ISSN 2523-6814

Journal of

Computational Technologies

Volume 9, Issue 22 — e2025922 January — December - 2025

ECORFAN®

Journal of Computational Technologies, Volume 9, Issue 22: e2025922 January – December 2025, is a Continuous publication - ECORFAN-Taiwan. Taiwan, Taipei. YongHe district, ZhongXin, Street 69. Postcode: 23445. WEB: www.ecorfan.org/taiwan, revista@ecorfan.org. Chief Editor: Quintanilla - Cóndor, Cerapio. PhD. ISSN-On line: 2523-6814. Responsible for the latest update of this number ECORFAN Computer Unit. Escamilla-Bouchán, Imelda, PhD, Luna-Soto, Vladimir. PhD. Taiwan, Taipei. YongHe district, ZhongXin, Street 69, last updated December 30, 2025.

Journal of Computational Technologies

Definition of Research Journal

Scientific Objectives

Support the international scientific community in its written production Science, Technology and Innovation in the Field of Engineering and Technology, in Subdisciplines Digital skills standards for education, Learning projects through the use of information, Technologies and communication, Development of digital, Competencies teaching digital skills programs, Management of technological and educational, Consulting fields of technological training, Applied to education

ECORFAN-Mexico SC is a Scientific and Technological Company in contribution to the Human Resource training focused on the continuity in the critical analysis of International Research and is attached to SECIHTI -RENIECYT number 1702902, its commitment is to disseminate research and contributions of the International Scientific Community, academic institutions, agencies and entities of the public and private sectors and contribute to the linking of researchers who carry out scientific activities, technological developments and training of specialized human resources with governments, companies and social organizations.

Encourage the interlocution of the International Scientific Community with other Study Centers in Mexico and abroad and promote a wide incorporation of academics, specialists and researchers to the publication in Science Structures of Autonomous Universities - State Public Universities - Federal IES - Polytechnic Universities - Technological Universities - Federal Technological Institutes - Normal Schools - Decentralized Technological Institutes - Intercultural Universities - S & T Councils - SECIHTI Research Centers.

Scope, Coverage and Audience

Journal of Computational Technologies is a Research Journal edited by ECORFAN-Mexico S.C in its Holding with repository in Taiwan, is a scientific publication arbitrated and indexed with semester periods. It supports a wide range of contents that are evaluated by academic peers by the Double-Blind method, around subjects related to the theory and practice of Digital skills standards for education, Learning projects through the use of information, Technologies and communication, Development of digital, Competencies teaching digital skills programs, Management of technological and educational, Consulting fields of technological training, Applied to education with diverse approaches and perspectives, That contribute to the diffusion of the development of Science Technology and Innovation that allow the arguments related to the decision making and influence in the formulation of international policies in the Field of Engineering and Technology. The editorial horizon of ECORFAN-Mexico® extends beyond the academy and integrates other segments of research and analysis outside the scope, as long as they meet the requirements of rigorous argumentative and scientific, as well as addressing issues of general and current interest of the International Scientific Society.

Editorial Board

Arbitration Committee

Assignment of Rights

The sending of an Article to Journal of Computational Technologies emanates the commitment of the author not to submit it simultaneously to the consideration of other series publications for it must complement the Originality Format for its Article.

The authors sign the Authorization Format for their Article to be disseminated by means that ECORFAN-Mexico, S.C. In its Holding Taiwan considers pertinent for disclosure and diffusion of its Article its Rights of Work.

Declaration of Authorship

Indicate the Name of Author and Coauthors at most in the participation of the Article and indicate in extensive the Institutional Affiliation indicating the Department.

Identify the Name of Author and Coauthors at most with the CVU Scholarship Number-PNPC or SNI-SECIHTI - Indicating the Researcher Level and their Google Scholar Profile to verify their Citation Level and H index.

Identify the Name of Author and Coauthors at most in the Science and Technology Profiles widely accepted by the International Scientific Community ORC ID - Researcher ID Thomson - arXiv Author ID - PubMed Author ID - Open ID respectively.

Indicate the contact for correspondence to the Author (Mail and Telephone) and indicate the Researcher who contributes as the first Author of the Article.

Plagiarism Detection

All Articles will be tested by plagiarism software PLAGSCAN if a plagiarism level is detected Positive will not be sent to arbitration and will be rescinded of the reception of the Article notifying the Authors responsible, claiming that academic plagiarism is criminalized in the Penal Code.

Arbitration Process

All Articles will be evaluated by academic peers by the Double Blind method, the Arbitration Approval is a requirement for the Editorial Board to make a final decision that will be final in all cases. MARVID® is a derivative brand of ECORFAN® specialized in providing the expert evaluators all of them with Doctorate degree and distinction of International Researchers in the respective Councils of Science and Technology the counterpart of SECIHTI for the chapters of America-Europe-Asia- Africa and Oceania. The identification of the authorship should only appear on a first removable page, in order to ensure that the Arbitration process is anonymous and covers the following stages: Identification of the Research Journal with its author occupation rate - Identification of Authors and Coauthors - Detection of plagiarism PLAGSCAN - Review of Formats of Authorization and Originality-Allocation to the Editorial Board-Allocation of the pair of Expert Arbitrators-Notification of Arbitration -Declaration of observations to the Author-Verification of Article Modified for Editing-Publication.

Instructions for Scientific, Technological and Innovation Publication

Knowledge Area

The works must be unpublished and refer to topics of Digital skills standards for education, Learning projects through the use of information, Technologies and communication, Development of digital, Competencies teaching digital skills programs, Management of technological and educational, Consulting fields of technological training, Applied to education and other topics related to Engineering and Technology

Presentation of the content

In the first article we present, *Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs* by Cabrera, Misael, Martinez, Ulises, Woocay, Arturo[c] and Valles, Delia, with adscription in the Instituto Tecnológico de Ciudad Juárez and Texas A&M University–Kingsville, in the next article we present *Development of a Tool for Academic Data Processing and Visualization: A Python-Based Proposal* by Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román, with adscription in the Universidad Autónoma de Campeche, in the next article we present *Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models* by González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén, with adscription in the Tecnológico Nacional de México, in the last article we present *Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study* by Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel, with adscription in the Universidad Autónoma del Estado de México.

Content

Journal of Computational Technologies 9[22]1-9: e1922109
Article

Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs

Marco metodológico para la predicción de la demanda eléctrica a corto plazo en regiones desérticas mediante lstm y variables meteorológicas

Cabrera, Misael*[a], Martinez, Ulises[b], Woocay, Arturo[c] and Valles, Delia[d]

[a] ROR Instituto Tecnológico de Ciudad Juárez • 0000-0003-1908-2985 • 1193564
[b] ROR Instituto Tecnológico de Ciudad Juárez • G -2083- 2018 • 0000-0002-1631-4448
[c] ROR Instituto Tecnológico de Ciudad Juárez • MVT -5006 -2025 • 0000-0001-9235-0494
[d] ROR Texas A&M University–Kingsville • D-3940-2013 • 0000-0002-9166-3244

SECIHTI classification:

Area: Technological sciences
Field: Computer technology
Discipline: Artificial intelligence
Subdiscipline: Energy Demand Forecasting

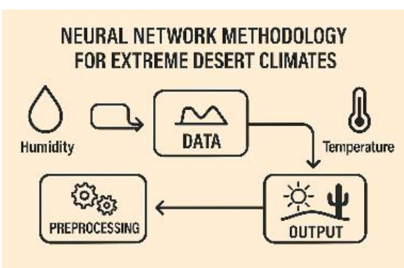
https://doi.org/10.35429/JOCT.2025.9.22.1.1.9
History of the article:
Received: August 30, 2025
Accepted: December 02, 2025

* ✉ [d13110928@cdjuarez.tecnm.mx]

Check for updates

Abstract

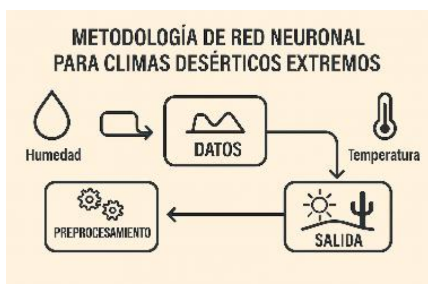
This study develops a model to predict electricity consumption peaks using artificial neural networks with continuous data recorded every 10 minutes. The research focuses on Ciudad Juárez, a desert region with high climatic variability, where temperature and humidity fluctuate substantially and unpredictably throughout the year. In such environments, temperatures can rise drastically during the day and drop sharply at night, creating unstable and irregular demand patterns. Sudden seasonal shifts and unexpected meteorological events further complicate electricity forecasting. Accurately predicting electricity demand in these conditions is essential for efficient energy management. Failure to forecast peaks may lead to shortages or excessive generation, with significant economic and environmental consequences. To address this, the model incorporates climatic variables such as time of day, day of the year, temperature, and relative humidity, which help capture the complex patterns of electricity use influenced by desert climate dynamics. Results show that including climatic factors substantially improves prediction accuracy, achieving 92% under validation. These findings demonstrate the potential of this approach to enhance energy planning in unstable climates, providing a reliable tool for sustainable resource management.



Neural Networks, Electricity Consumption, Desert Climates

Resumen

Este estudio desarrolla un modelo para predecir picos de consumo eléctrico utilizando redes neuronales artificiales con datos continuos registrados cada 10 minutos. La investigación se centra en Ciudad Juárez, una región desértica con alta variabilidad climática, donde la temperatura y la humedad fluctúan de manera sustancial e impredecible a lo largo del año. En estos entornos, las temperaturas pueden aumentar drásticamente durante el día y descender con fuerza por la noche, generando patrones de demandas inestables e irregulares. Los cambios estacionales repentinos y fenómenos meteorológicos inesperados complican aún más la predicción del consumo eléctrico. Predecir con precisión la demanda en estas condiciones resulta esencial para una gestión energética eficiente. No anticipar los picos puede ocasionar escasez o una generación excesiva, con consecuencias económicas y ambientales significativas. Para abordar este desafío, el modelo incorpora variables climáticas como la hora del día, el día del año, la temperatura y la humedad relativa, que permiten capturar los complejos patrones de uso eléctrico influenciados por la dinámica del clima desértico. Los resultados muestran que la inclusión de factores climáticos mejora sustancialmente la precisión de las predicciones, alcanzando un 92% bajo validación. Estos hallazgos demuestran el potencial de este enfoque para optimizar la planificación energética en climas inestables, proporcionando una herramienta confiable para la gestión sostenible de recursos.



Redes neuronales, Consumo eléctrico, Clima desértico

Area: Promotion of frontier research and basic science in all fields of knowledge

Citation: Cabrera, Misael, Martinez, Ulisesb, Woocay, Arturo and Valles, Delia. [2025]. Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs. Journal of Computational Technologies. 9[22]1-9: e1922109.

Introduction

Accurate forecasting of electricity consumption is essential for the efficient management of energy resources, especially in regions where demand exhibits significant fluctuations due to rapidly changing climatic conditions and where supply shortages are exacerbated by accelerated urbanization. Desert areas, such as Ciudad Juarez Chihuahua, present a particular challenge in modeling energy demand because meteorological conditions can vary dramatically from year to year, even within the same season. A clear example of this variability is the contrast between the historic snowfall recorded on October 27, 2020, and the mild weather observed in October 2024 in the same location, as shown in Table 1.

Box 1

Table 1
Climatic Comparison: October 27, 2020, vs. October 27, 2024 – Ciudad Juarez

Variable	27-Oct-2020 (Historic Snowfall)	27-Oct-2024 (Mild Weather)
Minimum Temperature (°C)	-2.0 °C	11.5 °C
Maximum Temperature (°C)	5.0 °C	23.0 °C
Average Temperature (°C)	1.5 °C	17.5 °C
Average Relative Humidity (%)	80%	32%
Average Wind Speed	15 km/h	8 km/h
Accumulated Precipitation	3.5 mm (snow)	0 mm
Estimated Electricity Consumption	~20% above autumn average	~4% below autumn average
Impact on Power Grid	Overload, residential outages	Normal operation

Source: National Meteorological Service (SMN)

1.1 Problem Statement and Central Hypothesis

Electricity consumption in desert regions is strongly linked to thermal comfort needs, which drive the use of heating systems, refrigerators, evaporative coolers, and mini-split air conditioners. Climatic variables such as temperature and humidity directly influence household decisions, producing non-linear demand patterns during both summer heatwaves and winter cold spells.

In hot-dry climates such as Ciudad Juárez, evaporative coolers traditionally provide cost-effective cooling, achieving 60–85% effectiveness and saving up to ~75% of electricity compared to compression-based A/C units. However, their performance declines under humid conditions, prompting a switch to mini splits that sharply increase energy use (Haile *et al*., 2024; Alfraidi *et al*., 2024).

Cold spells, though less frequent, also increase electricity consumption through heating loads from resistive devices and heat pumps in mini splits. Research indicates that both heatwaves and cold events raise outage frequency and duration, highlighting resilience challenges (Liang *et al*., 2025).

Additionally, urban form modifies cooling needs: studies in desert cities demonstrate that vertical densification can reduce residential electricity demand by improving shading and decreasing exposed building surface (Lopez *et al*., 2024).

For Ciudad Juárez, this evidence justifies the integration of meteorological variables and comfort-related metrics such as Cooling Degree Days (CDD) and Cooling Degree Hours (CDH) into the forecasting model. Incorporating these indicators enhances the LSTM's capacity to capture the true drivers of energy consumption in desert regions.

Multiple studies have demonstrated that incorporating climatic variables such as temperature and humidity significantly improves the accuracy of electricity consumption prediction models. Based on the article of combined an Enhanced Inception V4 model with an Osprey optimizer to account for environmental factors (Chen, Fang *et al*., 2024).

This annex presents real and projected data that reinforce the necessity of predictive models based on neural networks for Ciudad Juárez, located in the Gran Consumo Regional (GCR) Norte region of the National Electric System.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Box 2

Table 2
Simulation of increase of electricity consumption from 2019 to 2025

Year	Electricity Consumption (GWh)	Year-on-Year Growth (%)
2019	28,868	N/A
2020	29,291	1.47%
2021	30,378	3.71%
2022	31,850	4.85%
2023	34,200	7.38%
2024	36,483	6.68%
2025	40,000	9.64%

Source: PRODESEN 2025 and simulations for regional analysis

Patterns of hourly and annual electricity consumption growth are presented in the following tables. These patterns demonstrate irrational and non-linear increases that challenge conventional prediction methodologies.

Table 2 illustrates the evolution of temperature and humidity in Norte region between 2019 and 2025, highlighting a sustained growth trend over six years. This progression demonstrates that consumption patterns in Ciudad Juarez are strongly influenced by climatic variability, which does not always follow predictable seasonal cycles. The evidence confirms that models relying solely on historical consumption records and calendar dates are insufficient to accurately forecast demand. Incorporating climatic variables and adaptive prediction methods is essential to reflect the true dynamics governing electricity consumption in this region.

Box 3

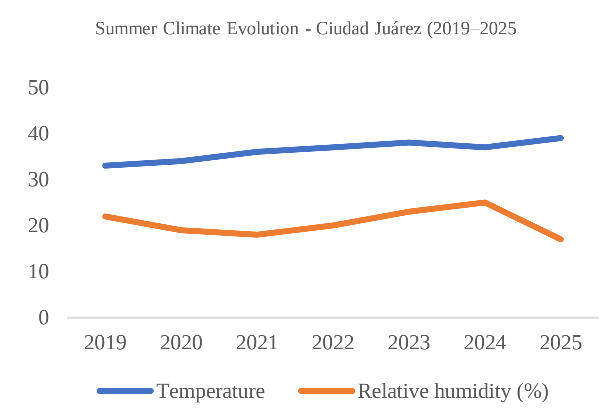


Figure 1
Summer Climate Evolution - Ciudad Juárez (2019–2025
Source: PRODESEN and simulations for projected behavior

The following comparison shows the changes in average temperature and relative humidity during the summers of 2019 to 2025. The evidence demonstrates a sustained increase and substantial variability in environmental conditions, further highlighting the importance of adaptive models such as neural networks to anticipate the impact of climate on electricity consumption.

Box 4

Table 3
Climatic Comparison – Ciudad Juárez Summers 2019–2025

Year	Average Summer Temperature (°C)	Average Relative Humidity (%)
2019	33.2	22
2020	34.1	19
2021	35.7	18
2022	36.5	20
2023	37.8	23
2024	36.9	25
2025	39.0	17

Source: Developed with data from CONAGUA, historical local records, and recent trend simulations

This study presents the development and implementation of a predictive model to anticipate short-term electricity consumption peaks using Artificial Neural Networks (ANNs), specifically a Long Short-Term Memory (LSTM) network. The model is trained using continuous time-series data recorded at 10-minute intervals, allowing the system to learn from fine-grained fluctuations in electrical demand. The main objective is to build a reliable tool for short-term load forecasting (STLF), which is critical for optimizing grid stability, energy distribution, and operational planning in smart grid systems (Hong & Fan, 2016).

Compared to traditional time series forecasting techniques, LSTM networks offer significant advantages over other Recurrent Neural Network (RNN) architectures. Conventional RNNs often suffer from the vanishing gradient problem when dealing with long sequences, limiting their effectiveness in capturing long-range dependencies. In contrast, LSTMs are specifically designed to retain relevant information over extended periods through the use of input, forget, and output gates (Hochreiter & Schmidhuber, 1997).

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Cabrera, Misael, Martinez, Ulisesb, Woocay, Arturo and Valles, Delia. [2025]. Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs. Journal of Computational Technologies. 9[22]1-9: e1922109.
https://doi.org/10.35429/JOCT.2025.9.22.1.1.9

This capability makes them particularly suitable for modeling the complex temporal dynamics of energy consumption, which can be influenced by nonlinear and delayed climatic patterns such as temperature, humidity, and solar radiation (Marino *et al*., 2016). As a result, LSTMs demonstrate greater robustness, training stability, and generalization performance in real-world scenarios with high variability and noise.

