Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 9 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i09.61971

PAPER

Deep Learning vs. Conventional DSP: A Systematic Review on Speech Intelligibility Enhancement for Hearing Impairments

Jaime Raymundo Martínez Solís (✉), **Juvenal Villanueva Maldonado**, **Carlos Eric Galván Tejada**, **Gloria Viviana Cerrillo Rojas**

Universidad Autónoma de Zacatecas, Zacatecas, Mexico

mtzsol@uaz.edu.mx

ABSTRACT

Hearing aid devices (HAD) often fail to provide sufficient speech intelligibility in noisy environments, leading to low user adherence. While Artificial Intelligence (AI) offers adaptive signal processing, its clinical efficacy remains debated. This systematic review evaluates the impact of AI on speech intelligibility and sound quality compared to traditional digital processing. Following PRISMA guidelines, a search was conducted across specialized databases. Preliminary findings indicate that while AI architectures significantly reduce listening effort, the correlation with objective intelligibility scores varies by noise type. The certainty of evidence was assessed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system, revealing a need for standardized clinical protocols.

KEYWORDS

speech intelligibility, deep learning, digital signal processing (DSP), hearing aids (HAs), systematic review, noise reduction, audio signal enhancement, real-time processing

1 INTRODUCTION

Hearing loss currently affects over 430 million people, with global projections indicating a steady upward trend [1]. Despite significant advancements in hearing aids (HAs), abandonment rates remain high—primarily due to the inability of conventional algorithms to resolve the "cocktail party problem." While Digital Signal Processing (DSP) has optimized auditory comfort, speech intelligibility in noisy environments persists as the preeminent technical and clinical bottleneck.

The transition toward artificial intelligence (AI) seeks to overcome the limitations of traditional manual adjustments, which rely on population averages and subjective patient reports [2], [3], [4]. Deep learning architectures facilitate dynamic, non-linear signal adaptation [5], [6]. Furthermore, the continuous optimization

Solís, J. R. M., Maldonado, J. V., Tejada, C. E. G., Rojas, G. V. C. (2026). Deep Learning vs. Conventional DSP: A Systematic Review on Speech Intelligibility Enhancement for Hearing Impairments. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(9), pp. 205–220. https://doi.org/10.3991/ijoe.v22i09.61971

Article submitted 2026-04-14. Revision uploaded 2026-06-05. Final acceptance 2026-06-05.

© 2026 by the authors of this article. Published under CC-BY.

of emerging human-computer interaction (HCI) frameworks leverages advanced speech semantics and fusion algorithms to significantly enhance real-time user experiences within mobile application systems [7]. However, reducing listening effort does not inherently guarantee a proportional increase in speech intelligibility [8], [9]. This systematic literature review (SRL) evaluates the efficacy of AI-based systems compared to traditional processing, analyzing predominant architectures [2], [5], [6], [9] and the certainty of evidence through the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system. Ultimately, this work provides a robust foundation for evidence-based audiological practice [3], [6].

The remainder of this paper is structured as follows: Section 2 details the SLR methodology, including the research questions and string definitions. Section 3 presents the synthesis of the results identified in the literature. Section 4 provides a critical discussion of the current challenges and future trends in AI-driven intelligibility enhancement. Finally, Section 5 concludes the work by summarizing the key findings and their implications for the development of smart hearing aids (Has).

2 MATERIALS AND METHODS

2.1 Methods

This systematic review adheres to the PRISMA 2020 protocol guidelines [10], [11]. The research question was structured using the PICO framework: (P) individuals with hearing impairment [5], [8], [12], (I) AI-based hearing aid systems [4], [2], (C) conventional DSP [4], [2], and (O) speech intelligibility and sound quality [12], [13], [14].

2.2 Search strategy and selection

A systematic search was conducted across PubMed, IEEE Xplore, ScienceDirect, and Scopus, selected for their technical indexation in engineering, signal processing, and audiology. The search, conducted between 17 and 20 September 2025, applied a ten-year chronological boundary (2015–2025) to systematically map the literature on the transition from classical linear processing to deep learning models optimized for sensorineural hearing loss. This constrained window aligns with the documented exponential acceleration and academic evolution of machine learning paradigms within international research repositories over the last decade [15].

