methodology," *Machine Learning and Knowledge Extraction*, vol. 3, no. 2, pp. 392–413, Apr. 2021. [Online]. Available: https://doi.org/10.3390/make3020020

[19] A. Febriana, K. Muchtar, R. Dawood, and C.-Y. Lin, "USK-Coffee dataset: A multi-class green arabica coffee bean dataset for deep learning," in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*. IEEE, Jun. 2022, pp. 469–473. [Online]. Available: https://doi.org/10.1109/CyberneticsCom55287.2022.9865489

[20] Patricio Bolívar Betancourt Ludeña, "Lojano Arabica coffee," *Zenodo*, 2025. [Online]. Available: https://doi.org/10.34740/kaggle/dsv/13947455

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Oct. 2019. [Online]. Available: http://dx.doi.org/10.1007/s11263-019-01228-7

[22] H. L. Gope and H. Fukai, "Peaberry and normal coffee bean classification using CNN, SVM, and KNN: their implementation in and the limitations of Raspberry Pi 3," *AIMS Agriculture and Food*, vol. 7, no. 1, pp. 149–167, 2022. [Online]. Available: https://doi.org/10.3934/agrfood.2022010

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
pISSN: 1390-650X / eISSN: 1390-860X

Guidelines for publication in Ingenius Journal

1. General Information

Ingenius is a scientific publication of the *Universidad Politécnica Salesiana* of Ecuador, published since January 2007, with a fixed biannual periodicity, specialized in Mechanical Engineering, Electrical Engineering, Electronics, Computer Science and its integration in what is now known as Mechatronics; these lines of action strengthen areas such as automation, control, robotics, among others..

It is a scientific journal, which uses the peer-review system, under double-blind review methodology, according to the publication standards of the Institute of Electrical and Electronics Engineers (IEEE). Compliance with this system allows authors to guarantee an objective, impartial and transparent review process, which facilitates the publication of their inclusion in reference databases, repositories and international indexing.

Ingenius is indexed in the directory and selective catalog of the Regional Online Information System for Scientific Journals of Latin America, the Caribbean, Spain and Portugal (Latindex), in the Directory of Journals of Open Access DOAJ, In the Information Matrix for the Analysis of Journals, MIAR, In the Ibero-American Network of Innovation and Scientific Knowledge, REDIB and in repositories, libraries and specialized catalogs of Latin America.

The journal is published in a double version: printed (ISSN: 1390-650X) and digital (e-ISSN: 1390-860X), in Spanish, each work being identified with a DOI (Digital Object Identifier System. The articles sent to Ingenius magazine must comply with the following criteria:

2. Scope and policy

2.1. Theme

Original contributions in Mechanical Engineering, Electrical and Electronic Engineering, Computer Science and its integration in what is now known as Mechatronics, as well as related areas: Automation, Control, Domotics, Robotics in their different fields of action and all those related disciplines with the same central theme.

All the work carried out by national or foreign researchers may be published once they meet the required scientific quality criteria.

2.2. Contributions

Ingenius Journal preferably publishes articles related to empirical research, and also reports of technological development, proposals for models and innovations, products for the elaboration of graduate and postgraduate thesis that contribute to the field of science and technology, as well as select revisions of literature. (state-of-the-art).

- **Research:** 5,000 to 6,500 words of text, including title, abstracts, descriptors, charts and references.
- **Reports:** 5,000 to 6,500 words of text, including title, abstracts, charts and references.
- **Reviews:** 6,000 to 7,000 words of text, including charts and references. Current, selective and justified references, would be specially valued from among 40 works

The Ingenius Journal publishes original and unpublished works written in Spanish and English, they may not have been published

through any printed or electronic media, nor be in the process of arbitration or publication.

Every article will be subjected to a rigorous arbitration process; the evaluation of the article will be made according to criteria of originality, relevance, relevance, contributions, scientific rigor and compliance with established editorial guidelines.

Being an arbitrated publication, the Editorial Board approves its publication based on the concept of specialized pairs. The reception of a document does not imply commitment of publication.

It is essential to present a letter of presentation and grant of rights which can be downloaded from: <
urlhttps://goo.gl/ZNkMRD>.

Contributions must be exclusively sent and through the OJS (Open Journal System) <`https://goo.gl/JF7dWT`>. In which all authors must previously register as a user. For any consultation of the procedure you should contact:

<revistaingenius@ups.edu.ec>,
<jcalle@ups.edu.ec> ó
<mquinde@ups.edu.ec>.

To promote diversity in publications, the author(s) may not publish more than one (1) article per issue, nor in consecutive issues. In order for the author(s) to submit their research again to Ingenius, a minimum of 3 published issues must have elapsed.