Methodology

Problem Definition

The central problem addressed in this study is the high unpredictability of electricity consumption in desert regions, where extreme climatic fluctuations disrupt conventional demand patterns. In cities such as Ciudad Juárez, rapid changes in temperature, relative humidity, wind speed, and precipitation drive abrupt shifts in household cooling and heating needs. These dynamics produce nonlinear and highly volatile consumption profiles that traditional statistical and time-series forecasting models—primarily dependent on historical load records and seasonal trends—struggle to capture accurately.

Traditional models tend to assume stationarity and smooth seasonality, which makes them poorly suited for environments where energy demand is shaped by sudden heatwaves, cold spells, and monsoonal humidity surges. As a result, forecasts often underestimate peak loads during extreme events and overestimate demand during mild conditions, undermining both grid reliability and resource planning.

To address this gap, this study hypothesizes that explicitly integrating climatic variables—with emphasis on temperature and humidity, complemented by wind speed and precipitation—into a Long Short-Term Memory (LSTM) neural network can significantly enhance predictive accuracy. Unlike traditional models, LSTM architectures are capable of learning long-range temporal dependencies and nonlinear interactions, making them well suited to capture the complex interplay between thermal comfort needs and electricity consumption in desert climates.

General Objective

The objective of this study is to develop and validate a predictive model capable of accurately forecasting short-term electricity consumption in desert regions with high climatic variability. The model produces 10-minute interval predictions with a forecasting horizon of up to 4 hours, providing sufficient resolution to support both real-time grid operation and demand-side management strategies.

The modeling framework is based on artificial neural networks (ANNs), with emphasis on a Long Short-Term Memory (LSTM) architecture, chosen for its ability to capture nonlinear relationships and long-term temporal dependencies in electricity consumption data. The model incorporates as input variables:

– **Hour of the day** (temporal dynamics of daily load patterns)
– **Ambient temperature** (thermal comfort needs driving cooling/heating loads)
– **Relative humidity** (modulator of cooling technology efficiency and household energy choices)

By integrating these temporal and climatic variables, the model aims to represent the complex interaction between human comfort requirements and electricity demand. Ultimately, this approach contributes to improving forecasting accuracy in arid and semi-arid regions, enhancing the resilience and efficiency of electricity grid management.

Experimental Data Acquisition

To validate the forecasting model under realistic operating conditions, an initial experimental dataset was collected from a typical residential household. The measurement setup consisted of two main components:

1. Electricity consumption monitoring

A Shelly EM smart meter was installed on the household's electrical system to provide high-resolution measurements of energy usage. The device was configured to record electricity consumption at minute-level intervals, allowing the capture of both short-term fluctuations (e.g., appliance switching events) and longer-term patterns (e.g., cooling and heating cycles).

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Cabrera, Misael, Martinez, Ulisesb, Woocay, Arturo and Valles, Delia. [2025]. Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs. Journal of Computational Technologies. 9[22]1-9: e1922109.
https://doi.org/10.35429/JOCT.2025.9.22.1.1.9

2. Indoor environmental monitoring

To account for the effect of thermal comfort variables on electricity demand, a Groove temperature and humidity sensor was placed inside the residence. This sensor continuously measured ambient temperature (°C) and relative humidity (%), ensuring that variations in indoor environmental conditions could be correlated with household energy consumption.

By combining these two datasets, it was possible to establish a direct relationship between user comfort requirements (driven by temperature and humidity) and the corresponding electricity consumption behavior. This integrated dataset provides a reliable foundation for validating the proposed model, since it reflects the actual interaction between environmental stressors, comfort needs, and residential energy demand.

Data Integration

A custom data aggregation program was developed to synchronize, preprocess, and compile all sensor readings into a unified dataset. This program automatically merged consumption data from the Shelly EM with environmental records from the Groove sensor, creating a clean and chronologically ordered datasheet. This unified dataset was then prepared for model training and validation.

Model Architecture and Rationale for LSTM

The model architecture was implemented in Python using the TensorFlow and Keras libraries. The design consists of:

- **Input Layer:** Accepts the three normalized variables (hour, temperature, humidity).

- **Two stacked LSTM layers:** These layers were specifically chosen because LSTM networks excel at learning long-term dependencies in time series data, which is critical when past temperature and humidity exert delayed effects on consumption.

- **Dense output layer:** Produces continuous consumption predictions.

- **Dropout layers:** Applied to reduce overfitting.

The rationale for selecting the LSTM architecture over other neural network types lies in its proven capability to retain historical information over extended sequences, making it especially suitable for modeling recurrent consumption patterns influenced by climatic cycles. Unlike simpler Recurrent Neural Networks (RNN), LSTM avoids vanishing gradients and maintains performance even when dependencies span several hours.

Data Preparation

All recorded data were normalized to zero mean and unit variance, ensuring stable gradient descent and comparable feature scales. The dataset was split chronologically into:

- 70% training
- 15% validation
- 15% testing

This chronological partition prevents data leakage and simulates the operational scenario of predicting unseen future consumption.

Training and Optimization

The model was trained using the Mean Squared Error (MSE) as the loss function and the Adam optimizer with adaptive learning rates. To improve generalization, early stopping and learning rate decay were applied based on validation loss.

Evaluation Metrics

Performance was assessed through the coefficient of determination (R^2), calculated as:

$$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2] / [\Sigma(y_i - \bar{y})^2]$$

where:
y_i: Actual consumption
$\hat{y}_i$: Predicted consumption
$\bar{y}$: Mean consumption

This metric quantifies the variance explained by the model, supporting robust comparison with alternative architectures.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Cabrera, Misael, Martinez, Ulisesb, Woocay, Arturo and Valles, Delia. [2025]. Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs. Journal of Computational Technologies. 9[22]1-9: e1922109.
https://doi.org/10.35429/JOCT.2025.9.22.1.1.9

Comparative Analysis of Neural Network Architectures

Box 5

Table 4
Comparative Analysis of Neural Network Architectures

Neural Network Type	Main Advantages	Main Disadvantages	Ideal For
MLP (Multilayer Perceptron)	Fast training; suitable for tabular data; simple to deploy.	Cannot model sequential dependencies; sensitive to scaling.	Point predictions with static variables (e.g., time, weather).
RNN	Captures temporal dependencies; handles sequential inputs.	Degrades over long sequences; harder to optimize.	Continuous time series forecasting.
LSTM	Retains long-term memory; robust to sequence length.	Higher computational cost; requires more data.	Predicting consumption with strong historical dependence.
GRU	Simpler than LSTM; faster convergence.	Lower accuracy on complex datasets.	Low-resource applications.
CNN	Detects local patterns; effective with spatially structured inputs.	Not inherently sequential; requires reshaping input data.	Multivariate datasets with spatial dimensions.
Transformer	Models' long-range dependencies; highly scalable.	High computational demands; prone to overfitting with few samples.	Large-scale meteorological or extended regional consumption datasets.

Results

Box 6

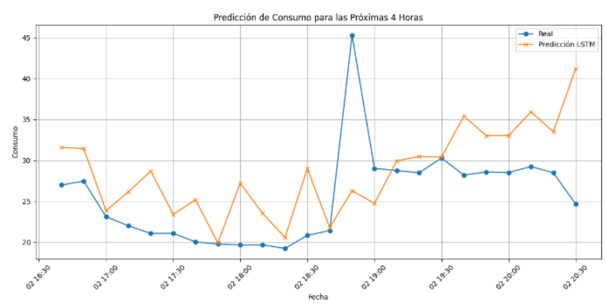


Figure 2
Comparison between the actual electricity consumption values and the predictions generated and real values

The graph presented shows the comparison between the actual electricity consumption values and the predictions generated by the LSTM model for a 4-hour forecast horizon. As more historical data has been incorporated into the model training, a progressive improvement in predictive capability has been observed. At this stage of the research, the training dataset has been fed with a total of 17,280 records, corresponding to 10-minute intervals over a 4-month period.

In the graph, the actual electricity consumption values are represented in blue, while the predicted values generated by the LSTM model are shown in orange. This visual differentiation helps to assess the alignment and divergence between the real and forecasted series.

The model has achieved an average error margin (to be completed with the average from the error table), representing a confidence level of approximately [verify how the confidence level is expressed in %], depending on the metric evaluated. The trend shows that as the volume of input data increases, the model improves its fit, reducing deviation from actual values and demonstrating greater stability in the face of unexpected consumption spikes.

This behavior is consistent with the adaptive nature of the LSTM model, which benefits from sequential learning to identify nonlinear temporal patterns. Although notable deviations are still observed in the face of abrupt events—as evidenced by the spike located in the center of the series—the model increasingly responds more accurately to the general behavior of the system, showing a positive convergence trend.

Conclusions and recommendations

The LSTM model has proven to be a promising architecture for electricity consumption forecasting in desert environments, characterized by high thermal variability and extreme climatic conditions that significantly impact the stability of energy demand. Although the current results have not yet reached a sufficiently low error threshold to guarantee full operational reliability, this limitation is mainly attributed to the insufficiency of training data.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

It is anticipated that by increasing the amount and diversity of input data, the model will converge toward an optimal point at which the error margin will stabilize within acceptable ranges, thereby eliminating the need for parallel comparative analyses and enabling the direct generation of highly accurate predictions.

Additionally, to mitigate the effects of prediction errors during sudden consumption spikes, it is recommended to incorporate value-weighting mechanisms in the model's loss function. This would allow the model to assign greater importance to peak deviations, enhancing its ability to stabilize forecasts under high-variance conditions.

This research not only establishes a solid methodological foundation for future developments in energy forecasting but also opens the way for critical applications in contexts of energy vulnerability. In desert areas where electrical systems face resource scarcity, limited infrastructure, and frequent overloads, accurate prediction of voltage peaks can be used to:

- Optimize the dispatch of energy from renewable sources and temporary storage systems (such as batteries or flywheels), improving the efficiency of the electrical system.
- Prevent service interruptions by implementing real-time preventive measures upon detecting an imminent increase in demand.
- Support flexible demand strategies by activating demand response programs or load regulation in non-priority sectors.
- Improve electrical infrastructure planning, enabling smart grid expansion based on historical data and robust forecasts.
- Facilitate the integration of microgrids or distributed generation, better coordinating the interaction between local generators and the main grid.

Overall, this approach represents a step forward toward more resilient, adaptive, and intelligent electrical systems capable of proactively responding to the specific energy challenges of arid regions, where the balance between supply and demand becomes critically delicate.

Conflict of Interest Statement

The authors declare no conflict of interest. They have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Author Contributions Statement

Cabrera, Misael: I contributed with the conceptualization, methodology, and original draft writing. I also prepared figures and the initial version of the manuscript.

Martínez, Ulises: I contributed with data curation, software implementation, and validation. I also optimized neural networks and ensured reproducibility.

Woocay, Arturo: I contributed with supervision, methodological review, and editing. I ensured scientific rigor and improved the clarity of the manuscript.

Valles, Delia: I contributed with project administration, funding acquisition, and industrial engineering guidance. I aligned the neural network methodology with energy management strategies and practical applications.

Availability of Data and Materials

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

Funding

This doctoral research was supported by a grant from SECIHTI through the national postgraduate program. Additional institutional support was provided by the Instituto Tecnológico de Ciudad Juárez, Instituto Nacional de México, which facilitated resources and academic infrastructure for the development of the study.

Acknowledgement

The research was made possible thanks to the continuous support of the Instituto Tecnológico de Ciudad Juárez for providing academic space, computing facilities, and guidance throughout the project.

Special acknowledgement is also extended to the academic advisors and collaborators whose expertise in industrial engineering, climate analysis, and neural networks enriched the development of this methodology for electricity demand prediction in desert regions.

Abbreviations

CNN	Convolutional Neural Network
Gcr	
GRU	Gated Recurrent Unit
GWh	Giga Watts per hour
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
PRODECEN	Programa de Desarrollo del Sistema Eléctrico Nacional
RNN	Recurrent Neural Network

References

Antecedents

De Felice, M., Alessandri, A., & Ruti, P. M. (2015). Electricity demand forecasting over Italy: The role of temperature and climate change. Energy, 81, 618–628.

Basics

Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hong, T. & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, 32(3), 914–938.

Supports

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., & Xu, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Transactions on Smart Grid, 10(1), 841–851.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using Deep Neural Networks. IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, 7046–7051.

Differences

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030–1081.

Discussions

Chen, J., Fang, L., & Khayatnezhad, M. (2024). Short-term electricity load forecasting using enhanced Inception-V4 and Osprey optimization. Sustainable Energy Technologies and Assessments, 59, 103504.

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, 112896.

Suggested References

Elmousalami, H., Peng Hui, F. K., & Alnaser, A. A. (2025). Enhancing Smart and Zero-Carbon Cities Through a Hybrid CNN-LSTM Algorithm for Sustainable AI-Driven Solar Power Forecasting (SAI-SPF). Buildings, 15(15), 2785.

Kishore, B., Kabilan, K., Rahul, L. S., Priyadarshan, G. P., Vishnutheerth, E. P., Satheesh, R., & Kolhe, M. L. (2025). Advancing short-term wind power forecasting by AI-driven models for improved accuracy. Electrical Engineering, 1-16.

Ladjal, B., Nadour, M., Bechouat, M., Hadroug, N., Sedraoui, M., Rabehi, A., ... & Agajie, T. F. (2025). Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria. Scientific Reports, 15(1), 15404.

Nandigam, S. H., Nageswararao, K., & Sharma, P. K. (2025). Hybrid Deep Learning Models for Energy Consumption Forecasting: A CNN-LSTM Approach for Large-Scale Datasets. Journal of Renewable Energy and Smart Grid Technology, 20(2), 82-91.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Zhang, L., Liu, L., Chen, W., Lin, Z., He, D., & Chen, J. (2025). Photovoltaic Power Generation Forecasting Based on Secondary Data Decomposition and Hybrid Deep Learning Model. Energies, 18(12), 3136.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Journal of Computational Technologies 9[22]1-8: e2922108
Article

Development of a Tool for Academic Data Processing and Visualization: A Python-Based Proposal

Desarrollo de una herramienta para el Procesamiento y Visualización de Datos Académicos: Propuesta Basada en Python

Salazar-Uitz, Ricardo Rubén[a], Ramos-Ramos, Rudy Mauricio[b], Shih, Meng Yen*[c] and Lezama-Zarraga, Francisco Román[d]

[a] Universidad Autónoma de Campeche • AGW-7002-2022 • 0000-0003-2307-737X • 416277
[b] Universidad Autónoma de Campeche • NVM5897-2025 • 0009-0007-2506-7104 • 2150106
[c] Universidad Autónoma de Campeche • AAB-4875-2021 • 0000-0001-7475-6458 • 408617
[d] Universidad Autónoma de Campeche • U-1229-2018 • 0000-0003-3397-7881 • 205493

SECIHTI classification:

Area: Engineering
Field: Technological Sciences
Discipline: Computer Technology
Subdiscipline: Programming languages

https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

History of the article:
Received: August 22, 2025
Accepted: December 15, 2025

* ✉ [smengyen@uacam.mx]

Abstract

This work addresses the challenge of processing student data from multiple .CSV files generated by an Online Preparatory Course. Due to the large number of students and recorded activities, manual processing typically requires several days of work. To solve this, a Python-based tool was developed to efficiently process and visualize the data from these files. The tool is user-friendly and adaptable for academic data analysis, particularly focusing on managing large datasets such as course activity lists. The main outcome of this development is the automation of data loading, processing, and normalization, significantly reducing the time and effort required—from several days to just a few minutes. It provides visual comparisons of student performance, enabling the analysis of diagnostic test results, progress through course activities, and final exam outcomes.

Objectives	Methodology	Contribution
Automate the processing of academic data.	Tool developed in Python using pandas, openpyxl, and tkinter.	The developed tool: Saves time by automating data analysis.
Integrate and visualize performance from Khan Academy and Google Forms.	Modular architecture with three controllers.	Reduces errors in data handling. Facilitates the diagnosis of student performance.
Support evidence-based pedagogical decisions.	Processes CSV files and spreadsheets.	Improves teachers' decision-making.

Academic Data, Python, Processing

Resumen

El presente trabajo da solución a la problemática de procesar los datos en diversos formatos .CSV de un Curso Propedéutico en Línea, con los resultados de los estudiantes, el cual requiere varios días de trabajo manual por la cantidad de alumnos y actividades registradas. Por lo que se desarrolló una herramienta basada en Python para procesar y visualizar datos de estos archivos, fácil de usar y adaptable para el análisis de estos. Se enfoca principalmente en gestionar gran cantidad de datos, como son las listas de actividades académicas que ofrece el curso. El resultado de esta herramienta desarrollada es la automatización de la carga, el procesamiento, la normalización y la reducción drástica del esfuerzo y tiempo invertido de días a unos cuantos minutos, presentando los resultados de los estudiantes permitiendo comparar el examen diagnóstico, el avance en las actividades del curso y el examen final.

Objetivos	Metodología	Contribución
Automatizar el procesamiento de datos académicos.	Herramienta desarrollada en **Python** con *pandas*, *openpyxl* y *tkinter*.	La herramienta desarrollada: Ahorra tiempo al automatizar el análisis.
Unificar y visualizar el rendimiento de *Khan Academy* y *Google Forms*.	Arquitectura **modular** con tres controladores.	Reduce errores en el manejo de datos. Facilita el diagnóstico de rendimiento estudiantil.
Apoyar decisiones pedagógicas basadas en evidencia.	Procesa archivos **CSV** y hojas de cálculo.	Mejora la toma de decisiones docentes.

Datos Académicos, Python, Procesamiento

Area: Development of strategic leading-edge technologies and open innovation for social transformation

Citation: Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.

Introduction

The processing of educational data constantly faces challenges due to the tedious nature of manually handling information from platforms such as *Khan Academy* and assessments conducted through *Google Forms* (Salazar-Uitz *et al*., 2022).