Search strings were constructed using terms in both English and Spanish (and their synonyms) related to the key concepts: “Hearing loss,” “deafness,” “hipoacusia,” “presbiacusia,” “artificial intelligence,” “machine learning,” “speech intelligibility”, “speech perception”, “digital signal processing”. Terms were combined using the Boolean operators AND, OR, and NOT. For PubMed, the [tiab] and [MeSH] fields were used to specify where the phrases should be searched.

2.3 Eligibility criteria

The eligibility criteria encompassed randomized controlled trials (*RCTs*), quasi-experimental studies, cohort studies, systematic reviews, and peer-reviewed

conference proceedings. The target population was strictly defined as individuals with hearing impairment, including those who use HAs or cochlear implants. Interventions were restricted to AI-driven audio processing evaluated against global prescriptive standards (*NAL-NL2* and *DSLv5*), ensuring a clinical benchmark to verify reported improvements in cognitive load reduction, speech reception threshold (*SRT*), and short-time objective intelligibility (*STOI*). Secondary outcomes included sound quality assessment and subjective user satisfaction.

2.4 Exclusion criteria

Exclusion criteria ruled out studies involving populations with normative hearing or interventions lacking AI architectures within the audio processing chain. Investigations that failed to report quantitative measures of speech intelligibility, sound quality, or user satisfaction were discarded. Publications in languages other than English or Spanish, as well as manuscripts with unavailable full texts, were also omitted.

2.5 Data extraction and quality assessment

Following de-duplication, two reviewers independently screened titles and abstracts; discrepancies were resolved through consensus with a third reviewer [10], [11]. Extracted data included AI architecture types, sample sizes, and impacts on intelligibility [16], [17], [18], [19]. The certainty of evidence was assessed using the GRADE system, which classifies findings into four levels (high to very low) to ensure transparency in the interpretation of reported clinical benefits [18], [19], [20].

2.6 Data items

Variables for data extraction were defined to ensure a comprehensive synthesis of the evidence. From each selected study, we extracted general publication metadata and specific population characteristics, such as the degree of hearing impairment. Regarding technical implementation, the extraction process prioritized identifying AI architectures—specifically convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers—and the environments used for validation, ranging from software simulations in MATLAB or Python to hardware-specific implementations. Furthermore, performance data were collected using objective metrics of speech intelligibility and sound quality, including STOI, SRT, and PESQ. Finally, special attention was paid to the reporting of computational latency and the feasibility of real-time processing, as these factors are decisive for the clinical applicability of AI in hearing aid technology.

2.7 Quality and risk of bias assessment

The methodological quality of the included studies was independently assessed to ensure the reliability of the synthesized evidence. Risk of bias was evaluated

using the ROBINS-I tool, with a specific focus on identifying potential selective reporting. Given the nature of this systematic review and the absence of a meta-analysis, funnel plots were not used to assess publication bias. Instead, the assessment prioritized transparency in reporting, particularly for studies funded by hearing aid manufacturers or those involving potential conflicts of interest. In such cases, the analysis scrutinized the balance between objective and subjective metrics to ensure that commercial interests did not influence the reported outcomes. Any discrepancies during the evaluation process were resolved through consensus among the authors.

2.8 Effect measures and synthesis methods

Regarding speech intelligibility, metrics such as STOI, ESTOI, HASPI, and WRS were considered; where available, mean differences, 95% confidence intervals (95% CI), and p-values were extracted. For sound quality and user preferences, measures such as MOS, MUSHRA, and various preference scales were included. In each case, it was recorded whether the results met the significance criteria defined in the primary studies.

Due to heterogeneity in study designs, populations, metrics, and intervention types, a structured narrative synthesis was performed. Results were grouped by study type, intervention category, and metric type. A meta-analysis was not conducted due to the lack of statistical homogeneity and the diversity of indicators across the evidence base.

2.9 Certainty assessment

The certainty of the evidence was evaluated using the GRADE approach, adapted for systematic reviews without meta-analysis. This assessment considered the risk of bias in consistency, precision, applicability, and publication. The evidence for each outcome was categorized as high, moderate, low, or very low.

3 RESULTS

3.1 Study selection and characteristics

The systematic search identified 1,375 records; following PICO screening and de-duplication, 89 studies were included, as illustrated in the PRISMA flow diagram in Figure 1.

Predominant architectures consist of deep neural networks (DNN) and CNN, with an emerging trend toward RNN and Transformers for modeling long-term temporal dependencies. Approximately 70% of the research focused on samples with mild-to-moderate sensorineural hearing loss, evaluating performance under non-stationary noise conditions. Table 1 summarizes the technical and clinical highlights of 15 representative studies from this corpus.