3. Presentation and structure of the manuscripts

For those works that are empirical investigations, the manuscripts will follow the IMRDC structure (Introduction, Materials and Methods, Results and Discussion and Conclusions), being optional the Notes and Supports. Those papers that, on the contrary, deal with reports, studies, proposals and reviews may be more flexible in their epigraphs, particularly in material and methods, analysis, results, discussion and conclusions. In all typologies of works, references are mandatory.

Articles may be written on Microsoft Word (.doc or .docx) or LaTeX(.tex). The template to be used can be downloaded from the journal's website, a, <`https://goo.gl/gtCg6m`>, while for LaTeX in <`https://goo.gl/hrHzzQ`>, it is necessary that the file be anonymised in Properties of File, so that the author(s) ID is not displayed.

Figures, Graphs and/or Illustrations, as well as Charts shall be numbered sequentially including an explanatory description for each. The equations included in the article must also be numbered; the figures, charts and equations must be cited in the text.

Use space after point, commas and question marks.

Use "enter" at the end of each paragraph and title heading. Do not use .$^{\text{en}}$ter.$^{\text{a}}$nywhere else, let the word processor program automatically break the lines.

Do not center headings or subheadings as they should be aligned to the left.

Charts must be created in the same program used for the document body, but must be stored in a separate file. Use tabs, not spaces, to create columns. Remember that the final size of printed pages will be 21 x 28 cm, so the tables must be designed to fit the final print space.

3.1. Structure of the manuscripts

3.1.1. Presentation and cover letter

1. **Título (español) / Title (inglés):** Concise but informative, in Spanish on the front line and in English on the second, when the article is written in Spanish and vice versa if it is written in English.

2. **Authors and affiliations:** Full name and surname of each author, organized by order of priority and their institutional affiliation with reference to the end of the first sheet, where it must include: Dependency to which belongs within the institution, Institution to which he/she belongs, country, ORCID. A maximum of 5 authors will be accepted, although there may be exceptions justified by the complexity and extent of the topic.
3. **Abstract (Spanish) / Abstract (English):** It will have a maximum extension of 230 words, first in Spanish and then in English. : 1) Justification of the topic; 2) Objectives; 3) Methodology and sample; 4) Main results; 5) Main conclusions.
4. **Keywords (Spanish) / Keywords (English):** 6 descriptors must be presented for each language version directly related to the subject of the work. The use of the key words set out in UNESCO's Thesaurus will be positively valued.
5. **Presentation (Cover Letter):** A statement that the manuscript is an original contribution, not submission or evaluation process in another journal, with the confirmation of the signatory authors, acceptance (if applicable) of formal changes in the manuscript according to the guidelines and partial assignment of rights to the publisher, according to the format established in: <https://goo.gl/ZNkMRD>

3.1.2. Manuscript

1. **Título (español) / Title (inglés):** Concise but informative, in Spanish on the front line and in English on the second, when the article is written in Spanish and vice versa if it is written in English.
2. **Authors and affiliations:** Full name and surname of each author, organized by order of priority and their institutional affiliation with reference to the end of the first sheet, where it must include: Dependency to which belongs within the institution, Institution to which he/she belongs, country, ORCID. A maximum of 5 authors will be accepted, although there may be exceptions justified by the complexity and extent of the topic.
3. **Abstract (Spanish) / Abstract (English):** It will have a maximum extension of 230 words, first in Spanish and then in English. : 1) Justification of the topic; 2) Objectives; 3) Methodology and sample; 4) Main results; 5) Main conclusions.
4. **Keywords (Spanish) / Keywords (English):** 6 descriptors must be presented for each language version directly related to the subject of the work. The use of the key words set out in UNESCO's Thesaurus will be positively valued.
5. **Introduction:** It should include the problem statement, context of the problem, justification, rationale and purpose of the study, using bibliographical citations, as well as the most significant and current literature on the topic at national and international level.
6. **Material and methods:** It must be written so that the reader can easily understand the development of the research. If applicable, it will describe the methodology, the sample and the form of sampling, as well as the type of statistical analysis used. If it is an original methodology, it is necessary to explain the reasons that led to its use and to describe its possible limitations.
7. **Analysis and results:** It will try to highlight the most important observations, describing, without making value judgments, the material and methods used. They will appear in a logical sequence

in the text and the essential charts and figures avoiding the duplication of data.