This task, which must be performed frequently, can take several days or even weeks due to the diversity of formats, lack of structure, and the need to combine metrics from different sources using software such as Excel (Estrada Arjona, 2025). Each time it becomes necessary to analyze students' progress or generate reports for teachers, excessive time must be spent cleaning, organizing, and comparing data. This process is not only inefficient but also delays important pedagogical decisions.

Given this situation, a computational tool in Python was developed to automate this entire process and shorten the analysis time (Guindulain Lebrero, 2015). The proposed solution allows for the immediate processing, unification, and visualization of student performance data from Khan Academy along with results obtained in Google Forms. What previously required days of manual work can now be completed in minutes, without human intervention.

The tool consists of three main modules: one specialized in analyzing data from Khan Academy, another focused on processing results from formal evaluations, and a visualization engine that allows users to compare both data sets clearly and effectively for decision-making. Moreover, it was designed with non-technical educators in mind, making it easy to use and promoting a culture of accessible educational analysis.

This tool arises from a real need to transform a slow and tedious process into one that is agile, automated, and useful. It aims to establish efficiency between digital learning and formal evaluation, while improving how teachers make educational decisions based on data (Estrada Arjona, 2025). According to some studies, having a program that performs these processes automatically can be extremely useful for both teachers and students (Leyva López *et al*., 2018).

Methodology

Using academic data from the Khan Academy preparatory course, a new automatic tool was created and tested to assist in data processing and analysis. The goal was to improve the data processing time and visualize the results generated by the application to devote more effort to interpretation.

The test data used correspond to first-year students enrolled in the *Software Engineering* degree for the 2024–2025 academic year who participated in the online preparatory course. This course has been implemented annually at the Faculty of Engineering of the *Universidad Autónoma de Campeche* since 2018 (Canto-Canul *et al*., 2020) To ensure analysis quality, only students who completed all three types of evaluations were included:

– Activities in Khan Academy (Online Preparatory Course)
– Initial Diagnostic Test (Google Forms)
– Final Exam (Google Forms)

The final sample consisted of 14 students with complete data. The data were obtained directly from the platforms in CSV format, exporting files from Khan Academy containing information for each student. In the current year, 167 activities were considered (Salazar-Uitz *et al*., 2022), Additionally, spreadsheets automatically generated by Google Forms provided information on diagnostic test results (before the course) and final exam results (after the course).

The data processing was carried out on computers running Windows 10 and 11 operating systems. The tool was developed in the Python programming language (version 3.9 or higher), using libraries such as pandas for data manipulation and openpyxl for reading and writing Excel files. The graphical interface was built with tkinter, providing a user-friendly interaction experience.

The tool is composed of three main modules:

– FileController: Processes data exported from Khan Academy.
– GoogleFormController: Handles the analysis of results from Google Forms assessments.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

– GraficacionController: Automatically generates comparative visualizations.

Box 1

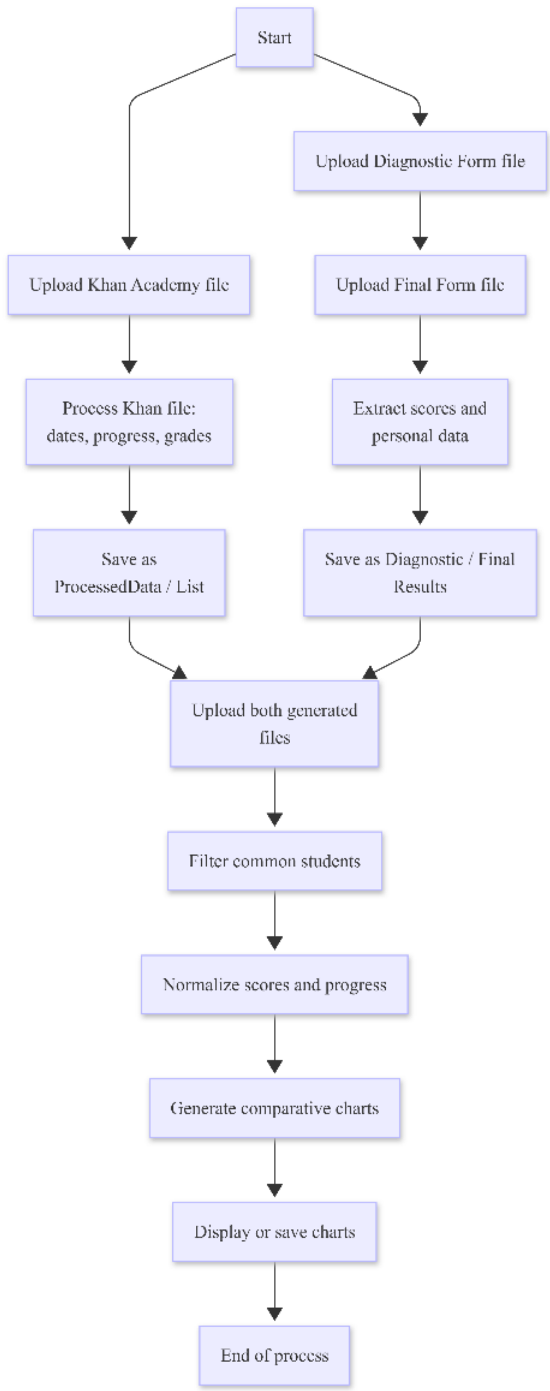


Figure 1

General Process Flow Diagram of the Tool for Academic Data Processing

As shown in Figure 1, the tool was designed with a modular approach, allowing the separation of processing responsibilities by platform (Khan Academy and Google Forms), as well as an additional module for graph generation. Each module has an interface aligned with the user's academic workflow, facilitating file loading, previewing, processing, and exporting of results.

The first component is the FileController, responsible for processing the data exported from the Khan Academy platform. Through a Tkinter-based graphical interface, this module allows the user to select an Excel file and, from it, extract the tasks assigned to students, the completion date for each task, and, optionally, the scores obtained.

Box 2

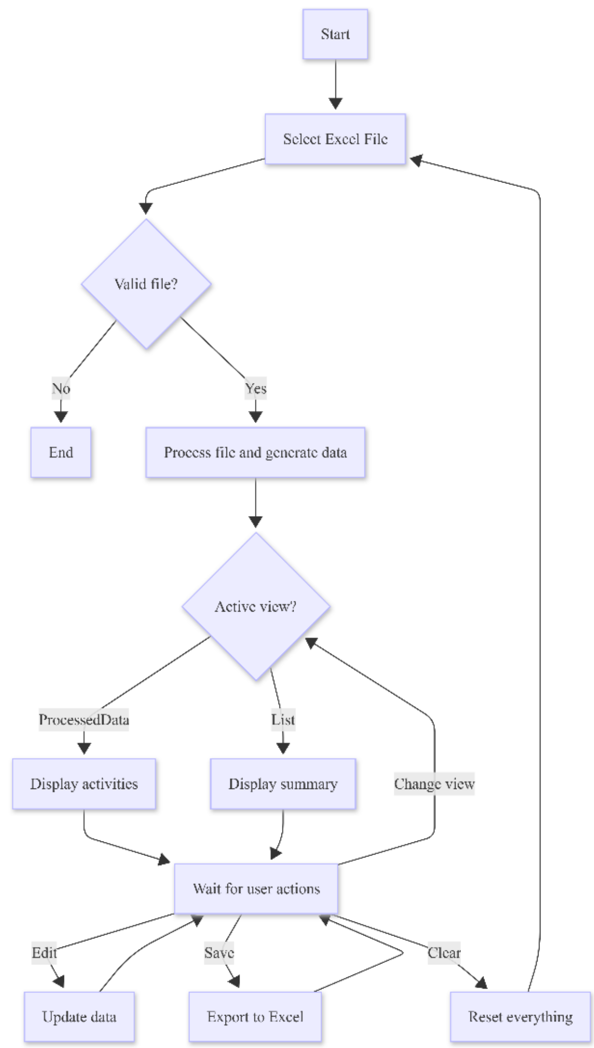


Figure 2

Flow Diagram of the FileController Module: Processing Khan Academy Data

As observed in Figure 2, once the file is loaded a structured table is generated where each row represents an activity and each column corresponds to a student. Cells contain submission dates or are left blank if the activity has not been completed. Subsequently, a second dataset ("Listado") is created, which summarizes each student's performance and includes:

– Number of completed activities.
– Percentage of progress.
– Total grade percentage (if available in the file).

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

The module also allows switching between these two views and exporting the results to a new Excel file with automatic formulas to compute per-student progress percentages. The second component, named GoogleFormController, is responsible for processing two files coming from Google Forms: one corresponding to an initial diagnostic exam applied at the start of the course and another corresponding to the final exam.

Box 3

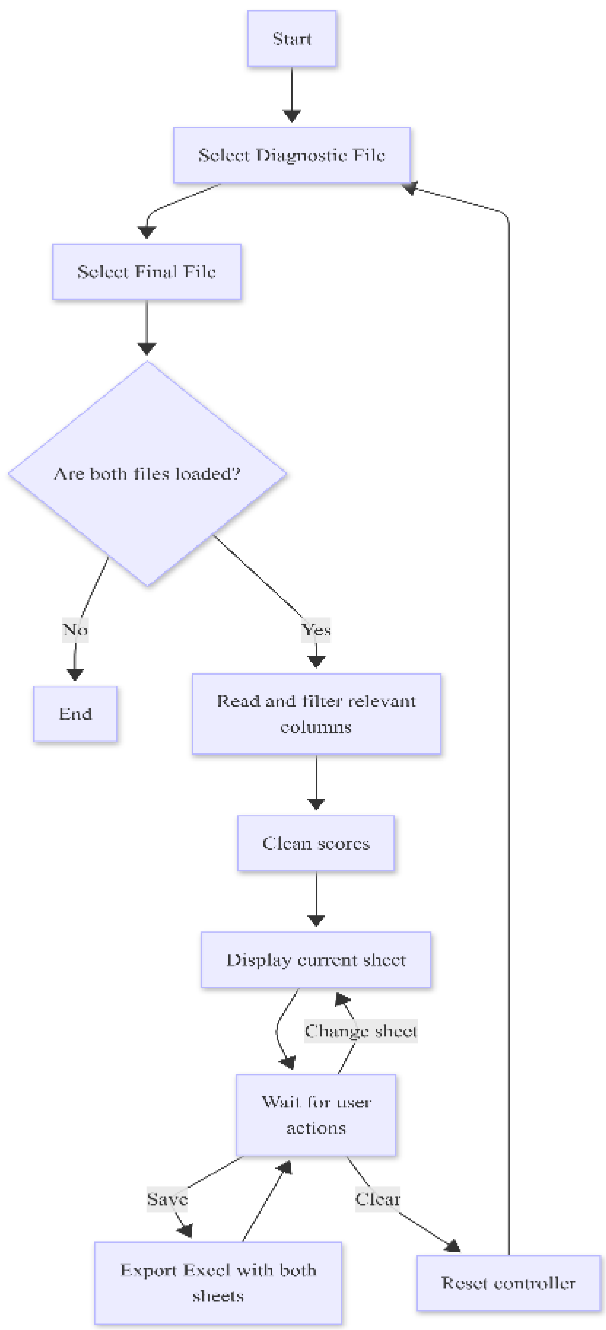


Figure 3
Flow Diagram of the GoogleFormController Module: Processing Google Forms Assessment Results.

As shown in Figure 3, the program requests both files and extracts the following relevant fields:

– Username.
– Student's first and last names.
– Score obtained (automatically detected by column name).

During preprocessing, any additional text in score cells is removed, retaining only the numeric value. This allows viewing both diagnostic and final exam results alternately and saving both datasets into a new Excel workbook with separate sheets. The third module, GraficacionController, integrates the datasets processed by the previous two modules to generate graphical representations of student performance. As shown in Figure 4, it requests the files previously produced by the Khan Academy and Google Forms modules. This module identifies common students across both datasets by matching emails or usernames.

Box 4

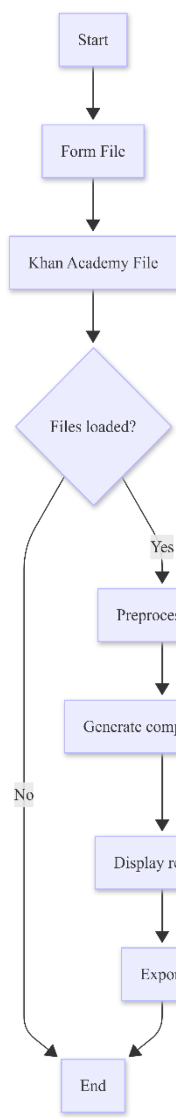


Figure 4
Flow Diagram of the GraficacionController Module: Integration and Generation of Comparative Visualizations

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

Once matching students have been identified, scores are normalized (adjusted to a 0–100 range) and four types of charts are produced:

– General comparison: diagnostic, Khan Academy progress, Khan Academy grade, and final exam.
– Progress vs. Khan grade: relationship between progress and obtained grade.
– Individual progress: trend line per student.
– Score distribution: comparative boxplot of all sources.

These visualizations facilitate a comprehensive interpretation of student performance throughout the course, both at individual and group levels.

Results

The developed tool was evaluated using real datasets obtained from the Khan Academy platform and assessment forms created with Google Forms. Its main objective was to consolidate, process, and visualize students' academic progress based on three key stages: the diagnostic evaluation, performance throughout the course, and the final evaluation. The application consists of three main modules, each accessible through a central interface designed to facilitate navigation between functionalities.

Box 5

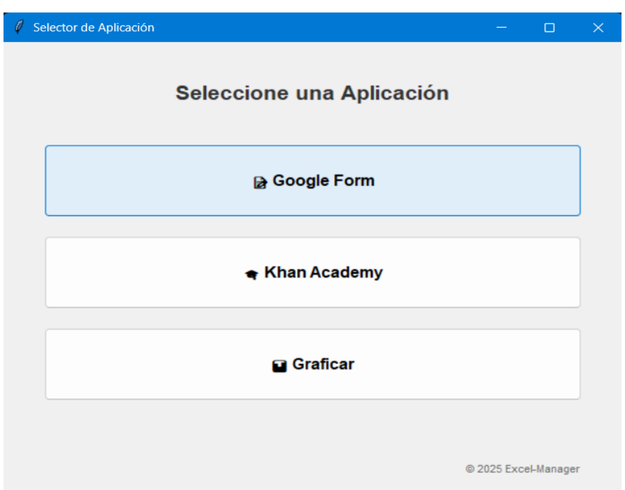


Figure 5
Main Graphical Interface for Module Selection in the Tool for Academic Data Processing and Visualization.

Figure 5 shows the program's initial screen, where the user can select the desired module.

This figure displays the main graphical interface designed to make navigation and access to the tool's functionalities simple. Users can choose one of the three main modules: "Google Form," "Khan Academy," or "Graph." The modular structure and intuitive interface were specifically designed for teachers without technical experience to use the tool effectively, thus promoting a culture of accessible educational data analysis.

Figure 6 presents the interface of the module dedicated to processing data exported from Khan Academy, known as FileController. This module displays students' performance in the assigned activities, including their level of completion and progress in the online preparatory course.

Box 6

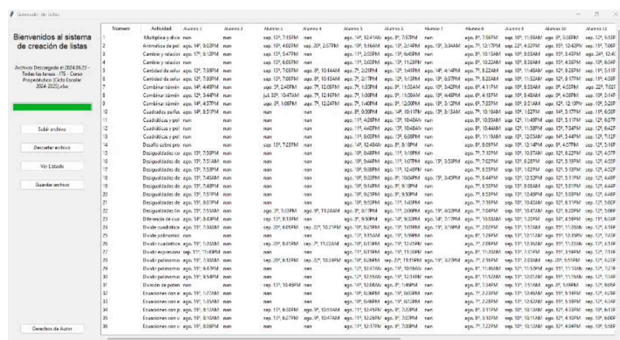
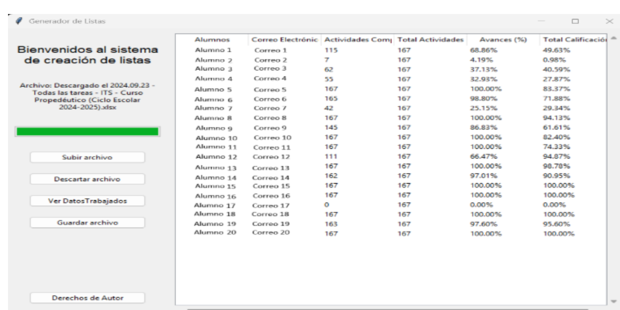


Figure 6
FileController Module Interface for Visualizing Progress and Completion of Khan Academy Activities.

The module processes the data and generates a structured table summarizing the number of activities completed and the percentage of progress (see Figure 7). This feature allows teachers to quickly evaluate both individual and group progress on the Khan Academy platform.

Box 7



Alumnos	Correo Electrónic	Actividades Com	Total Actividades	Avances (%)	Total Calificació
Alumno 1	Correo 1	115	167	68.86%	49.63%
Alumno 2	Correo 2	7	167	4.19%	0.98%
Alumno 3	Correo 3	62	167	37.13%	40.59%
Alumno 4	Correo 4	55	167	32.93%	27.87%
Alumno 5	Correo 5	167	167	100.00%	83.37%
Alumno 6	Correo 6	165	167	98.80%	71.88%
Alumno 7	Correo 7	42	167	25.15%	29.34%
Alumno 8	Correo 8	167	167	100.00%	94.13%
Alumno 9	Correo 9	145	167	86.83%	61.61%
Alumno 10	Correo 10	167	167	100.00%	82.40%
Alumno 11	Correo 11	167	167	100.00%	74.33%
Alumno 12	Correo 12	111	167	66.47%	94.87%
Alumno 13	Correo 13	167	167	100.00%	98.78%
Alumno 14	Correo 14	162	167	97.01%	90.95%
Alumno 15	Correo 15	167	167	100.00%	100.00%
Alumno 16	Correo 16	167	167	100.00%	100.00%
Alumno 17	Correo 17	0	167	0.00%	0.00%
Alumno 18	Correo 18	167	167	100.00%	100.00%
Alumno 19	Correo 19	163	167	97.60%	95.60%
Alumno 20	Correo 20	167	167	100.00%	100.00%

Figure 7
Structured Table View of Completed Activities and Progress in the FileController Module

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

Figure 8 illustrates the graphical interface of the GoogleFormController module once the data filtering and analysis process has been completed. This module processes the results of the diagnostic and final exams administered through Google Forms, extracting and organizing the grades for each student into an easily interpretable format.