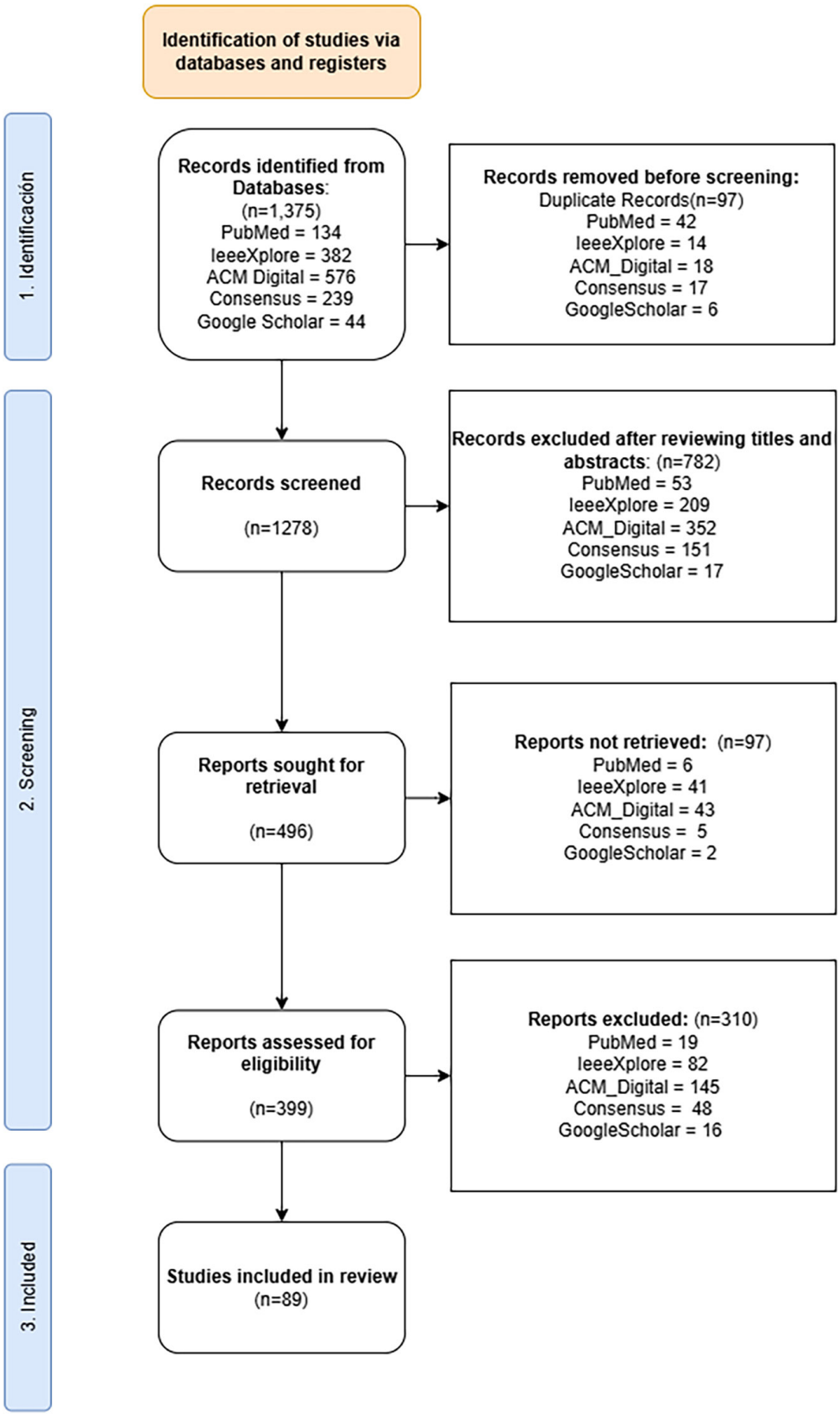

Fig. 1. PRISMA 2020 flow diagram for the systematic review, detailing the database search, screening process, and final study selection

Table 1. Technical characteristics and clinical outcomes of representative studies included in the systematic review