8. **Discussion and Conclusions:** It will summarize the most important findings, relating the observations themselves to relevant studies, indicating contributions and limitations, without adding data already mentioned in other sections. It should also include deductions and lines for future research.
9. **Supports and acknowledgments (optional):** The Council Science Editors recommends the author (s) to specify the source of funding for the research. Priority will be given to projects supported by national and international competitive projects.
10. **The notes (optional):** will go, only if necessary, at the end of the article (before the references). They must be manually annotated, since the system of footnotes or the end of Word is not recognized by the layout systems. The numbers of notes are placed in superscript, both in the text and in the final note. The numbers of notes are placed in superscript, both in the text and in the final note. No notes are allowed that collect simple bibliographic citations (without comments), as these should go in the references.
11. **References:** Bibliographical citations should be reviewed in the form of references to the text. Under no circumstances should references mentioned in the text not be included. Their number should be sufficient to contextualize the theoretical framework with current and important criteria. They will be presented sequentially in order of appearance, as appropriate following the format of the IEEE.

3.2. Guidelines for Bibliographical references

Journal articles:

[1] J. Riess, J. J. Abbas, "Adaptive control of cyclic movements as muscles fatigue using functional neuromuscular stimulation". IEEE Trans. Neural Syst. Rehabil. Eng vol. 9, pp.326–330, 2001. [Onine]. Available: https://doi.org/10.1109/7333.948462

Books:

[1] G. O. Young, "Synthetic structure of industrial plastics" in Plastics, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw–Hill, 1964, pp. 15–64.

Technical reports:

[1] M. A. Brusberg and E. N. Clark, "Installation, operation, and data evaluation of an oblique–incidence ionosphere sounder system," in "Radio Propagation Characteristics of the Washington–Honolulu Path," Stanford Res. Inst., Stanford, CA, Contract NOBSR–87615, Final Rep., Feb. 1995, vol. 1

Articles presented in confeences (unpubished):

[1] Vázquez, Rolando, Presentación curso "Realidad Virtual". National Instruments. Colombia, 2009.

Articles of memories of Conferences (Published):

[1] L. I. Ruiz, A. García, J. García, G. Taboada. "Criterios para la optimización de sistemas eléctricos en refinerías de la industria petrolera: influencia y análisis en el equipo eléctrico," IEEE CONCAPAN XXVIII, Guatemala 2008.

Thesis:

[1] L.M. Moreno, "Computación paralela y entornos heterogéneos," Tesis doctoral, Dep. Estadística, Investigación Operativa y Computación, Universidad de La Laguna, La Laguna, 2005.

Guidelines:

[1] IEEE Guide for Application of Power Apparatus Bushings, IEEE Standard C57.19.100–1995, Aug. 1995.

Patents:

[1] J. P. Wilkinson, "Nonlinear resonant circuit devices," U.S. Patent 3 624 125, July 16, 1990.

Manuals:

[1] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

Internet resources:

[1] E. H. Miller, "A note on reflector arrays" [Online]. Available. https://goo.gl/4cJkCF

3.3. Epigraphs, Figures and Charts

The epigraphs of the body of the article will be numbered in Arabic. They should go without a full box of capital letters, neither underlined nor bold. The numbering must be a maximum of three levels: 1. / 1.1. / 1.1.1. At the end of each numbered epigraph will be given an enter to continue with the corresponding paragraph.

The charts must be included in the text according to order of appearance, numbered in Arabic and subtitled with the description of the content, the subtitle should go at the top of the table justified to the left.

Figures can be linear drawings, maps or black and white halftone or color photographs in 300 dpi resolution. Do not combine photographs and line drawings in the same figure.

Design the figures so that they fit eventually to the final size of the journal 21 x 28 cm. Make sure inscriptions or details, as well as lines, are of appropriate size and thickness so that they are not illegible when they are reduced to their final size (numbers, letters and symbols must be reduced to at least 2.5 mm in height After the illustrations have been reduced to fit the printed page). Ideally, the linear illustrations should be prepared at about a quarter of their final publication size.

Different elements in the same figure should be spelled a, b, c, etc.

Photographs should be recorded with high contrast and high resolution. Remember that photographs frequently lose contrast in the printing process. Line drawings and maps should be prepared in black.

The text of the figures and maps must be written in easily legible letters.

If the figures have been previously used, it is the responsibility of the author to obtain the corresponding permission to avoid subsequent problems related to copyright.

Each figure must be submitted in a separate file, either as bitmap (.jpg, .bmp, .gif, or .png) or as vector graphics (.ps, .eps, .pdf).

4. Submission process

The manuscript must be sent through the OJS system of the journal, <`https://goo.gl/JF7dWT`>,the manuscript should be uploaded as an original file in .pdf without author data and anonymized according to the above; In complementary files the complete manuscript must be loaded in .doc or .docx (Word file), that is to say with the data of the author (s) and its institutional ascription; Also the numbered figures should be uploaded in independent files according to the corresponding in the manuscript (as bitmap .jpg, .bmp, .gif, or .png or as vector graphics .ps, .eps, .pdf). It is also obligatory to upload the cover letter and grant of rights as an additional file.