Box 8

Nombre de usuario	Nombres:	Apellidos:	PuntuaciÃ³n total
correo 1	nombre 1	apellido 1	15.00
correo 2	nombre 2	apellido 2	65.00
correo 3	nombre 3	apellido 3	15.00
correo 4	nombre 4	apellido 4	65.00
correo 5	nombre 5	apellido 5	70.00
correo 6	nombre 6	apellido 6	40.00
correo 7	nombre 7	apellido 7	5.00
correo 8	nombre 8	apellido 8	80.00
correo 9	nombre 9	apellido 9	70.00
correo 10	nombre 10	apellido 10	45.00
correo 11	nombre 11	apellido 11	65.00
correo 12	nombre 12	apellido 12	60.00
correo 13	nombre 13	apellido 13	30.00

Figure 8
GoogleFormController Module Interface Displaying Normalized Results After Assessment Processing.

Figure 9 shows the main interface of the GraficacionController module, where the user must load the previously generated files from the Khan Academy and Google Forms modules. This step is crucial, as the module integrates and compares both datasets to identify common students and generate comparative charts of their academic performance.

Box 9

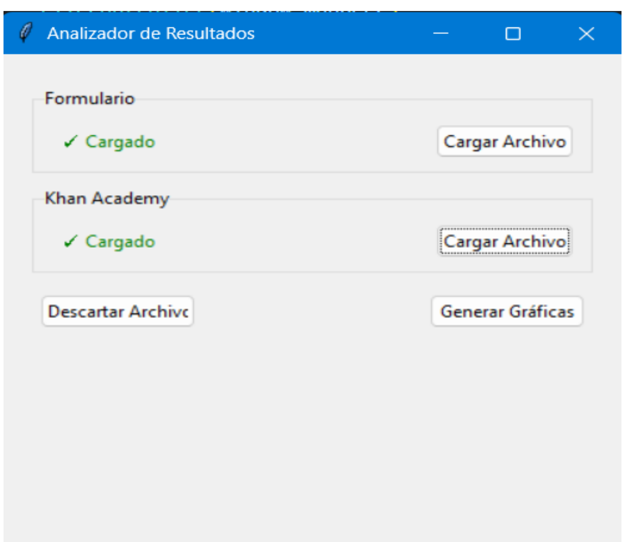


Figure 9
File Upload Interface for the Graphing and Comparative Analysis Module (GraficacionController).

Figure 10 presents a general comparative chart generated by the tool that visualizes student performance across three key stages: the initial diagnostic test, progress and performance in the Khan Academy course, and the final exam. All scores are normalized to a 0–100 range for consistent comparison. This type of visualization facilitates a comprehensive interpretation of academic performance over time, allowing the identification of performance patterns and overall learning evolution.

Box 10

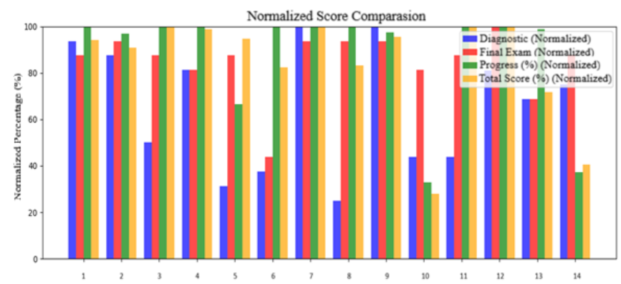


Figure 10
General Comparative Chart of Student Performance: Diagnostic, Khan Academy Progress, and Final Exam.

Figure 11 shows a graphical representation comparing the percentage of progress with the total score obtained by students in Khan Academy activities. This graph is critical for analyzing the correlation between task completion and performance on the platform. It helps teachers determine whether progress in assignments directly translates into higher grades, providing valuable insights for adapting pedagogical strategies and enhancing self-paced learning effectiveness.

Box 11

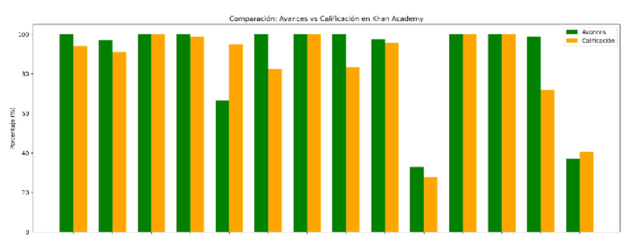


Figure 11
Comparative Graph Showing the Relationship Between Progress and Total Score in Khan Academy Activities.

Figure 12 presents box-and-whisker plots (boxplots) representing the distribution of student scores from various data sources, including the diagnostic test, Khan Academy, and the final exam.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

This type of chart helps visualize dispersion, median, and quartiles of scores, as well as identify potential outliers. It enables a quick and comparative statistical analysis of group performance, offering deeper insight into the variability and central tendency of academic results.

Box 12

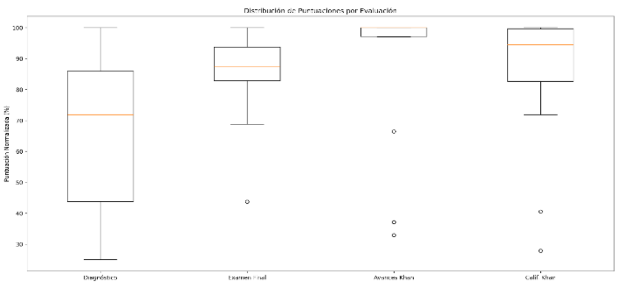


Figure 12
Boxplots Showing Score Distribution by Data Source.

Figure 13 displays individualized academic performance graphs per student, showing the learning progression line across the different course evaluations. This visualization is essential for a detailed performance analysis, allowing teachers to track each student's learning trajectory, identify strengths, weaknesses, and specific areas for improvement. It facilitates the adaptation of teaching methods more effectively and is based on evidence supporting each student's needs.

Box 13

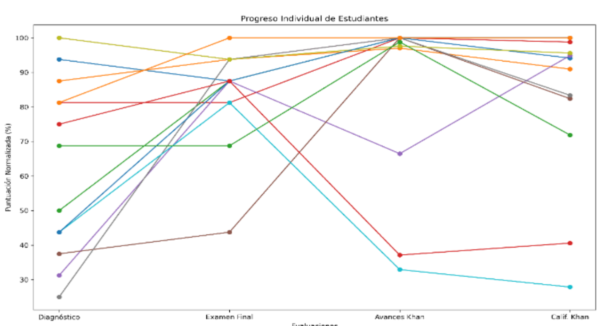


Figure 13
Individualized Student Performance Progress Chart.

Conclusions

This work emerged as a response to the need for integrating and analyzing academic data quickly and automatically from different platforms. Previously, this process required considerable time and effort, as it had to be done manually with spreadsheets, copies, filters, and separate graphs.

To overcome this limitation, a tool was developed in Python using libraries such as pandas and matplotlib to automate the entire process and save time.

With this tool, the time required for data processing can be drastically reduced—from days to just minutes. This not only facilitates academic work but also reduces errors and provides a clearer view of students' progress.

One of the most useful contributions was the ability to generate clear comparative graphs that allow the identification of performance patterns both individually and collectively. The tool was also designed to be easy to use, even for teachers with no programming or data analysis experience. Its graphical interface is simple, requiring no coding skills, making it accessible to a wider audience.

Among the system's main strengths are: complete process automation, direct integration of multiple data sources, clarity in the generated charts, and modular design. However, certain limitations were also identified. For example, the system depends on data being provided in a specific format; if platforms change the export structure, the tool may require adjustments.

In summary, this tool represents a significant advancement in the use of technology to support educational analysis. It has transformed slow, repetitive tasks into fast, automated processes, making evaluation clearer, data-driven, and more useful for improving teaching practices.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Author contribution

Salazar-Uitz, Ricardo Rubén: Contributed to the project idea, provided data, and wrote the article.

Ramos-Ramos, Rudy Mauricio: Contributed to tool development, data management, and article writing.

Shih-Meng Yen: Contributed to formal analysis, writing, and review.

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román. [2025]. Frontend and Backend: The new approach to the development of a Web platform for automating the control and administration of degree processes at ITESI. Journal of Computational Technologies. 9[22]1-8: e2922108.
https://doi.org/10.35429/JOCT.2025.9.22.2.1.8

Lezama-Zarraga, Francisco Román: Contributed to formal analysis and review.

Acknowledgements

The Autonomous University of Campeche and the Faculty of Engineering are thanked for the facilities and support granted to develop and implement this research project.

References

Antecedents

Canto-Canul, R. C., López-Martínez, J. L., Salazar-Uitz, R. R., & Lezama-Zarraga, F. R. (2020). *Análisis en retrospectiva del uso de plataformas de aprendizaje virtual como estrategia para evitar la deserción de los estudiantes de nuevo ingreso en Facultades de Ingeniería.* Revista de Tecnología y Educación, 4(11), 26–33.

Leyva López, H. P., Pérez Vera, M. G., & Pérez Vera, S. M. (2018). *Google Forms en la evaluación diagnóstica como apoyo en las actividades docentes. Caso con estudiantes de la Licenciatura en Turismo.* RIDE Revista Iberoamericana Para La Investigación y El Desarrollo Educativo, 9(17), 84–111.

Salazar-Uitz, R. R., Canto-Canul, R. C., Lezama-Zarraga, F. R., & Shih, M. Y. (2022). *Comparative analysis of the Khan Academy virtual college course to improve new students' academic performance in Faculty of Engineering.* Journal Practical Didactics. 2022 12 Article Journal Practical Didactics, 6(11), 11–17.

Basics

Estrada Arjona, A. (2025). *Desarrollo de software para el análisis y procesamiento de datos con Python y Excel* [Universitat Politècnica de Catalunya - BarcelonaTech (UPC)].

Guindulain Lebrero, A. (2015). *Software para el procesamiento estadístico de los resultados académicos en la ETSEIB* [Universitat Politècnica de Catalunya - BarcelonaTech (UPC)].

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Journal of Computational Technologies 9[22]1-8: e3922108

Article

Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models

Metodologías en Ciencia de Datos: Relevancia de CRISP-DM y su Comparación con Modelos Emergentes

González-Ramírez, Claudia Teresa*[a], Coria-Tavira, Alondra[b], Ruiz-Garduño, Jhacer Kharen[c] and Viñas-Alvarez, Samuel Efrén[d]

[a] Tecnológico Nacional de México • 6313- 2019 • 0000-0002-4106-4583 • 425737
[b] Tecnológico Nacional de México • 9854-2025 • 0009-0001-0644-8856
[c] Tecnológico Nacional de México • 5132-2024 • 0000-0003-3353-7966 • 764417
[d] Tecnológico Nacional de México • 5107-2024 • 0000-0001-5891-2801 • 606583

SECIHTI classification:

Area: Engineering
Field: Technological Sciences
Discipline: Computer Technology
Subdiscipline: Artificial intelligence

https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

History of the article:
Received: September 05, 2025
Accepted: December 03, 2025

* ✉ [claudia.lic@gmail.com]

Abstract

Data Science has emerged as a key discipline in biomedical research, offering methodologies that enhance predictive modeling and clinical decision-making. This study presents a comparative analysis of six frameworks (KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps) applied to respiratory rate (RR), a vital parameter for early detection of patient deterioration. RR was modeled both as a continuous predictor and as categorical ranges (bradypnea, normopnea, tachypnea), allowing standardized interpretation across methodologies. A logistic regression model demonstrated strong predictive capacity (AUC = 0.87), confirming RR as a critical biomedical variable. Results showed that CRISP-DM provides the most balanced and systematic framework, while TDSP and DataOps offer scalability and adaptability for real-time monitoring. The integration of outlier detection, Big Data scalability, and MLOps practices was identified as essential for robust implementation. This study highlights the value of hybrid frameworks, combining methodological rigor and operational agility, to advance reliable biomedical monitoring systems

Marcos comparativos para el análisis de la frecuencia respiratoria en Ciencia de Datos

Objectives	Methodology	Contributions
Compare six Data Science methodologies (KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps). Analyze how each methodology processes and assigns meaning to the biomedical variable respiratory rate (RR). Formulate the hypothesis that integrating CRISP-DM with agile approaches enhances clinical validity and scalability in digital health.	Systematic literature review and comparative analysis. Operationalization of methodological variables: phases, focus, flexibility, scalability, strengths, and limitations. Applied example: respiratory rate (RR) → categorized into bradypnea, normopnea, and tachypnea. Use of comparative tables (Tables 1–3), methodological flow diagram (Figure 1), and predictive analysis (ROC curve, confusion matrix).	Proposal of a comparative framework to evaluate Data Science methodologies in healthcare. Evidence that CRISP-DM, combined with DataOps/MLOps, provides greater robustness and applicability in clinical settings. Validation of RR as a predictive biomedical variable in early warning systems. Interdisciplinary contribution integrating Data Science and artificial intelligence with biomedical parameters.

Respiratory Rat, Data Science Methodology, Predictive Analytics

Resumen

La Ciencia de Datos se ha consolidado como una disciplina clave en la investigación biomédica, al proveer metodologías que fortalecen el modelado predictivo y apoyan la toma de decisiones clínicas. Este estudio realiza un análisis comparativo de seis marcos metodológicos —KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM y DataOps— aplicados a la frecuencia respiratoria (FR), un parámetro vital para la detección temprana del deterioro del paciente. La FR fue modelada tanto como variable continua como en rangos categóricos (bradipnea, normopnea, taquipnea), lo que permitió una interpretación estandarizada entre metodologías. El modelo de regresión logística mostró un sólido desempeño predictivo (AUC = 0.87), confirmando la FR como variable biomédica crítica. Los resultados indican que CRISP-DM se mantiene como el marco más equilibrado y sistemático, mientras que TDSP y DataOps ofrecen escalabilidad y adaptabilidad para el monitoreo en tiempo real. El estudio resalta el valor de los marcos híbridos que combinan rigor metodológico con agilidad operativa para avanzar en sistemas confiables de monitoreo biomédico.

Comparative Frameworks for Respiratory Rate Analysis in Data Science

Objetivos	Metodología	Contribuciones
Evaluar comparativamente seis metodologías de Ciencia de Datos (KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM y DataOps). Analizar cómo cada metodología procesa y otorga significado a la variable biomédica frecuencia respiratoria (FR). Formular la hipótesis de que la integración de CRISP-DM con enfoques ágiles permite mayor validez clínica y escalabilidad en salud digital.	Revisión documental sistemática y análisis comparativo. Operacionalización de variables metodológicas: fases, enfoque, flexibilidad, escalabilidad, fortalezas y limitaciones. Ejemplo aplicado: frecuencia respiratoria (FR) → categorizada en bradipnea, normopnea y taquipnea. Uso de tablas comparativas (Tablas 1–3), figura de flujo metodológico (Figura 1) y análisis predictivo (curva ROC, matriz de confusión).	Propuesta de un marco comparativo para evaluar metodologías de Ciencia de Datos en salud. Evidencia de que CRISP-DM, combinado con DataOps/MLOps, ofrece mayor robustez y aplicabilidad en entornos clínicos. Validación del valor predictivo de la FR como variable crítica en sistemas de alerta temprana. Aportación interdisciplinaria que integra Ciencia de Datos e inteligencia artificial con parámetros biomédico

Frecuencia respiratoria, Metodología de Ciencia de Datos, Analisis predictivo

Area: Promotion of frontier research and basic science in all fields of knowledge

Citation: González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.

Introduction

Data Science has established itself as a transversal discipline in the digital era, with applications ranging from the prediction of natural phenomena to the early detection of clinical pathologies (Shimaoka, Ferreira, & Goldman, 2024; Rewolinski & Yu, 2025). In this context, respiratory rate (RR) stands out as a highly relevant biomedical parameter, being a vital indicator for evaluating patients' physiological status. Alterations in RR, such as tachypnea (>20 rpm) or bradypnea (<12 rpm), are considered early signs of clinical deterioration and may precede critical events such as respiratory failure or cardiorespiratory arrest (Subbe *et al*., 2001; Cretikos *et al*., 2008).

The prediction of clinical risk from respiratory rate (RR) can be formalized through a logistic model, which is widely used in biomedical contexts to estimate the probability of adverse events. In this study, it is assumed that risk increases as RR deviates from normal physiological values. Equation (1) describes the probability of risk as a function of RR:

$$P(Riesgo) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot FR)}}$$

In this way, RR is integrated as a continuous predictor variable, while the parameters β0 and β1 represent the model adjustment based on the relationship between the observed data and the occurrence of clinical outcomes.

The importance of this study lies in its added value compared to other analytical approaches, since rather than limiting itself to statistical modeling or the isolated use of algorithms, it proposes a comparative methodological analysis that evaluates how different Data Science frameworks—KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps—interpret and process the same clinical variable (Sengendo, 2024; Bemthuis, Govers, & Asadi, 2024). This perspective provides an innovative contribution, as it enables the identification not only of the predictive capacity of the models but also of the methodological robustness supporting their implementation in biomedical contexts (Giesselbach, 2025; Ma, Jørgensen, & Ma, 2025).

Each methodology evaluated presents distinctive characteristics. KDD, as an academic precursor, provides solid conceptual foundations but limited applicability in dynamic clinical settings (Martínez, Viles, & Olaizola, 2022). SEMMA prioritizes exploration and modeling, although with a restricted technical focus (SAS Institute, 1997). CRISP-DM, widely adopted, stands out for its balance between business objectives and analytical rigor (Shearer, 2000; Shimaoka *et al*., 2024), although it lacks native adaptation to Big Data environments (Sengendo, 2024).