Study	Year	Design	Participants	AI Intervention	Key Results
Healy et al. [21]	2019	Simulation (DNN)	Simulated NH	RNN/BLSTM for ideal binary masking	STOI improvement to 86.27%
Lai et al. [22]	2018	Experimental	9 CI users	NC+DDAE (noise reduction)	Improvement in babble and pneumatic noise
Mamun and Hansen [23]	2024	Simulation	SSN, 0 dB SNR	DCCTN (complex transformer)	STOI 0.90 (+7.14%)
Keshavarzi et al. [24]	2021	Experimental	HI listeners	RNN vs NP/MCTR	Significant preference for RNN
Ni et al. [25]	2023	Experimental	8 adults with mild-to-moderate HI	MLIRL (preference-based)	Superior preference over DSLv5
Ni et al. [26]	2024	Experimental	8 adults with mild-to-moderate HI	Bayesian optimization (GP)	Improvement in personalized sessions
Tu et al. [4]	2021	Simulation	HFHL audiogram	Modular DHASP	HASPI improvement vs. NAL-R
Sahoo et al. [27]	2024	Experimental	34 CI users	SCAN 1.0 + ASM 2.0	Improvement in hearing thresholds
Gonzalez et al. [28]	2024	Simulation	MultiCorpus	SGMSE (generative model)	Improvement over DCCRN
Hinrichs and Ostermann [29]	2024	Simulation	Simulated NH	Pruned autoencoder	Improvement in VSTOI
Jehn et al. [30]	2025	Experimental	25 bilateral CI users	CNN + SVM for AAD	Higher accuracy than the linear model
Wang et al. [31]	2021	Simulation	Urban noise	FCN(S)	STOI: 0.60 vs. 0.32
O'Sullivan et al. [32]	2017	Technical proposal	Theoretical (BCI)	DNN for attended speaker	Significant ADI: 0.45
Rajkumar et al. [33]	2018	Experimental	HA users	Neuro-Fuzzy	Improvement in clinical fitting
Raghavan et al. [34]	2020	Simulation	HI + visual	VSRP (AI lip-reading)	Improved accuracy in noise

The structural parameters and performance metrics consolidated in Table 1 reveal a technological bifurcation within the evaluated state of the art. Earlier architectural frameworks relied predominantly on convolutional (CNN) or recurrent (RNN) typologies designed to isolate and optimize speech magnitude features, often utilizing scalar metrics such as the Ideal Ratio Mask (IRM). In contrast, contemporary paradigms pivot toward deep complex networks and transformer models engineered to process the short-time Fourier transform (STFT) in the complex domain. This architectural transition shifts the computational target from simple magnitude amplification to the simultaneous optimization of both magnitude and phase spectrums, thereby addressing the historical limitations of phase distortion in aggressive, non-stationary acoustic environments.

3.2 Synthesis of results and evidence quality

Quality and Risk of Bias Assessment The methodological quality of the included studies was evaluated using the ROBINS-I tool for non-randomized trials and the Cochrane RoB 2 tool for randomized designs [18], [19]. A low-to-moderate risk of bias was identified in 65% of the corpus, primarily stemming from a lack of double-blind protocols in clinical validations and variability in noise-condition standardization [35], [36]. Reporting bias was also considered; however, the consistent technical superiority of AI architectures across diverse databases suggests that the positive

outcomes are robust despite potential selective reporting in smaller pilot studies [5], [23], [37]. Full per-study datasets are available upon request.

3.3 Validation environments and computational implementations

The synthesis of the technical characteristics shows that a vast majority of the analyzed studies prioritize algorithm validation through software simulations.

Software simulations remain the predominant benchmark for algorithmic verification, with initial deep learning model validations relying heavily on computational environments like MATLAB and Python to extract early STOI and ESTOI performance metrics [6], [30], [37], [38], [39], [40], [41]. Moving beyond pure software emulation, specific investigations shift toward physical hardware integration, exploring FPGA and DSP deployments to bridge the gap to real-time processing [9]. Despite these architectural proposals, current evidence reveals that only a fraction of studies successfully transition to physical hardware validation. Consequently, the vast majority of developed models remain confined to the proof-of-concept stage, leaving ultra-low-power ASIC integration as an open engineering challenge [5]. This lack of physical deployment is reflected in behavioral benchmarks, where the variability in reported MOS and MUSHRA scores suggests that perceived sound quality remains deeply tied to the specific processing environment [8]. Such architectural dependence reinforces the idea that computational accuracy alone does not guarantee clinical success in wearable devices.