All authors must enter the required information on the OJS platform and only one of the authors will be responsible for correspondence.

Once the contribution has been sent the system will automatically send the author for correspondence a confirmation email of receipt

of the contribution.

5. Editorial process

Once the manuscript has been received in OJS, a first check by the editorial team of the following points:

- The topic is in accordance with the criteria of the journal.
- Must have the IMRDC structure.
- Must be in the INGENIUS format.
- Must use the IEEE citation format.
- All references should be cited in the text of the manuscript as well as charts, figures and equations.
- The manuscript is original; for this, software is used to determine plagiarism.

The assessment described above can take up to 4 weeks.

If any of the above is not complete or there is inconsistency, an email will be sent to the author to make the requested corrections.

The author will make the corrections and resend the contribution through an email in response to the notification and will also upload the corrected manuscript into OJS supplementary files.

The editorial team will verify that the requested corrections have been incorporated, if it complies, the manuscript will start the second part of the process that may be followed by the author through OJS, otherwise the author will be notified and the manuscript will be archived.

The second phase of the process consists of the evaluation under the methodology of double-blind review, which includes national and foreign experts considering the following steps:

- The editor assigns two or more reviewers for the article.
- After reviewing the article, the reviewers will submit the evaluation report with one of the following results.
  - Publishable
  - Publishable with suggested changes
  - Publishable with mandatory changes
  - Non publishable
- The editor once received the evaluation by the reviewers will analyze the results and determine if the article is accepted or denied.
- If the article is accepted, the author will be notified to make corrections if required and the corresponding editorial process will be continued.
- If the article is denied, the author will be notified and the manuscript will be archived.
- In the two previous cases the result of the evaluation of the reviewers and their respective recommendations will be sent.

The second phase of the process lasts at least 4 weeks, after which they will be notified to the author giving instructions to continue with the process.

6. Publication

The INGENIUS Journal publishes two issues per year, on January 1st and July 1st, so it is important to consider the dates for sending the articles and their corresponding publication. Articles received until October will be considered for the January publication and those received until April for the July publication.

7. Information on the Use of Artificial Intelligence

Should artificial intelligence be used at any stage of the research presented in the article, authors are required to clearly highlight this in the cover letter associated with the article, specifying the section or sections where artificial intelligence has been used. The purpose

of this requirement is to inform readers about the sections where this technology has been employed, providing greater transparency and understanding of its application in the presented research.

INGENUS, Revista de Ciencia y Tecnologia, recognizes the importance of maintaining high ethical standards in scientific research, particularly in the use of artificial intelligence (AI).

The decision to accept a publication that has utilized artificial intelligence rests at the discretion of the editorial team.

UNIVERSIDAD POLITÉCNICA SALESIANA DEL ECUADOR
Juan Cárdenas Tapia, sdb,
Rector

Turuhuayco 3-69 y Calle Vieja
Postal code 2074
Cuenca, Ecuador
Teléfono: (+593 7) 205 00 00
Fax: (+593 7) 408 89 58
Email: srector@ups.edu.ec

Exchange

Exchange with other periodicals is accepted.

Address:
Secretaría Técnica de Comunicación
Universidad Politécnica Salesiana
Turuhuayco 3-69 y Calle Vieja
Postal code 2074
Cuenca, Ecuador
Phone: (+593 7) 205 00 00 Ext. 1182
Fax: (+593 7) 408 89 58
Email: rpublicas@ups.edu.ec
www.ups.edu.ec
Cuenca – Ecuador

Ingenius, Journal Science of Technology,
Issue 35
january/june 2026
John Calle Siguencia, Editor in chief
revistaingenius@ups.edu.ec

Printed
Centro Gráfico Salesiano: Antonio Vega Muñoz 10-68 y General Torres.
Phone: (+593 7) 283 17 45
Cuenca – Ecuador
Email: centrograficosalesiano@lns.com.ec

OTHER PERIODIC PUBLICATIONS OF THE UNIVERSITY

UNIVERSITAS, Journal of Social and Human Sciences.

LA GRANJA, Journal of the Sciences.

ALTERIDAD, Journal of Education.

RETOS,Journal of Administration Sciences and Economics.

UTOPÍA, University Youth Ministry Magazine.

SOPHIA, Collection of Philosophy of Education.

ABYA YALA | UNIVERSIDAD POLITÉCNICA SALESIANA

ECUADOR SALESIANOS DON BOSCO