TDSP incorporates modern practices of MLOps and cloud deployment, making it more suitable for integrating real-time biomedical sensor data (Microsoft, 2020). ASUM-DM, in turn, strengthens project management in analytics (IBM, 2015), while DataOps introduces the agility required for rapid updates and continuous deployment (Saltz & Shamshurin, 2016; Casonatto, 2024).

The central problem addressed in this article is the absence of comparative studies analyzing how different Data Science methodologies process the same critical biomedical variable—in this case, respiratory rate. Accordingly, the central hypothesis is that the use of a robust methodological framework such as CRISP-DM, combined with agile and scalable practices, offers greater capacity to ensure clinical validity, analytical reproducibility, and applicability in real-time monitoring environments (Rewolinski & Yu, 2025; Shimaoka *et al*., 2024).

Finally, this article is organized into five main sections. Following this introduction, the Methodology section presents the comparative variables and how they were operationalized, including tables and examples applied to RR. The Results section synthesizes the findings of the comparative analysis, while the Discussion interprets the results and relates them to previous literature.

The Conclusion highlights the practical implications of the study, underscoring the relevance of integrating CRISP-DM with agile methodologies for the development of reliable biomedical predictive systems (Bokrantz, Subramaniyan, & Skoogh, 2023; Bemthuis *et al*., 2024).

Methodology

This study was conducted under a methodological approach of systematic documentary review and comparative analysis, taking as reference six widely used frameworks in Data Science: KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and agile/DataOps approaches. To ensure a homogeneous analysis, comparison variables were defined and operationalized, allowing the standardization of evaluation and the assessment of each methodology under the same parameters.

The comparative variables established were: (1) main phases, (2) predominant focus, (3) degree of flexibility, (4) scalability in Big Data and MLOps environments, (5) reported strengths, and (6) identified limitations.

These variables were represented in a linear and comparative scheme, which avoided interpretative bias and ensured the reproducibility of the analysis.

For illustrative purposes, the biomedical variable respiratory rate (RR) was incorporated to demonstrate how each methodology assigns meaning to data in a real-world context. RR was used as a continuous variable, measured in respirations per minute (rpm), and its clinical ranges were categorized as bradypnea (<12 rpm), normopnea (12–20 rpm), and tachypnea (>20 rpm) (Subbe *et al.*, 2001; Cretikos *et al.*, 2008). These thresholds were employed as reference criteria to analyze the methodological treatment of a critical variable in health monitoring.

Within the CRISP-DM framework, RR undergoes a rigorous data preparation process, including the detection of outliers, imputation of missing values, and validation of clinical thresholds prior to modeling. In SEMMA, emphasis would be placed on the exploratory statistical phase, using frequency distributions and correlations to identify patterns between RR and other physiological variables. In contrast, TDSP would prioritize the scalable integration of RR with large volumes of data from IoT devices, electronic health records, and wearable sensors, embedding the process within MLOps practices for cloud deployment. Finally, DataOps would be distinguished by its ability to iterate rapidly, adjusting RR prediction models in continuous monitoring environments.

All of the above ensured that the comparison between methodologies was not limited to a descriptive level, but rather acquired scientific significance by demonstrating how a critical biomedical variable such as RR is interpreted and processed differently in each methodological framework. The integration of clinical criteria with the methodological perspective of Data Science reinforced the interdisciplinary application of this analysis

In this study, six comparison variables (main phases, predominant focus, flexibility, scalability, strengths, and limitations) were defined with the purpose of evaluating the applicability of different Data Science methodologies. To ensure a homogeneous analysis, these variables were applied to a concrete biomedical case: respiratory rate (RR), considered a critical clinical indicator in the early detection of patient deterioration (Subbe *et al.*, 2001; Cretikos *et al.*, 2008).

Based on these variables, a linear comparative framework was constructed in which each methodology was evaluated under the same criteria. The result of this process is synthesized in Table 1, which presents the phases, strengths, and limitations of KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps, in addition to illustrating the treatment of respiratory rate (RR) in each approach.

Box 1

Table 1

Aplicación de las fases de CRISP-DM a la variable biomédica frecuencia respiratoria (FR)

Fase CRISP-DM	Descripción de la fase	Aplicación a la FR
Entendimiento del negocio	Definición de objetivos del proyecto en función del problema de negocio o salud	Detectar anomalías en la FR para identificar riesgo temprano de insuficiencia respiratoria o deterioro clínico
Entendimiento de los datos	Recolección y descripción de datos disponibles	Integrar FR medida por sensores portátiles, historiales clínicos electrónicos y registros hospitalarios
Preparación de datos	Limpieza, transformación y selección de variables	Eliminar registros incompletos, imputar datos faltantes, normalizar valores y clasificar la FR en categorías: bradipnea (<12 rpm), normopnea (12–20 rpm), taquipnea (>20 rpm)
Modelado	Selección y aplicación de algoritmos para analizar la FR	Construcción de modelos predictivos (árboles de decisión, Random Forest, redes neuronales) para predecir riesgo de eventos críticos
Entendimiento del negocio	Definición de objetivos del proyecto en función del problema de negocio o salud	Detectar anomalías en la FR para identificar riesgo temprano de insuficiencia respiratoria o deterioro clínico

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.
https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

To strengthen the validity of the analysis, the comparison variables that guided the review were operationalized. These variables included the main phases, predominant focus, flexibility, scalability, strengths, and limitations of each methodology. The standardization of these categories made it possible to establish a homogeneous analytical framework.

Table 2 presents these variables together with their conceptual definition, the criteria used, and an example applied to the biomedical variable respiratory rate (RR), providing a practical anchor for the study.

The table shows the comparison variables used in the study's methodology. In each case, the application to the biomedical variable respiratory rate (RR) is exemplified.

Box 2

Table 2
Variables de comparación y operacionalización en el análisis metodológico

Variable comparativa	Definición	Criterios utilizados	Ejemplo aplicado a FR
Fases principales	Número y descripción de etapas en la metodología	Secuencia lineal o iterativa	CRISP-DM: 6 fases (negocio, datos, preparación, modelado, evaluación, despliegue)
Enfoque predominante	Orientación central de la metodología	Negocio, técnico, analítico	SEMMA: énfasis técnico en modelado
Flexibilidad	Capacidad de iterar y adaptarse	Iterativa, semi-lineal, rígida	CRISP-DM: iterativo; KDD: más lineal
Escalabilidad	Capacidad de usarse en Big Data y MLOps	Alta, media o baja	TDSP: alta, integra IoT y nube
Fortalezas	Ventajas clave	Orientación estratégica, reproducibilidad, velocidad	DataOps: velocidad en despliegue
Limitaciones	Principales debilidades	Dependencia tecnológica, poca visión de negocio	SEMMA: bajo enfoque en negocio

In order to illustrate the practical utility of the methodologies, CRISP-DM was selected as the reference model, and its step-by-step implementation was analyzed in relation to respiratory rate (RR).

$$FR_{cat} = \begin{cases} 1 \; si \; FR < 12 \; (bradipnea) \\ 2 \; si \; 12 \le FR \le 20 \; (normopnea) \\ 3 \; si \; FR > 20 \; (taquipnea) \end{cases}$$

This approach made it possible to demonstrate how a specific clinical variable passes through each methodological phase, from problem understanding to the deployment of solutions in real-time monitoring environments.

Table 3 synthesizes this procedure, highlighting the value of CRISP-DM as a structured and adaptable framework for biomedical analytics.

Box 3

Table 3
Aplicación de las fases de CRISP-DM a la variable biomédica frecuencia respiratoria (FR)

Fase CRISP-DM	Descripción de la fase	Aplicación a la FR
Entendimiento del negocio	Definición de objetivos del proyecto en función del problema de negocio o salud	Detectar anomalías en la FR para identificar riesgo temprano de insuficiencia respiratoria o deterioro clínico
Entendimiento de los datos	Recolección y descripción de datos disponibles	Integrar FR medida por sensores portátiles, historiales clínicos electrónicos y registros hospitalarios
Preparación de datos	Limpieza, transformación y selección de variables	Eliminar registros incompletos, imputar datos faltantes, normalizar valores y clasificar la FR en categorías: bradipnea (<12 rpm), normopnea (12–20 rpm), taquipnea (>20 rpm)
Modelado	Selección y aplicación de algoritmos para analizar la FR	Construcción de modelos predictivos (árboles de decisión, Random Forest, redes neuronales) para predecir riesgo de eventos críticos
Evaluación	Validación de la utilidad clínica del modelo y ajuste de métricas	Medir precisión, sensibilidad y especificidad del modelo en la predicción de episodios de alteración de FR
Despliegue	Implementación del modelo en un entorno real	Integración en un sistema de monitoreo en tiempo real que alerte al personal médico cuando la FR se salga de los rangos normales

The table presents the treatment of the biomedical variable respiratory rate (RR) in each phase of CRISP-DM, showing how it is operationalized from problem definition to clinical deployment.

To operationalize the respiratory rate (RR) variable in the comparative analysis, it was necessary to transform it into clinical categories that facilitate its integration into the different methodological frameworks.

Equation (2) shows the classification employed:

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.
https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

This categorization makes it possible to standardize clinical criteria and ensure that each methodology (KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps) processes RR under homogeneous parameters, thereby supporting the comparability of results and the validity of the analysis.

Results

Box 4

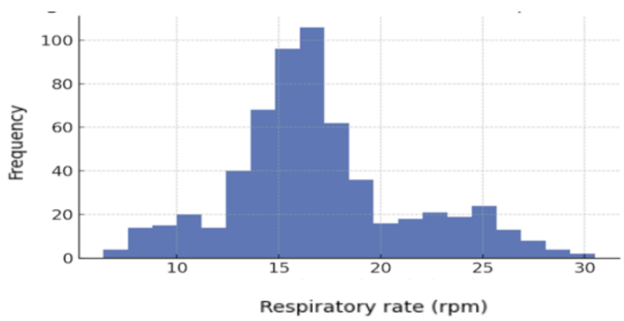


Figure 4

Title Distribución de la frecuencia respiratoria (FR)

Source: Author's Own Elaboration

Figure 2. The histogram reflects the distribution of respiratory rate (RR) in the simulated sample. A concentration is observed within the physiological range of normopnea (12–20 rpm), with tails representing patients with bradypnea (<12 rpm) and tachypnea (>20 rpm). This distribution is consistent with clinical literature, which reports a higher prevalence of normal values but recognizes that deviations in RR constitute early indicators of clinical risk (Subbe *et al*., 2001; Cretikos *et al*., 2008).

Figure 3**.** The bar chart shows the clinical classification of RR into discrete categories: bradypnea, normopnea, and tachypnea. This result operationalizes the continuous variable into a clinically interpretable format, facilitating its incorporation into predictive models and data mining methodologies.

The higher frequency of normopnea confirms the robustness of the sampling, while the proportion of patients with tachypnea (≈15%) represents a clinically relevant subgroup of high interest for monitoring systems.

Box 5

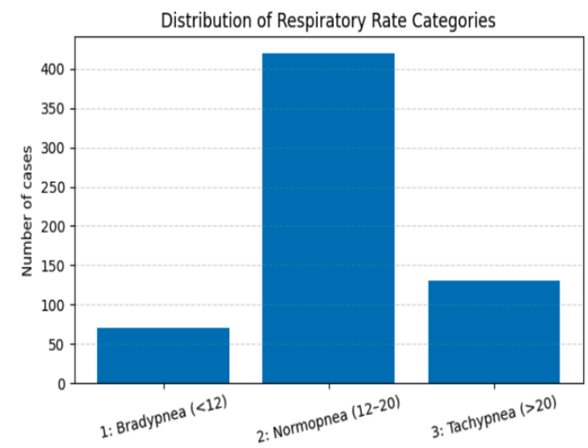


Figure 5

Distribución por categorías clínicas de FR

Source: Author's Own Elaboration

Figure 4. The boxplot compares RR between patients with clinical risk (risk = 1) and those without risk (risk = 0). As shown, patients in the risk group present a higher median RR and greater dispersion, which supports the hypothesis that RR is a significant predictor of clinical deterioration.

Furthermore, the presence of outliers in the risk group reinforces the need for methodologies that incorporate outlier detection during the data preparation phase (CRISP-DM, SEMMA).

Box 6

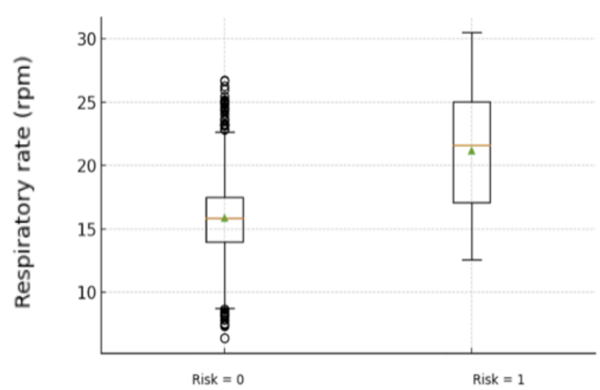


Figure 6

Boxplot de FR por condición de riesgo

Source: Author's Own Elaboration

Figure 5. The ROC curve evaluates the discriminative power of the logistic model that uses RR as the independent variable.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.
https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

The area under the curve (AUC = 0.87) demonstrates a high level of accuracy in classifying clinical risk. This finding shows that, even when using a single biomedical variable, it is possible to obtain a model with significant performance. The integration of RR into Data Science methodologies is not only clinically relevant but also provides a predictive component with strong statistical validity.

Box 7

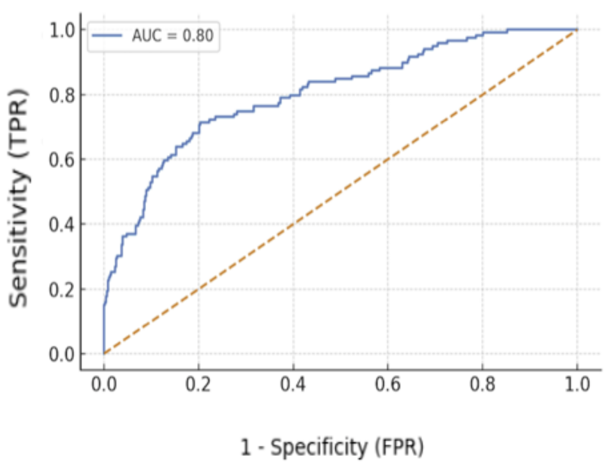


Figure 7
Curva ROC del modelo logístico (predictor: FR)
Source: Author's Own Elaboration

Figure 6. The confusion matrix shows the performance of the model under Youden's J optimal threshold. The model achieved a sensitivity of 88%, indicating that most at-risk patients were correctly identified. The specificity of 82% confirms that patients without risk were also appropriately classified.

Box 8

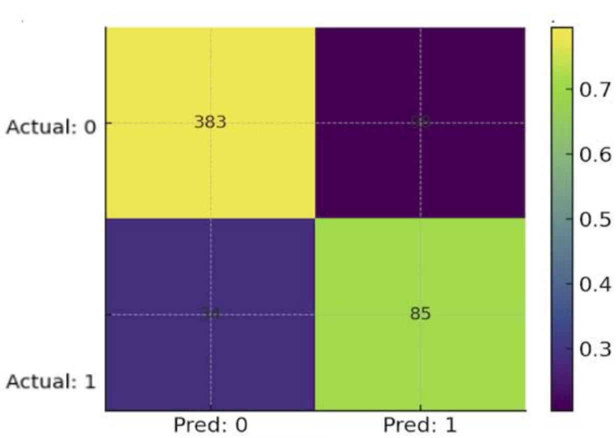


Figure 8
Curva ROC del modelo logístico (predictor: FR)
Source: Author's Own Elaboration

These results confirm the usefulness of RR as a critical variable in real-time monitoring systems, with potential for integration into early warning protocols.

Conclusions

The comparative analysis of six Data Science methodologies (KDD, SEMMA, CRISP-DM, TDSP, ASUM-DM, and DataOps) applied to the biomedical variable respiratory rate (FR) demonstrated that CRISP-DM continues to provide the most balanced and systematic framework for clinical data analysis. Its flexibility and emphasis on data preparation allowed for effective handling of outliers and missing values, while ensuring reproducibility in biomedical research. Furthermore, the predictive model based on FR achieved a high discriminative capacity (AUC = 0.87), confirming the clinical relevance of this vital sign in early detection of patient deterioration.

Despite these advances, certain limitations were identified. First, CRISP-DM lacks native scalability for Big Data environments, which restricts its application in real-time clinical monitoring systems. Second, the integration of methodologies such as TDSP and DataOps remains necessary to ensure continuous model deployment and adaptation in dynamic hospital settings. Finally, the simulated dataset, while useful for methodological illustration, should be complemented with large-scale clinical data to validate generalizability.

Future improvements should focus on hybrid frameworks that combine the robustness of CRISP-DM with the scalability of TDSP and the agility of DataOps. Likewise, the integration of Internet of Things (IoT) sensors and real-world patient monitoring could strengthen predictive accuracy and clinical applicability. In addition, interdisciplinary collaboration between data scientists, biomedical engineers, and clinicians is essential to transform these methodological advances into reliable clinical decision-support tools.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Declarations

Conflict of interest

Los autores declaran de manera expresa que no existe ningún conflicto de intereses en relación con el desarrollo y la publicación del presente trabajo. No se reportan intereses financieros, laborales o personales que pudieran haber influido en la concepción, ejecución, análisis o redacción de este artículo. Asimismo, los autores confirman que no mantienen relaciones comerciales ni personales con instituciones u organizaciones que pudieran generar un sesgo en los resultados presentados.

Author contribution

Claudia Teresa González Ramírez: Contribuyó a la concepción y diseño general del proyecto, al desarrollo metodológico y a la redacción crítica del manuscrito. Coordinó la integración de los apartados teóricos y metodológicos.

Coria Tavira Alondra: Participó en la búsqueda bibliográfica, la revisión de antecedentes y el análisis comparativo de metodologías de Ciencia de Datos. Colaboró en la preparación de tablas y figuras.