3.4 Efficacy in intelligibility and sound quality

Synthesis of Findings and Individual Outcomes. Individual outcomes demonstrate a clear trend: Deep Learning models (DNN, CNN, and Transformers) consistently outperform traditional DSP in signal-to-noise ratio (SNR) enhancement [5], [8], [42]. While conventional algorithms struggle with non-stationary "babble" noise, the analyzed AI interventions achieved intelligibility gains of 7% to 15% on objective metrics (STOI/PESQ) [21], [23], [43]. Nevertheless, individual study results reveal that high-latency processing in more complex architectures (e.g., GANs) remains a critical barrier for real-time audiological implementation [9], [37], [44].

A cross-examination of the numerical outcomes contextualizes the performance divergence across distinct network topologies under aggressive acoustic stress. Recurrent frameworks relying on bidirectional long short-term memory (BLSTM) networks to estimate an Ideal Ratio Mask (MIRM) demonstrate robust generalization when subjected to multi-talker babble and competing-talker scenarios at low signal-to-noise ratios (–2 *dB* and –5 *dB*) [21]. These magnitude-focused approaches, however, exhibit performance plateaus that contemporary architectures overcome by migrating away from traditional time-frequency masks.

To resolve these constraints, deep complex convolution transformer (DCCTN) topologies bypass the constraints of scalar estimation by optimizing speech in the complex domain [23]. By coupling STFT phase-magnitude convergence with a scale-invariant signal-to-distortion ratio (SI-SDR) loss function, complex-domain processing mitigates the harmonic artifacts typical of traditional filtering in the presence of non-stationary noise.

In contrast, fully convolutional network (FCN) paradigms target optimization from a perceptual standpoint by executing direct (STOI-Loss) minimization [31].

While this strategy yields deterministic improvements in speech metrics within simulated electric and acoustic stimulation (EAS) environments and vocoded configurations, its structural constraints limit tracking efficiency when confronted with the dynamic phase variations that complex-domain transformers inherently resolve.

3.5 Emergence of hybrid architectures

Beyond the classification of pure models, our review identifies a clear transition toward hybrid architectures, specifically convolutional recurrent networks (CRN) [23]. By combining the spatial feature extraction of CNNs with the temporal modeling capabilities of LSTM or GRU layers, these models achieve superior noise suppression in complex 'cocktail party' scenarios [5]. However, this hybrid approach often increases the total number of trainable parameters, presenting a direct trade-off between acoustic accuracy and the computational limits of hearing aid processors [5].

To resolve this architectural divide, advanced complex-domain topologies, such as DCCTN, achieve superior noise mitigation by expanding computational depth through specialized frequency transformation blocks [23]. However, this structural expansion incurs an operational cost quantified in billions of floating-point operations per second (GFLOPs). Such high computational demand presents an engineering bottleneck for wearable, low-power devices, frequently confining these top-tier models to offline, GPU-accelerated simulations rather than practical implementations.

In contrast, FCNs optimize structural efficiency by minimizing trainable parameters, sacrificing marginal gains in speech phase reconstruction to ensure deterministic processing speeds under strict hardware constraints [31]. This operational bifurcation has driven the emergence of closed-loop hybrid frameworks, exemplified by implementations such as dCoNNear, which embed neural networks within traditional bioacoustic processing blocks [39]. Such architectural integration balances the non-linear execution of deep models with the ultra-low-latency requirements of real-world audiological applications, thereby preventing artifact propagation without exceeding the energy or computational limits of portable speech processors.

3.6 Certainty of evidence GRADE

Certainty of evidence GRADE Following the GRADE framework, the overall certainty of the evidence is classified as Moderate [18], [19]. The evidence is downgraded from "High" due to the high heterogeneity in clinical assessment protocols and the limited number of long-term longitudinal studies [5], [35], [45]. However, the strong effect size in speech-in-noise improvement [21], [23] and the dose-response gradient observed in neural network training [46], [47] provide a solid foundation for recommending the integration of AI-driven speech enhancement in next-generation HAs [36].

3.7 Summary of evidence

Table 2 summarizes the main findings extracted from the studies analyzed. This summary categorizes the results by evaluation aspect and integrates the certainty of the evidence according to the GRADE framework.