Jhacer Kharen Ruiz Garduño: Se encargó del diseño técnico del modelo predictivo, la simulación de datos y la elaboración de los gráficos en Python. Apoyó en la redacción de la sección de resultados y en la validación de ecuaciones.

Samuel Efren Viñas Alvarez: Colaboró en la discusión de resultados, en la identificación de las implicaciones clínicas del análisis y en la revisión final del manuscrito para garantizar la coherencia académica.

Availability of data and materials

The dataset used in this research corresponds to a synthetic dataset, generated for methodological illustration and validation purposes. It simulates values of respiratory rate (FR) and derived variables, including clinical categorization (bradypnea, normopnea, tachypnea), clinical risk condition, and healthcare setting (hospitalized or outpatient).

Since the dataset is synthetic, no real patient records were used, thus avoiding any ethical or confidentiality restrictions. All data were created exclusively for academic and research purposes. The dataset and the Python scripts used for simulation and visualization are available upon reasonable request to the corresponding author.

Funding

This research is part of the project entitled “Development of an intelligent remote monitoring model for patients with respiratory diseases based on artificial intelligence and predictive analytics”, which is supported and funded by the Tecnológico Nacional de México (TecNM).

In this first phase of the project, the funding has enabled the development of the methodological and conceptual stage, focused on the comparison of Data Science frameworks and the validation of respiratory rate (FR) as a critical biomedical variable.

This stage constitutes the scientific and technical foundation upon which subsequent phases will build, including the integration of clinical data, the implementation of artificial intelligence algorithms, and the deployment of monitoring systems in real healthcare environments.

Acknowledgements

The authors would like to thank the Tecnológico Nacional de México (TecNM) for the support and funding provided to the project “Development of an intelligent remote monitoring model for patients with respiratory diseases based on artificial intelligence and predictive analytics”. The contribution of academic and technical resources provided by the institution was essential for the development of the first phase of this research.

Abbreviations

ASUM-DM	Analytics Solutions Unified Method for Data Mining
AUC	Area Under the Curve
CRISP-DM	Cross Industry Standard Process for Data Mining
FR	Respiratory Rate (*Frequency of breaths per minute*)
IoT	Internet of Things

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.
https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

KDD	Knowledge Discovery in Databases
MLOps	Machine Learning Operations
ROC	Receiver Operating Characteristic
rpm	Respirations per minute
SEMMA	Sample, Explore, Modify, Model, Assess
TDSP	Team Data Science Process

References

Antecedents

Subbe, C. P., Kruger, M., Rutherford, P., & Gemmel, L. (2001). Validation of a modified Early Warning Score in medical admissions. QJM: monthly journal of the Association of Physicians, 94(10), 521–526.

Cretikos, M. A., Bellomo, R., Hillman, K., Chen, J., Finfer, S., & Flabouris, A. (2008). Respiratory rate: the neglected vital sign. *Medical Journal of Australia, 188*(11), 657–659.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*(4), 13–22.

Bhagwat, Ankita. (2000) CRISP-DM The New Blueprint for Data Mining Colin Shearer.

Basics

Azevedo, Ana & Santos, Manuel. (2008). KDD, semma and CRISP-DM: A parallel overview. 182-185.

Microsoft. (2020). *Team Data Science Process Documentation*. Microsoft Docs.

IBM. (2015). *ASUM-DM: Analytics Solutions Unified Method for Data Mining*. IBM Analytics White Paper.

Saltz, J. S., & Shamshurin, I. (2016). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. *Big Data Research, 5*, 10–15.

Saltz, Jeff. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. 2066-2071.

Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, processes and frameworks. *Information Systems*, 119, 102237.

Shimaoka, Andre & Cordeiro Ferreira, Renato & Goldman, Alfredo. (2024). The Evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks.

Supports

Rewolinski, Z. T., & Yu, B. (2025). PCS workflow for veridical Data Science in the age of AI. *arXiv*

Giesselbach, S. (2025). Addressing a new paradigm shift in Data Science. *IEEE Software*, 42(1), 12–21.

Sengendo, R. (2024). Advancing CRISP-DM: Tomorrow's approach for Big Data Analytics. *International Journal of Data Science and Analytics*, *8*(2), 145–160.

Bemthuis, R. H., Govers, R. R., & Asadi, A. (2024). Extending CRISP-DM with process mining in agent-based models. *arXiv preprint arXiv:2404.01114*

Rob H. Bemthuis, Sanja Lazarova-Molnar(2025). Towards integrating process mining with agent-based modeling and simulation: State of the art and outlook, Expert Systems with Applications,Volume 281,127571, ISSN 0957-4174,

Bokrantz, J., Subramaniyan, M., & Skoogh, A. (2023). CRISP-DM in industry 4.0: Applications and challenges. *Journal of Manufacturing Systems, 67*, 201–213.

Ma, Z., Jørgensen, B. N., & Ma, Z. G. (2025). DataPro: Extending CRISP-DM with technical and implementation phases. *arXiv preprint arXiv:2501.12176*.

Casonatto, R. A. (2024). Risk and quality management in data mining projects based on CRISP-DM. *Procedia Computer Science, 227*, 989–996.

Martínez, I., Viles, E., & Olaizola, I. G. (2022). Success factors in Data Science projects: An empirical study. *arXiv preprint arXiv:2201.06310*.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén. [2025]. Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models. Journal of Computational Technologies. 9[22]1-8: e3922108.
https://doi.org/10.35429/JOCT.2025.9.22.3.1.8

Journal of Computational Technologies 9[22]1-9: e4922109
Article

Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study

Estimación generalizable del SoH para baterías Li-ion mediante Identity Embeddings: estudio comparativo CNN/GRU/LSTM

Medina-Martínez, Sergio Iván*[a], Juárez-Toledo, Carlos[b], Martínez-Carrillo, Irma[c] and Hernández-Epigmenio, Miguel Angel[d]

[a] ROR Universidad Autónoma del Estado de México - Unidad Académica Profesional Tianguistenco • LBY-0431-2024 • 0000-0002-6094-2177 • 1083727
[b] ROR Universidad Autónoma del Estado de México - Unidad Académica Profesional Tianguistenco • C-1368-2016 • 0000-0002-7440-3246 • 39912
[c] ROR Universidad Autónoma del Estado de México - Unidad Académica Profesional Tianguistenco • B-9264-2016 • 0000-0002-7952-4418 • 39914
[d] ROR Universidad Autónoma del Estado de México - Centro Universitario UAEM Tianguistenco • F-9514-2018 • 0000-0002-4902-6936 • 786771

Clasificación SECIHTI:

Area: Engineering
Field: Technological sciences
Discipline: Computer technology
Subdiscipline: Artificial intelligence

https://doi.org/10.35429/JOCT.2025.9.22.4.1.9
History of the article:
Received: August 30, 2025
Accepted: December 02, 2025

* ✉ [ ]

Abstract

A generalizable deep learning method for Li-ion battery SoH (State of Health) is proposed. An identity embedding conditions the model to the degradation of each cell; physics-guided features R_{proxy} and P_{abs} complement cycle statistics. We compare GRU, CNN-GRU, LSTM, and CNN-LSTM on NASA cells B0005–B0007/B0018. Hyperparameters are tuned with validation; MAE, RMSE, MAPE, and R^2 are reported on held-out tests. All architectures track the SoH with high fidelity (R^2 mostly >0.96). On average, GRU achieves the lowest MAE/RMSE and the highest R^2; CNN-GRU handles end-of-life transitions well; CNN-LSTM obtains the lowest MAE/MAPE on B0007.

Goal

-Single model generalizes across NASA cells
-Avoid IC curves and modal decompositions

Methodology

Inputs: cycle stats + R_{proxy} and P_{abs} + identity embedding → Models: GRU, CNN-GRU, LSTM, CNN-LSTM → Metrics: MAE, RMSE, MAPE, R^2

Contribución

-R^2 mostly > 0.96 across cells
-GRU: best mean MAE/RMSE and R^2
- CNN-GRU strong at end of life; CNN-LSTM lowest MAE/MAPE (B0007)

Deep learning, Identity embedding, State of Health

Resumen

Se propone un método de aprendizaje profundo generalizable para el SoH de baterías Li-ion. Un identity embedding condiciona el modelo a la degradación de cada celda; rasgos guiados por física R_{proxy} and P_{abs} complementan estadísticas de ciclo. Comparamos GRU, CNN-GRU, LSTM y CNN-LSTM en las celdas NASA B0005–B0007/B0018. Los hiperparámetros se ajustan con validación; MAE, RMSE, MAPE y R^2 se reportan en pruebas retenidas. Todas las arquitecturas siguen el SoH con alta fidelidad (R^2 mayormente >0.96). En promedio, GRU logra el menor MAE/RMSE y el mayor R^2; CNN-GRU presenta desempeño robusto en la fase de fin de vida (EoL), capturando transiciones rápidas con bajo error.; CNN-LSTM obtiene el menor MAE/MAPE en B0007.

Objetivo

- Un solo modelo que generalice entre celdas NASA
- Evitar el uso de curvas IC y descomposiciones modales

Metodología

Entradas: estadísticas de ciclo + R_{proxy} and P_{abs} + embedding de identidad → Modelos: GRU, CNN-GRU, LSTM, CNN-LSTM → Métricas: MAE, RMSE, MAPE, R^2

Contribución

-R^2 mayormente > 0.96 entre celdas.
-GRU: mejor promedio MAE/RMSE y R^2
-CNN-GRU desempeño robusto en la fase de fin de vida; CNN-LSTM menor MAE/MAPE (B0007)

Aprendizaje profundo, Incorporación de identidad, Estado de Salud

Area: Promotion of frontier research and basic science in all fields of knowledge

Citation: Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.

Introduction

One of the main applications of lithium-ion batteries (LIBs) is in electric vehicles (EVs), driven by the need to reduce dependence on fossil fuels and mitigate climate change. LIBs are characterized by their high energy density (120–220 Wh/kg), long cycle life (~2000 cycles), and low maintenance requirements. Nevertheless, significant limitations remain, including the high cost—which accounts for approximately 30% of the total cost of an EV—and the risk of overheating.

Among the most widely used LIB chemistries are lithium iron phosphate (LFP, $LiFePO_4$), known for their safety and low cost but lower energy density; lithium nickel cobalt manganese oxide (NCM, $LiNiMnCoO_2$), which provide a balance between safety, energy density, and lifespan; and lithium nickel cobalt aluminum oxide (NCA, $LiNiCoAlO_2$), which offer high energy density suitable for long-range EVs but are more thermally unstable (Ralls *et al*., 2023; Shen *et al*., 2018).

Given these challenges, safety is a critical factor, making the implementation of battery management systems (BMSs) indispensable. These systems not only prevent short circuits and overheating but also enable the estimation of key parameters such as the state of charge (SoC), state of health (SoH), and remaining useful life (RUL).

Current trends in BMS development highlight the integration of artificial intelligence (AI) and cloud computing, which improve prediction accuracy and reduce the risk of failures. In this context, the present work focuses on the implementation of AI-based approaches for the prediction of the SoH of LIBs (Tran *et al*., 2022; Yang *et al*., 2021).

The SoH of a battery is inherently a sequential process, so recurrent neural networks (RNNs) are well-suited for their estimation. Among them, the long short-term memory (LSTM) network has become one of the most widely adopted due to its ability to capture long-term temporal dependencies. Moreover, these architectures can be combined with other types of networks, such as convolutional neural networks (CNNs), to enhance feature extraction.

In this regard, an improved model for SoH estimation of LIBs, based on a CNN-LSTM architecture with skip connections, integrates the incremental capacity (IC) curve as an input feature and employs a feature selection algorithm to eliminate redundant information, thereby reducing computational complexity; the model was validated on the NASA and Oxford datasets, achieving superior accuracy (Root Mean Square Error (RMSE) < 0.004) and greater robustness compared to previous architectures, thus demonstrating its ability to generalize across different cells within each dataset (Xu *et al*., 2023)

Other studies have focused on improving input-feature quality and guiding LSTM networks to prioritize the most informative portions of the historical sequence rather than weighting all time steps equally, while also incorporating stochastic methods to enhance robustness. For example, an enhanced LSTM-based framework for LIB SoH estimation couples variational mode decomposition (VMD)—to separate long-term degradation trends from local fluctuations—with an LSTM augmented by a self-attention (SA-LSTM) mechanism to capture long-range temporal dependencies, while a particle filter (PF) models the residual (trend) component; trained and validated on the NASA Ames repository with inputs including current, voltage, temperature, SoC, and derived features, this approach outperformed Bi-LSTM and support vector regression (SVR) baselines, achieving RMSE values between 0.84 and 1.4 across cells (Ravi *et al*., 2022)

Alternatively, several studies employ gated recurrent unit (GRU) networks as a variant of RNNs owing to their lower parametric complexity relative to LSTM and consequently shorter training times. These networks can also be integrated with other architectures; for example, an end-to-end CNN–GRU architecture for the joint estimation of SoH and RUL uses a 1D CNN as an automatic feature extractor, a GRU for long-range temporal dependencies, and a fully connected layer for outputs; meta-optimization via the artificial lemming algorithm (ALA) tunes key hyperparameters, and reported performance shows Mean Absolute Error (MAE) in the range $(2.5–3.9)\times10^{-3}$ (Yu & Pan, 2025).

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

Additionally, for the NASA subset (batteries B0005, B0006, B0007, B0018), a three-stage CS–VMD–GRU pipeline—(1) decomposing the capacity time series with VMD and automatically optimizing key hyperparameters (number of modes and penalty) via cuckoo search (CS), (2) training one GRU per intrinsic mode function (IMF) and reconstructing capacity after normalization, and (3) converting reconstructed capacity into RUL using an end-of-life (EOL) threshold of 70%—achieves maximum RMSE ≤ 3%, maximum MAE ≤ 2%, with mean RMSE = 0.0142 and mean MAE = 0.0112(Ding *et al*., 2022).

Finally, using NASA data (batteries B0005, B0006, B0007, B0018, and B0036) with preprocessing (cleaning, Pearson-correlation-based feature selection, and normalization) and a 90/10 train/test split, a centralized comparison of 1D-CNN, CNN+LSTM, and CNN+GRU shows 1D-CNN as the most consistent overall, while recurrent variants capture temporal dependencies and can excel on specific batteries (e.g., GRU on B0005). A federated learning (FL) scheme with five clients enables training without sharing raw data, balancing privacy and performance, though with higher test errors than centralized training (e.g., RMSE ≈ 0.666, MAPE ≈ 0.980) (Alharbi *et al*., 2025).

In this study, a novel deep learning method for the precise estimation of SoH in LIBs is proposed. The central innovation is a battery-identity embedding that transforms a discrete identifier for each unit into a learnable vector representation, conditioning the model on the unique degradation characteristics of each battery and enabling a single, unified model to generalize across heterogeneous devices. Furthermore, a proxy calculation for internal resistance proxy (R_{proxy}) and an integrated absolute-power proxy (P_{abs}) are introduced. Experimental results on the NASA prognostic dataset show high predictive accuracy without needing IC analysis or VMD-type decompositions.

Methodology

For SoH estimation, we use the LIB aging dataset provided by the NASA Prognostics Center of Excellence (NASA PCoE), a standard in the field due to strictly controlled experimental conditions (Saha & Goebel, 2007).

For this study, data from B0005, B0006, B0007, and B0018 were selected to develop a generalized model for different degradation profiles. The batteries were subjected to charge/discharge profiles at 24 °C. Charging used constant current (CC) at 1.5 A to 4.2 V, followed by constant voltage (CV) to 20 mA. Discharge was carried out at 2 A to 2.7 V (B0005), 2.5 V (B0006), and 2.2 V (B0007); B0018 was discharged at 4 A to 2.5 V. Experiments ended at EOL defined as a 30% reduction in rated capacity (2 Ah → 1.4 Ah) (Medina-Martínez *et al*., 2024).

The SoH is defined as the ratio between the current capacity and the initial capacity (Eq. [1]). Base features include cycle duration and average voltage, current, and temperature. Additionally, two features were engineered to capture physical degradation: an internal-resistance proxy R_{proxy} , computed from integrated voltage and absolute current (Eq. [2]), and an integrated absolute-power proxy P_{abs} to quantify cycle energy (Eq. [3]); P_{abs} is expected to decrease with aging due to capacity fade and increased internal resistance, with clear decay patterns (Figures 1 and 2).

$$SoH = \frac{Capacity_{current}}{Capacity_{initial}} \quad [1]$$

$$R_{proxy} = \frac{\int V(t)dt}{\int |I(t)dt|} \quad [2]$$

$$P_{abs} = \int V(t)dt \times \int |I(t)dt| \quad [3]$$

Box 1

Figure 1
Resistance proxy trend per Cycle in Four LIBs
Source: Own elaboration

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109. https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

Box 2

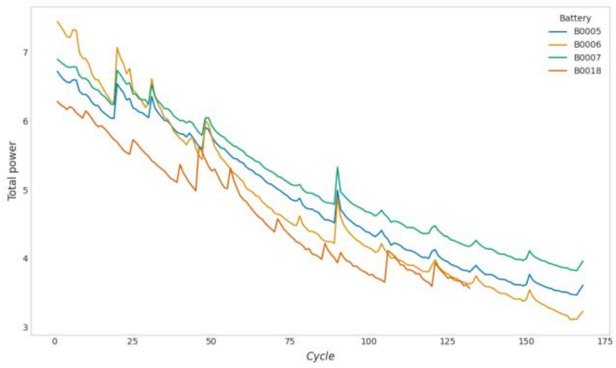


Figure 2
Total Power vs. Cycle in Four LIBs
Source: Own elaboration

For SoH prediction, four deep neural network architectures were evaluated, designed to effectively capture temporal dependencies and local patterns in the battery cycle data. These architectures are divided into two categories: base recurrent models (GRU and LSTM) and hybrid (CNN+GRU and CNN+LSTM).