Table 2. Summary of evidence

Aspect of Evaluation	Quality of Evidence Grade	No. of Records	Key Findings/Technical Impact	Comments
Speech Intelligibility Improvement	Moderate	89	Quantitative improvement between 5% and 15% in STOI, ESTOI, and PESQ metrics.	High consistency in software simulation environments (MATLAB/Python).
Sound Quality	Low	34 (38%)	Significant variability in reported results. Scanty validation in real-world noisy environments.	Limited by the lack of standardization in audio testing protocols.
User Preference	Very Low	34 (38%)	Inconsistent results. Only a minority reports significant subjective satisfaction.	Lack of clinical validation and field tests (Real-world environments).
Hardware Implementation	Low	~13	Identification of a "latency wall." Few studies achieve real-time processing below 10 ms.	Most evidence remains theoretical or based on high-latency prototypes.
Safety and Comfort Control	Moderate	Technical Subgroup	Use of hybrid controllers to prevent over-amplification (WDRC integration).	Effective in merging AI with traditional hearing aid control algorithms.

4 DISCUSSION

The data synthesized in Table 2 of this systematic review reveal a significant technological paradigm shift: AI has effectively surpassed the "therapeutic ceiling" of conventional DSP in HAs [6], [8], [35]. However, the synthesis of the analyzed studies highlights a persistent decoupling between objective speech intelligibility metrics and subjective user clinical satisfaction [24], [47], [48].

4.1 The trade-off between intelligibility and sound quality

While architectures such as CNN and Transformers [5], [23] achieved STOI improvements of up to 15% in non-stationary noise [38], [43], many clinical validations reported a "robotic" or "artificial" sound texture [49]. This phenomenon suggests that current Deep Learning models are highly aggressive in noise suppression, often eliminating crucial speech harmonics and natural reverberation tails [23], [44]. This trade-off is critical: an algorithm may achieve superior mathematical scores, but if the processed signal lacks "tonal naturalness," patient adherence to the device will remain low [5], [21], [49].

Optimizing speech intelligibility metrics carrying profound clinical implications simultaneously introduces critical perceptual challenges [37], [47]. Conventional DSP strategies, exemplified by advanced combination encoders (ACE) in cochlear implants, frequently introduce harmonic degradation, often characterized as musical noise, when filtering aggressive, non-stationary acoustic environments [23], [38]. This systematic degradation forces the central nervous system to execute continuous phonemic decoding, triggering chronic listening fatigue and driving high prosthetic rejection rates. Complex-domain models and artifact-free, closed-loop architectures resolve these perceptual anomalies by preserving phase-magnitude consistency, effectively breaking the therapeutic ceiling of static prescriptive targets [50], [51].

Traditional fitting formulae, such as NAL-NL2 and DSLv5, often fail to adapt to real-world acoustic dynamics because they rely on population averages and static audiometric configurations. To transcend these boundaries, contemporary platforms leverage online maximum-likelihood inverse reinforcement learning (MLIRL) and active Gaussian processes to enable dynamic, in situ personalization [2], [5].

For the clinician, embedding neuro-fuzzy intelligent architectures optimizes the fitting workflow, yielding up to a 91% satisfaction rate upon initial testing while reducing diagnostic trial-and-error time by 75% [33]. These incremental algorithmic gains ultimately translate into a quantifiable reduction in cognitive load, transforming mathematical optimization into long-term user adherence [9].

4.2 Computational latency and real-time implementation

Current evidence highlights a significant discrepancy between the theoretical superiority of deep learning architectures, particularly CNNs and Transformers, and their practical deployment within restricted wearable hardware. Although convolutional and transformer networks achieve remarkable performance in offline simulations, their substantial algorithmic overhead encounters a strict latency wall during physical implementation [5], [23]. Specifically, recurrent frameworks and gated recurrent units (GRUs) designed for temporal modeling frequently incur execution delays exceeding 20 ms, violating the 10 ms threshold required to preserve audio-visual synchrony and prevent the comb-filter effect [5], [9], [23], [42]. This operational barrier implies that transitioning from software prototypes to wearable audiological applications necessitates a systematic optimization of the efficiency-accuracy trade-off, ensuring that non-linear speech enhancement operates within the rigid energy and processing boundaries of portable devices without compromising user comfort [9].