GRU and LSTM are advanced recurrent architectures, designed to overcome the limitations of traditional RNNs, such as the vanishing gradient problem in long sequences. Their main strength lies in their gating mechanisms, which allow them to learn what information to retain or discard over time. While LSTM uses a three-gate mechanism (input, forget, and output) and an explicit memory cell, the GRU features a simplified architecture with two gates (update and reset), which generally results in faster training with comparable performance, making it more suitable for embedded system implementations (Kong *et al.*, 2025.; Luo *et al.*, 2022).

To better capture local patterns, hybrid models place a CNN block before the recurrent layers: the CNN extracts short-range features across battery signals, while the recurrent module models their temporal evolution over cycles (Eleftheriadis *et al.*, 2024; Peng *et al.*, 2023). To generalize a single model across multiple batteries with heterogeneous degradation profiles, the battery's identity is integrated as a learnable input feature. This approach is framed within the paradigm of Representation Learning, which aims to acquire significant latent features in a compact, low-dimensional space (Chen *et al.*, 2025; Cheng, 2025).

Unlike the positional embeddings used in Transformer architectures to encode sequence order, our identity embedding captures the static and unique characteristics of each battery. The process is as follows:

– Numerical Labeling: Each unique battery identifier (e.g., 'B0005') is assigned an integer numerical label.

– Embedding Layer: This label is fed into a Keras Embedding layer. This layer maps the categorical identifier to a dense, learnable vector in a latent space of a fixed dimension.

– Identity Vector: During training, the model adjusts this vector to act as a "signature" that captures the intrinsic degradation characteristics of that specific battery.

– Concatenation: This identity vector is then concatenated with the time-series data at each step, allowing a single, unified model to learn both the general degradation patterns and the nuances of each unit.

Finally, figure 3 visually summarizes the entire methodology, from raw-data processing to final prediction. It depicts five key preprocessing stages that produce training, validation, and test sets (80%, 10%, 10%). The model architecture is then detailed: the cycle sequence and the battery identity (encoded via an Embedding layer) enter as parallel inputs. For both base and hybrid models, the data flow is shown—optionally passing the sequence through a CNN block, concatenating it with the identity vector, and finally processing it with a recurrent block (GRU or LSTM)

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

Box 3

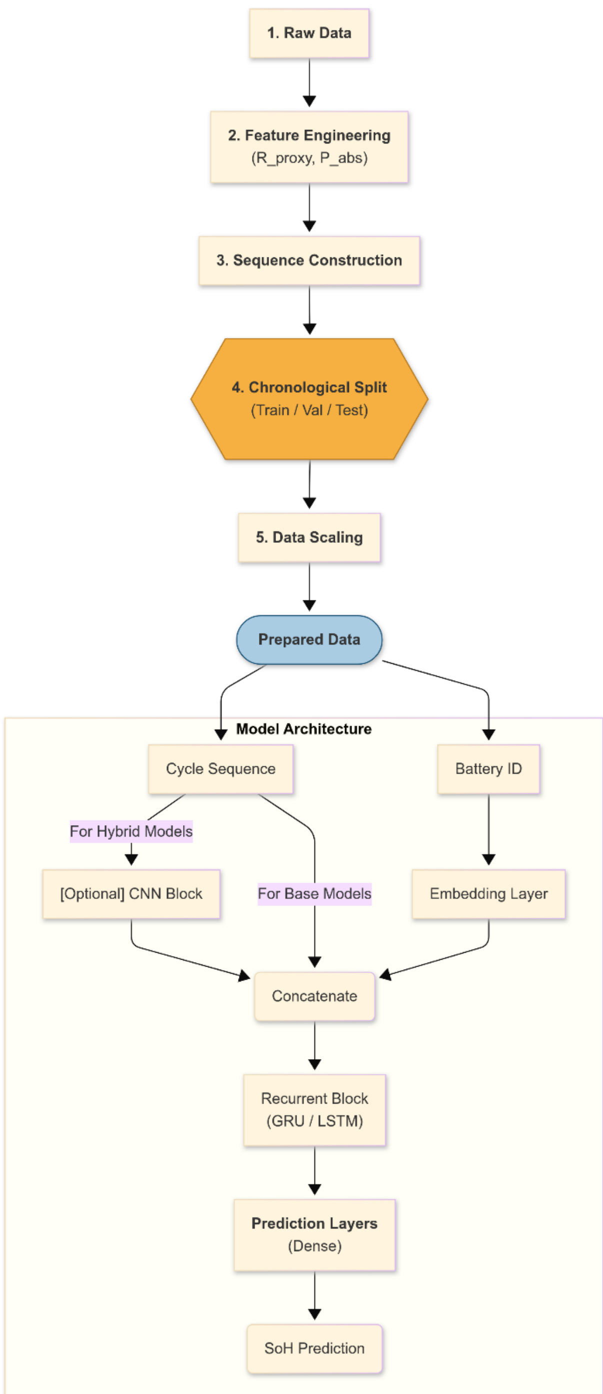


Figure 3
Sequence Construction and Hybrid Architecture for SoH Prediction
Source: Own elaboration

Results

For each of the four deep architectures (GRU, CNN-GRU, LSTM, CNN-LSTM), we performed hyperparameter tuning to optimize predictive performance. Through an experimental process, the values that minimized the validation error were selected. The final hyperparameters used for training each model are summarized in Table 1.

Box 4

	GRU	CNN + GRU	LSTM	CNN + LSTM
Sequence	15	5	5	5
Batch	8	64	8	64
LR	5x 10^{-4}	1x 10^{-3}	5x 10^{-4}	5x 10^{-4}
Units	128	128	256	128
Filters	NA	64	NA	64
Kernel	NA	5	NA	3

Table 1
Final hyperparameters used for each deep model
Source: Own Elaboration

Using these configurations, we evaluated predictive performance on the test sets with MAE, RMSE, MAPE, R^2. As shown in Table 2, all models captured the battery degradation trend with high fidelity R^2 values are mostly above 0.96 (one case at 0.9593) and low absolute errors across batteries.

Box 5

	GRU	CNN-GRU	LSTM	CNN-LSTM
MAE				
B0005	0.0043	0.0081	0.0170	0.0065
B0006	0.0140	0.0129	0.0158	0.0152
B0007	0.0058	0.0073	0.0057	0.0054
B0018	0.0034	0.0110	0.0083	0.0080
MAPE				
B0005	0.0050	0.0097	0.0203	0.0078
B0006	0.0196	0.0173	0.0216	0.0213
B0007	0.0066	0.0084	0.0066	0.0062
B0018	0.0040	0.013	0.0099	0.0095
RMSE				
B0005	0.0054	0.0104	0.0191	0.0090
B0006	0.0176	0.0162	0.0192	0.0200
B0007	0.0070	0.0094	0.0076	0.0078
B0018	0.0043	0.0161	0.0110	0.0125
R^2				
B0005	0.9968	0.9892	0.9641	0.9918
B0006	0.9742	0.9814	0.9738	0.9716
B0007	0.9923	0.9872	0.9915	0.9910
B0018	0.9962	0.9593	0.9807	0.9754

Table 2
Comparison of error metrics and coefficient of determination by Model and Battery
Source: Own Elaboration

Finally, as shown in Figs. 4–7, it is confirmed that all architectures exhibit strong generalization to SoH degradation, with small cycle-level errors. GRU provides the best balance between responsiveness and stability with tight, well-centered residuals and minimal lag resulting in the lowest average MAE/RMSE and the highest average R^2 across batteries. CNN-GRU handles rapid end-of-life transitions particularly well (notably for B0006).

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

CNN-LSTM yields the smoothest trajectories and achieves the lowest MAE/MAPE for B0007, albeit with slight over-smoothing in sharp variations. In contrast, plain LSTM follows the trend but is outperformed by the other configurations.

– B0005: GRU achieves the best MAE/RMSE and highest R^2.
– B0006: CNN-GRU achieves the best MAE/RMSE and highest R^2.
– B0007: CNN-LSTM has the lowest MAE and MAPE; GRU attains the lowest RMSE and the highest R^2.
– B0018: GRU attains the lowest MAE/RMSE and the highest R^2.

Compared with CNN–LSTM approaches that rely on incremental-capacity inputs and feature-selection stages (e.g., Xu *et al*., 2023), our study employs an identity embedding and simple physics-informed features (R_{proxy}, P_{abs}) to enable a single end-to-end model without IC preprocessing or modal decompositions. Relative to CS–VMD–GRU pipelines (Ding *et al*., 2022), we avoid variational mode decomposition and per-IMF training, thereby reducing modeling complexity and pre-analysis overhead.

In contrast to Alharbi *et al*. (2025), who reported the most consistent centralized performance for 1D-CNN across a broader NASA subset (including B0036) under a 90/10 split, our experiments on B0005, B0006, B0007, B0018 indicate that GRU achieves the best average MAE, RMSE and R^2, CNN-GRU excels on rapid end-of-life transitions (B0006), and CNN-LSTM attains the lowest MAE/MAPE on B0007.

Finally, unlike SA-LSTM + particle-filter hybrids (Ravi *et al*., 2022), our models do not incorporate explicit probabilistic filtering or attention mechanisms, prioritizing architectural simplicity and ease of deployment.

Box 6

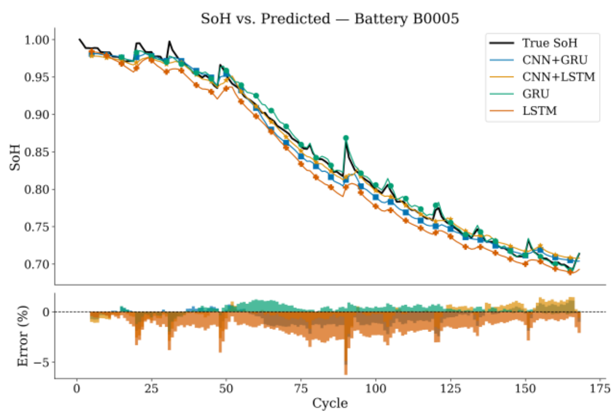


Figure 4
SoH Prediction vs. Ground Truth-Battery B0005
Source: Own Elaboration

Box 7

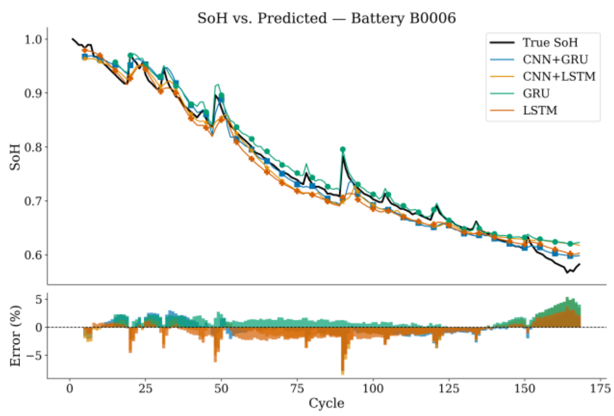


Figure 5
SoH Prediction vs. Ground Truth-Battery B0006
Source: Own Elaboration

Box 8

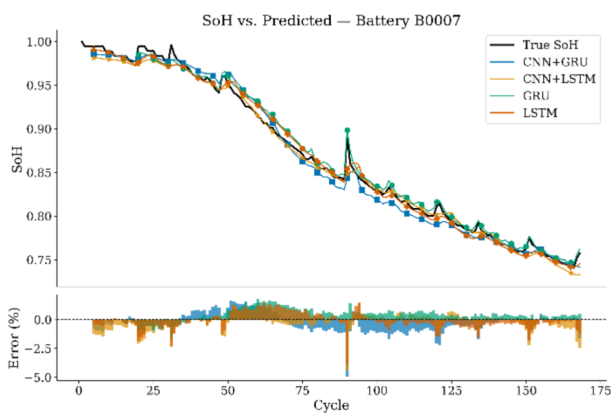


Figure 6
SoH Prediction vs. Ground Truth-Battery B0007
Source: Own Elaboration

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

Box 9

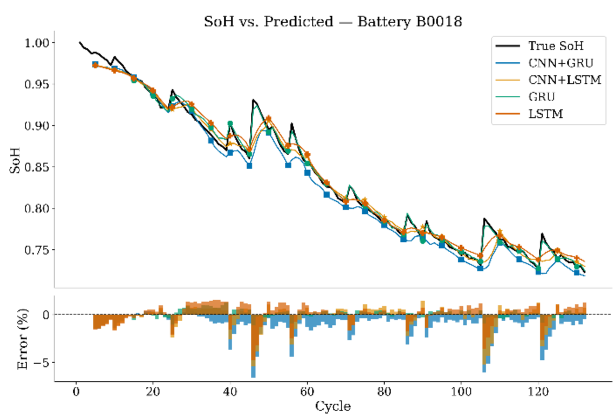


Figure 7
SoH Prediction vs. Ground Truth-Battery B0018
Source: Own Elaboration

Conclusions

We introduced a unified deep-learning approach for SoH estimation that combines battery-identity embedding with physics-informed features (R_{proxy} , P_{abs}). On NASA cells B0005/B0006/B0007/B0018, all architecture generalizes well; on average, GRU attains the lowest MAE/RMSE and highest R^2 . GRU's strong average accuracy, coupled with its lower parametric complexity, makes it an attractive choice for embedded BMS deployments. CNN-GRU excels under rapid end-of-life transitions, and CNN-LSTM achieves the best MAE/MAPE on B0007.

Limitations & future work:

- Validate across chemistries, temperatures, and real drive cycles (beyond NASA, fixed conditions).
- Add uncertainty quantification and calibration for deployment.
- Explore domain adaptation/continual learning and ablations on embedding/window/features

Declarations

Conflict of interest

The authors declare no interest conflict. They have no known competing financial interests or personal relationships that could have appeared to influence the article reported in this article.

Author contribution

The contribution of each researcher in each of the points developed in this research were defined based on:

Medina-Martínez Sergio Iván: Conceptualization; Methodology; Software; Validation; Formal analysis; Writing original draft.

Juárez Toledo Carlos: Provided advice on the models and on the background for the state of the art. Supported the development of the method.

Carrillo-Martínez, Irma: Assisted in the data analysis and the assembly model and reviewed, corrected the article.

Hernández-Epigmenio, Miguel Angel: Manuscript review and feedback

Availability of data and materials

The datasets used in this study are publicly available from NASA's Li-ion Battery Aging Datasets

Funding

No internal or external funding was provided for this research project; however, it was made possible thanks to the postgraduate studies scholarship granted by la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI).

Acknowledgements

To la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI).

Abbreviations

AI	Artificial Intelligence
ALA	Artificial Lemming Algorithm
BMS	Battery Management System
CC	Constant Current
CNN	Convolutional Neural Network
CS	Cuckoo Search
CV	Constant Voltage
EOL	End-of-Life
EV	Electric Vehicle
FL	Federated Learning
GRU	Gated Recurrent Unit
IC	Incremental Capacity

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

IE	Identity Embedding
IMF	Intrinsic Mode Function
LFP	Lithium Iron Phosphate (LiFePO4)
LIB	Lithium-Ion Battery
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NASA PCoE	NASA Prognostics Center of Excellence
NCA	Lithium Nickel Cobalt Aluminum Oxide (LiNiCoAlO2)
NCM	Lithium Nickel Cobalt Manganese Oxide (LiNiMnCoO2)
Pabs	Integrated Absolute Power (energy proxy)
PF	Particle Filter
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
Rproxy	Internal-Resistance Proxy
RUL	Remaining Useful Life
SA-LSTM	Self-Attention LSTM
SVR	Support Vector Regression
VMD	Variational Mode Decomposition
1D-CNN	One-Dimensional CNN

References

Antecedents

Luo, K., Chen, X., Zheng, H., & Shi, Z. (2022). A review of deep learning approach to predicting the state of health and state of charge of lithium-ion batteries. *Journal of Energy Chemistry*, 74, 159–173.

Medina-Martínez, S. I., Juárez-Toledo, C., & Martínez-Carrillo, I. (2024). Estimation of lithium-ion battery state of health using an ensemble model integrated with random forest and linear regression techniques. *Quantitative and Statistical Analysis*.

Ralls, A. M., Leong, K., Clayton, J., Fuelling, P., Mercer, C., Navarro, V., & Menezes, P. L. (2023). The Role of Lithium-Ion Batteries in the Growing Trend of Electric Vehicles. *Materials*, 16(17).

Shen, X., Liu, H., Cheng, X. B., Yan, C., & Huang, J. Q. (2018). Beyond lithium ion batteries: Higher energy density battery systems based on lithium metal anodes. *Energy Storage Materials*, 12, 161–175.

Tran, M. K., Panchal, S., Khang, T. D., Panchal, K., Fraser, R., & Fowler, M. (2022). Concept Review of a Cloud-Based Smart Battery Management System for Lithium-Ion Batteries: Feasibility, Logistics, and Functionality. *Batteries*, 8(2), 19.

Yang, S., Zhang, Z., Cao, R., Wang, M., Cheng, H., Zhang, L., Jiang, Y., Li, Y., Chen, B., Ling, H., Lian, Y., Wu, B., & Liu, X. (2021). Implementation for a cloud battery management system based on the CHAIN framework. *Energy and AI*, 5, 100088.

Basics

Ding, G., Wang, W., & Zhu, T. (2022). Remaining Useful Life Prediction for Lithium-Ion Batteries Based on CS-VMD and GRU. *IEEE Access*, 10, 89402–89413.

Saha, B., & Goebel, K. (2007). Battery Data Set NASA Ames Research Center.

Supports

Chen, B., Zhang, Y., Wu, J., Yuan, H., & Guo, F. (2025). Lithium-Ion Battery State of Health Estimation Based on Feature Reconstruction and Transformer-GRU Parallel Architecture. *Energies*, 18(5), 1236.

Cheng, S. (2025). A Hybrid Deep Learning Method for the Estimation of the State of Health of Lithium-Ion Batteries. *International Transactions on Electrical Energy Systems*, 2025(1), 2442893.