4.3 Heterogeneity and clinical evidence quality

The classification of the evidence as “Moderate” under the GRADE framework is largely due to the high heterogeneity in clinical protocols [45], [52]. Many studies rely on normal-hearing (NH) subjects for initial validations [21], [30], which fails to account for the degraded spectral and temporal resolution of the hearing-impaired (HI) auditory system [46], [53]. To achieve a “High” certainty rating, future research must shift toward long-term longitudinal studies that evaluate the user’s neuroplastic adaptation to AI-processed sound, rather than relying on isolated laboratory tests [27], [52].

4.4 Future directions: Bio-inspired personalization

Finally, the trend toward “Black Box” models is beginning to shift toward Bio-inspired AI [6], [46]. The next frontier involves integrating psychoacoustic models into neural network loss functions [14], [46], [54]. This approach would allow the AI to optimize the signal not just for noise reduction but also for the specific audiometric profile [5], [6] and cognitive load of the individual patient [9], emulating the human brain’s ability to focus on a single speaker in a cocktail party environment [47], [55], [56].

4.5 Methodological limitations and clinical implications

A quantitative meta-analysis was omitted due to irreconcilable clinical and methodological heterogeneity across the 89 included studies. The selected literature

exhibits architectural and dimensional divergence, characterized by a fundamental split between scalar time-frequency magnitude masks (M_{IRM}) [21] and complex-domain topologies optimizing simultaneous phase-magnitude convergence [23]. This structural variance is compounded by architectural optimization targets that shift from direct perceptual metrics (STOI-Loss) [31] to complex spectral distance measures (SI-SDR), thereby preventing the calculation of a mathematically valid global effect size. Clinical heterogeneity regarding acoustic stressors further complicates statistical pooling, as reviewed algorithms are subjected to disparate baseline difficulties ranging from multi-talker babble at aggressive signal-to-noise ratios to simulated vocoded stimulation. According to Cochrane Handbook guidelines [57], executing a statistical pooling under such structural and clinical incompatibility compromises the conceptual validity of the global estimator, reducing the evaluation to an arbitrary aggregation. To ensure maximum analytical rigor in line with the PRISMA 2020 declaration [10], this study formalizes a qualitative narrative synthesis within the Synthesis Without Meta-analysis (SWiM) [58] framework, tabulating and grouping findings to preserve methodological rigor and clinical translatability.

From a clinical standpoint, this technological heterogeneity severely challenges evidence-based prescriptive frameworks in audiology. While traditional fitting formulae, such as *NAL-NL2* and *DSLv5*, rely on standardized, static acoustic targets to optimize audibility and comfort, deep learning models introduce highly non-linear, dynamic transformations. The 5% to 15% objective improvements observed in metrics like STOI often come at the expense of computational artifacts that distort speech harmonics, resulting in a perceived "robotic" texture. Therefore, the clinical success of smart HAs hinges upon transitioning from purely mathematical optimization to hardware-aware, bio-inspired architectures that balance objective intelligibility gains with human-centric parameters to guarantee long-term user adherence in real-world environments.

5 CONCLUSION

This systematic review was conducted to synthesize the current impact of Deep Learning on hearing aid technology and identify the primary barriers to its practical integration. AI-driven architectures consistently outperform conventional DSP algorithms in speech intelligibility, with objective improvements ranging from 5% to 15%. According to the GRADE framework, there is a moderate certainty of evidence supporting these technical advancements. However, a significant gap remains regarding the clinical and subjective impact of these models. The certainty of evidence for sound quality is low, and it drops to very low for user preferences and satisfaction. This lack of subjective certainty is further highlighted by the fact that only 38% of the analyzed studies incorporated subjective metrics to validate their results.

6 ACKNOWLEDGMENTS

This study was supported by the Secretaría de Ciencias, Humanidades, Tecnologías e Innovación (Secihti), formerly CONAHCYT, through the doctoral scholarship granted to the first author (CVU: 316884). The authors also express their gratitude to the Posgrado en Ingeniería para la Innovación Tecnológica (PIIT) at the Universidad Autónoma de Zacatecas (UAZ) for the academic guidance and institutional support provided throughout this study.

6.1 Data availability statement

The data supporting the findings of this systematic review are derived from public domain sources and are included within the article (and/or its references). Further inquiries can be directed to the corresponding author.