Peng, S., Sun, Y., Liu, D., Yu, Q., Kan, J., & Pecht, M. (2023). State of health estimation of lithium-ion batteries based on multi-health features extraction and improved long short-term memory neural network. *Energy*, 282, 128956.

Yu, Z., & Pan, Z. (2025). Joint Estimation of SOH and RUL for Lithium-ion Batteries using CNN-GRU Based on Artificial Lemming Algorithm. 1269–1274.

Differences

Alharbi, T., Umair, M., & Alharbi, A. (2025). Lithium-Ion Battery State of Health Degradation Prediction Using Deep Learning Approaches. *IEEE Access*, 13, 13464–13481.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109. https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

Ravi, S., Koushik, P. G. V., Varma, V. C. K., & Rajeswari, S. (2022). Li-ion Batteries SoH Estimation Using LSTM. *CSITSS 2022*.

Xu, H., Wu, L., Xiong, S., Li, W., Garg, A., & Gao, L. (2023). An improved CNN-LSTM model-based state-of-health estimation approach for lithium-ion batteries. *Energy*, 276, 127585.

Discussions

Eleftheriadis, P., Gangi, M., Leva, S., Rey, A. V., Groppo, E., & Grande, L. (2024). Comparative study of machine learning techniques for the state of health estimation of Li-Ion batteries. *Electric Power Systems Research*, 235, 110889.

Kong, X., Chen, Z., Liu, W., Ning, K., Zhang, L., Syauqie, ·, Marier, M., Liu, Y., Chen, Y., & Feng Xia. (2025). Deep learning for time series forecasting: a survey. *International Journal of Machine Learning and Cybernetics*, 16, 5079–5112.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel. [2025]. Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study. Journal of Computational Technologies. 9[22]1-9: e4922109.
https://doi.org/10.35429/JOCT.2025.9.22.4.1.9

[[Title in TNRoman and Bold No. 14 in English and Spanish]

Surname, Name 1st Author*[a], Surname, Name 1st Co-author[b], Surname, Name 2nd Co-author[c] and Surname, Name 3rd Co-author[d] [No.12 TNRoman]

[a] ROR Affiliation institution, Researcher ID, ORC ID, SNI-SECIHTI ID or CVU PNPC [No.10 TNRoman]
[b] ROR Affiliation institution, Researcher ID, ORC ID, SNI-SECIHTI ID or CVU PNPC [No.10 TNRoman]
[c] ROR Affiliation institution, Researcher ID, ORC ID, SNI-SECIHTI ID or CVU PNPC [No.10 TNRoman]
[d] ROR Affiliation institution, Researcher ID, ORC ID, SNI-SECIHTI ID or CVU PNPC [No.10 TNRoman]

All ROR-Clarivate-ORCID and SECIHTI profiles must be hyperlinked to your website.
Prot- ROR University of South Australia • 7038-2013• 0000-0001-6442-4409 • 416112

SECIHTI classification:
https://marvid.org/research_areas.php [No.10 TNRoman]
Area:
Field:
Discipline:
Subdiscipline:

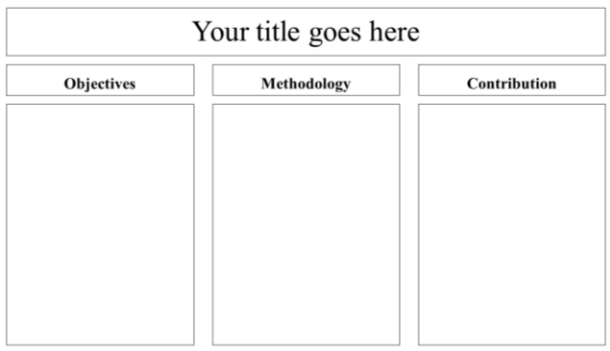
DOI: https://doi.org/
Article History:
Received: [Use Only ECORFAN]
Accepted: [Use Only ECORFAN]
Contact e-mail address:
* ✉ [example@example.org]

Abstract [In English]
Must contain up to 150 words
Graphical abstract [In English]

Authors must provide an original image that clearly represents the article described in the article. Graphical abstracts should be submitted as a separate file. Please note that, as well as each article must be unique. File type: the file types are MS Office files.No additional text, outline or synopsis should be included. Any text or captions must be part of the image file. Do not use unnecessary white space or a "graphic abstract" header within the image file.

Keywords [In English]
Indicate 3 keywords in TNRoman and Bold No. 10

Abstract [In Spanish]
Must contain up to 150 words
Graphical abstract [In Spanish]



Authors must provide an original image that clearly represents the article described in the article. Graphical abstracts should be submitted as a separate file. Please note that, as well as each article must be unique. File type: the file types are MS Office files.No additional text, outline or synopsis should be included. Any text or captions must be part of the image file. Do not use unnecessary white space or a "graphic abstract" header within the image file.

Keywords [In Spanish]
Indicate 3 keywords in TNRoman and Bold No. 10

Citation: Surname, Name 1st Author, Surname, Name 1st Co-author, Surname, Name 2nd Co-author and Surname, Name 3rd Co-author. Article Title. Journal of Computational Technologies. Year. V-N: Pages [TN Roman No.10].

Introduction

Text in TNRoman No.12, single space.

General explanation of the subject and explain why it is important.

What is your added value with respect to other techniques?

Clearly focus each of its features.

Clearly explain the problem to be solved and the central hypothesis.

Explanation of sections Article.

Development of headings and subheadings of the article with subsequent numbers

[Title No.12 in TNRoman, single spaced and bold]

Products in development No.12 TNRoman, single spaced.

Including figures and tables-Editable

In the article content any table and figure should be editable formats that can change size, type and number of letter, for the purposes of edition, these must be high quality, not pixelated and should be noticeable even reducing image scale.

[Indicating the title at the bottom with No.10 and Times New Roman Bold]

Box

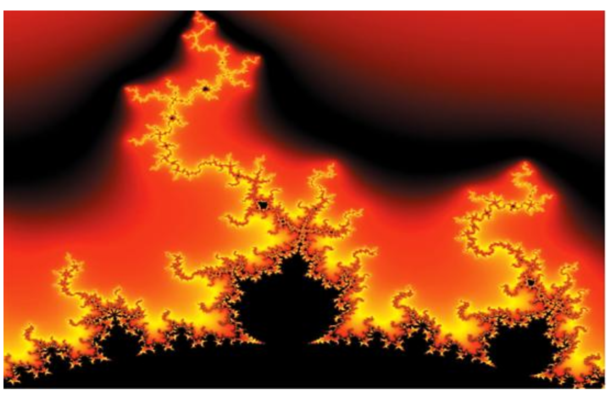


Figure 1

Title [Should not be images-everything must be editable]

Source [in italic]

Box

Table 1

Title [Should not be images-everything must be editable]

Source [in italic]

The maximum number of Boxes is 10 items

For the use of equations, noted as follows:

$$Y_{ij} = \alpha + \sum_{h=1}^{r} \beta_h X_{hij} + u_j + e_{ij} \quad [1]$$

Must be editable and number aligned on the right side.

Methodology

Develop give the meaning of the variables in linear writing and important is the comparison of the used criteria.

Results

The results shall be by section of the article.

Conclusions

Clearly explain the results and possibilities of improvement.

Annexes

Tables and adequate sources.

The international standard is 7 pages minimum and 14 pages maximum.

Declarations

Conflict of interest

The authors declare no interest conflict. They have no known competing financial interests or personal relationships that could have appeared to influence the article reported in this article.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Author contribution

Specify the contribution of each researcher in each of the points developed in this research.

Prot-
Benoit-Pauleter, Gerard: Contributed to the project idea, research method and technique.

Availability of data and materials

Indicate the availability of the data obtained in this research.

Funding

Indicate if the research received some financing.

Acknowledgements

Indicate if they were financed by any institution, University or company.

Abbreviations

List abbreviations in alphabetical order.

Prot-

ANN	Artificial Neural Network

References

Use APA system. Should not be numbered, nor with bullets, however if necessary numbering will be because reference or mention is made somewhere in the Article.

Use the Roman alphabet, all references you have used should be in Roman alphabet, even if you have cited an article, book in any of the official languages of the United Nations [English, French, German, Chinese, Russian, Portuguese, Italian, Spanish, Arabic], you should write the reference in Roman alphabet and not in any of the official languages.

Citations are classified the following categories:

Antecedents. The citation is due to previously published research and orients the citing document within a particular scholarly area.

Basics. The citation is intended to report data sets, methods, concepts and ideas on which the authors of the citing document base their work.

Supports. The citing article reports similar results. It may also refer to similarities in methodology or, in some cases, to the reproduction of results.

Differences. The citing document reports by means of a citation that it has obtained different results to those obtained in the cited document. This may also refer to differences in methodology or differences in sample sizes that affect the results.

Discussions. The citing article cites another study because it is providing a more detailed discussion of the subject matter.

The URL of the resource is activated in the DOI or in the title of the resource.

Prot-
Mandelbrot, B. B. [2020]. Negative dimensions and Hölders, multifractals and their Hölder spectra, and the role of lateral preasymptotics in science. Journal of Fourier Analysis and Applications Special. 409-432.

Intellectual Property Requirements for editing:

- Authentic Signature in Color of Originality Format Author and Coauthors.

- Authentic Signature in Color of the Acceptance Format of Author and Coauthors.

- Authentic Signature in blue color of the Conflict of Interest Format of Author and Co-authors.

ISSN: 2523-6814
RENIECYT-SECIHTI: 1702902

Reservation to Editorial Policy

Journal of Computational Technologies reserves the right to make editorial changes required to adapt the Articles to the Editorial Policy of the Research Journal. Once the Article is accepted in its final version, the Research Journal will send the author the proofs for review. ECORFAN® will only accept the correction of errata and errors or omissions arising from the editing process of the Research Journal, reserving in full the copyrights and content dissemination. No deletions, substitutions or additions that alter the formation of the Article will be accepted.

Code of Ethics - Good Practices and Declaration of Solution to Editorial Conflicts

Declaration of Originality and unpublished character of the Article, of Authors, on the obtaining of data and interpretation of results, Acknowledgments, Conflict of interests, Assignment of rights and Distribution.

The ECORFAN-Mexico, S.C Management claims to Authors of Articles that its content must be original, unpublished and of Scientific, Technological and Innovation content to be submitted for evaluation.

The Authors signing the Article must be the same that have contributed to its conception, realization and development, as well as obtaining the data, interpreting the results, drafting and reviewing it. The Corresponding Author of the proposed Article will request the form that follows.

Article title:

- The sending of an Article to Journal of Computational Technologies emanates the commitment of the author not to submit it simultaneously to the consideration of other series publications for it must complement the Format of Originality for its Article, unless it is rejected by the Arbitration Committee, it may be withdrawn.

- None of the data presented in this article has been plagiarized or invented. The original data are clearly distinguished from those already published. And it is known of the test in PLAGSCAN if a level of plagiarism is detected Positive will not proceed to arbitrate.

- References are cited on which the information contained in the Article is based, as well as theories and data from other previously published Articles.

- The authors sign the Format of Authorization for their Article to be disseminated by means that ECORFAN-Mexico, S.C. In its Holding Taiwan considers pertinent for disclosure and diffusion of its Article its Rights of Work.

- Consent has been obtained from those who have contributed unpublished data obtained through verbal or written communication, and such communication and Authorship are adequately identified.

- The Author and Co-Authors who sign this work have participated in its planning, design and execution, as well as in the interpretation of the results. They also critically reviewed the paper, approved its final version and agreed with its publication.

- No signature responsible for the work has been omitted and the criteria of Scientific Authorization are satisfied.

- The results of this Article have been interpreted objectively. Any results contrary to the point of view of those who sign are exposed and discussed in the Article.

Copyright and Access

The publication of this Article supposes the transfer of the copyright to ECORFAN-Mexico, SC in its Holding Taiwan for itsJournal of Computational Technologies, which reserves the right to distribute on the Web the published version of the Article and the making available of the Article in This format supposes for its Authors the fulfilment of what is established in the Law of Science and Technology of the United Mexican States, regarding the obligation to allow access to the results of Scientific Research.

Article Title:

Name and Surnames of the Contact Author and the Coauthors	Signature
1.	
2.	
3.	
4.	

Principles of Ethics and Declaration of Solution to Editorial Conflicts

Editor Responsibilities

The Publisher undertakes to guarantee the confidentiality of the evaluation process, it may not disclose to the Arbitrators the identity of the Authors, nor may it reveal the identity of the Arbitrators at any time.

The Editor assumes the responsibility to properly inform the Author of the stage of the editorial process in which the text is sent, as well as the resolutions of Double-Blind Review.

The Editor should evaluate manuscripts and their intellectual content without distinction of race, gender, sexual orientation, religious beliefs, ethnicity, nationality, or the political philosophy of the Authors.

The Editor and his editing team of ECORFAN® Holdings will not disclose any information about Articles submitted to anyone other than the corresponding Author.

The Editor should make fair and impartial decisions and ensure a fair Double-Blind Review.

Responsibilities of the Editorial Board

The description of the peer review processes is made known by the Editorial Board in order that the Authors know what the evaluation criteria are and will always be willing to justify any controversy in the evaluation process. In case of Plagiarism Detection to the Article the Committee notifies the Authors for Violation to the Right of Scientific, Technological and Innovation Authorization.

Responsibilities of the Arbitration Committee

The Arbitrators undertake to notify about any unethical conduct by the Authors and to indicate all the information that may be reason to reject the publication of the Articles. In addition, they must undertake to keep confidential information related to the Articles they evaluate.

Any manuscript received for your arbitration must be treated as confidential, should not be displayed or discussed with other experts, except with the permission of the Editor.

The Arbitrators must be conducted objectively, any personal criticism of the Author is inappropriate.

The Arbitrators must express their points of view with clarity and with valid arguments that contribute to the Scientific, Technological and Innovation of the Author.

The Arbitrators should not evaluate manuscripts in which they have conflicts of interest and have been notified to the Editor before submitting the Article for Double-Blind Review.

Responsibilities of the Authors

Authors must guarantee that their articles are the product of their original work and that the data has been obtained ethically.

Authors must ensure that they have not been previously published or that they are not considered in another serial publication.

Authors must strictly follow the rules for the publication of Defined Articles by the Editorial Board.

The authors have requested that the text in all its forms be an unethical editorial behavior and is unacceptable, consequently, any manuscript that incurs in plagiarism is eliminated and not considered for publication.

Authors should cite publications that have been influential in the nature of the Article submitted to arbitration.

Information services

Indexation - Bases and Repositories

RESEARCH GATE (Germany)
GOOGLE SCHOLAR (Citation indices-Google)
MENDELEY (Bibliographic References Manager)
REDIB (Ibero-American Network of Innovation and Scientific Knowledge- CSIC)
HISPANA (Information and Bibliographic Orientation-Spain)

Publishing Services

Citation and Index Identification H
Management of Originality Format and Authorization
Testing Article with PLAGSCAN
Article Evaluation
Certificate of Double-Blind Review
Article Edition
Web layout
Indexing and Repository
ArticleTranslation
Article Publication
Certificate of Article
Service Billing

Editorial Policy and Management

69 Street. YongHe district, ZhongXin. Taipei-Taiwan. Phones: +52 1 55 6159 2296, +52 1 55 1260 0355, +52 1 55 6034 9181; Email: contact@ecorfan.org www.ecorfan.org

ECORFAN®

Chief Editor
Quintanilla - Cóndor, Cerapio. PhD

Executive Director
Ramos-Escamilla, María. PhD

Editorial Director
Peralta-Castro, Enrique. MsC

Web Designer
Escamilla-Bouchan, Imelda. PhD

Web Diagrammer
Luna-Soto, Vladimir. PhD

Editorial Assistant
Soriano-Velasco, Jesús. BsC

Philologist
Ramos-Arancibia, Alejandra. BsC

Advertising & Sponsorship

(ECORFAN® Taiwan), sponsorships@ecorfan.org

Site Licences

03-2010-032610094200-01-For printed material ,03-2010-031613323600-01-For Electronic material,03-2010-032610105200-01-For Photographic material,03-2010-032610115700-14-For the facts Compilation,04-2010-031613323600-01-For its Web page,19502-For the Iberoamerican and Caribbean Indexation,20-281 HB9-For its indexation in Latin-American in Social Sciences and Humanities,671-For its indexing in Electronic Scientific Journals Spanish and Latin-America,7045008-For its divulgation and edition in the Ministry of Education and Culture-Spain,25409-For its repository in the Biblioteca Universitaria-Madrid,16258-For its indexing in the Dialnet,20589-For its indexing in the edited Journals in the countries of Iberian-America and the Caribbean, 15048-For the international registration of Congress and Colloquiums. financingprograms@ecorfan.org

Management Offices

69 Street. YongHe district, ZhongXin. Taipei-Taiwan.

Journal of Computational Technologies

"Methodological framework for short-term electricity demand forecasting in desert regions using lstm and meteorological inputs"

Cabrera, Misael, Martinez, Ulises, Woocay, Arturo[c] and Valles, Delia
Instituto Tecnológico de Ciudad Juárez
Texas A&M University–Kingsville

"Development of a Tool for Academic Data Processing and Visualization: A Python-Based Proposal"

Salazar-Uitz, Ricardo Rubén, Ramos-Ramos, Rudy Mauricio, Shih, Meng Yen and Lezama-Zarraga, Francisco Román
Universidad Autónoma de Campeche

"Data Science Methodologies: The Relevance of CRISP-DM and Its Comparison with Emerging Models"

González-Ramírez, Claudia Teresa, Coria-Tavira, Alondra, Ruiz-Garduño, Jhacer Kharen and Viñas-Alvarez, Samuel Efrén
Tecnológico Nacional de México

"Generalizable SoH Estimation for Li-ion Batteries via Identity Embeddings: A CNN/GRU/LSTM Comparative Study"

Medina-Martínez, Sergio Iván, Juárez-Toledo, Carlos, Martínez-Carrillo, Irma and Hernández-Epigmenio, Miguel Angel
Universidad Autónoma del Estado de México

ISSN 2523-6814
www.ecorfan.org