7 REFERENCES

[1] WHO, "Deafness and hearing loss," 2026. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[2] S. Akbarzadeh, E. Lobarinas, and N. Kehtarnavaz, "Online personalization of compression in hearing aids via maximum likelihood inverse reinforcement learning," *IEEE Access*, vol. 10, pp. 58537–58546, 2022. https://doi.org/10.1109/ACCESS.2022.3178594

[3] L. W. Balling, L. L. Mølgaard, O. Townend, and J. B. B. Nielsen, "The collaboration between hearing aid users and artificial intelligence to optimize sound," *Semin. Hear.*, vol. 42, no. 3, pp. 282–294, 2021. https://doi.org/10.1055/s-0041-1735135

[4] Z. Tu, N. Ma, and J. Barker, "DHASP: Differentiable hearing aid speech processing," in *ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 296–300. https://doi.org/10.1109/ICASSP39728.2021.9414571

[5] S. Ahmed *et al.*, "NeuroAMP: A novel end-to-end general purpose deep neural amplifier for personalized hearing aids," *IEEE Transactions on Artificial Intelligence*, vol. 7, no. 3, pp. 1610–1625, 2025. https://doi.org/10.1109/TAI.2025.3603538

[6] F. Drakopoulos and S. Verhulst, "A neural-network framework for the design of individualised hearing-loss compensation," *IEEE/ACM Transactions on Audio, Speech and Lang. Proc.*, vol. 31, pp. 2395–2409, 2023. https://doi.org/10.1109/TASLP.2023.3282093

[7] Y. Liu, "Design and optimization of human-computer interaction system for education management based on artificial intelligence," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 7, pp. 107–124, 2024. https://doi.org/10.3991/ijim.v18i07.48337

[8] P. U. Diehl *et al.*, "Restoring speech intelligibility for hearing aid users with deep learning," *Scientific Reports*, vol. 13, p. 2719, 2023. https://doi.org/10.1038/s41598-023-29871-8

[9] U. Anwar, X. Ni, A. Adetomi, K. Dashtipour, M. Gogate, and T. Arslan, "Cognitive load driven audio-visual speech enhancement using a neuro-fuzzy inference model and openmha platform," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 3267–3280, 2025. https://doi.org/10.1109/TASLPRO.2025.3594296

[10] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021. https://doi.org/10.1136/bmj.n71

[11] M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews", *BMJ*, vol. 372, p. n160, 2021. https://doi.org/10.1136/bmj.n160

[12] J. A. Johnson, J. Xu, and R. M. Cox, "Impact of hearing aid technology on outcomes in daily life II: Speech understanding and listening effort," *Ear and Hearing*, vol. 37, no. 5, pp. 529–540, 2016. https://doi.org/10.1097/AUD.0000000000000327

[13] J. M. Kates and K. H. Arehart, "The hearing-aid speech perception index (HASPI)," *Speech Communication*, vol. 65, pp. 75–93, 2014. https://doi.org/10.1016/j.specom.2014.06.002

[14] R. Sutherland, G. Close, T. Hain, S. Goetze, and J. Barker, "Using speech foundational models in loss functions for hearing aid speech enhancement," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, Lyon, France, 2024, pp. 421–425. https://doi.org/10.23919/EUSIPCO63174.2024.10714933

[15] Z. Genc *et al.*, "Analysis of documents published on mobile technology of hearing impaired in web of science database," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 23, pp. 169–181, 2020. https://doi.org/10.3991/ijet.v15i23.18829

[16] A. Shukla *et al.*, "Hearing loss, loneliness, and social isolation: A systematic review," *Otolaryngol–Head Neck Surg.*, vol. 162, no. 5, pp. 622–633, 2020. https://doi.org/10.1177/0194599820910377

[17] B. J. Lawrence, D. M. Jayakody, R. J. Bennett, R. H. Eikelboom, N. Gasson, and P. L. Friedland, "Hearing loss and depression in older adults: A systematic review and meta-analysis," *The Gerontologist*, vol. 60, no. 3, pp. e137–e154, 2020. https://doi.org/10.1093/geront/gnz009

[18] M. A. Ferguson, P. T. Kitterick, L. Y. Chong, M. Edmondson-Jones, F. Barker, and D. J. Hoare, "Hearing aids for mild to moderate hearing loss in adults," *Cochrane Database of Systematic Reviews*, no. 9, 2017. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012023.pub2/abstract

[19] M. S. K. Lakshmi, A. Rout, and C. R. O'Donoghue, "A systematic review and meta-analysis of digital noise reduction hearing aids in adults," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 2, pp. 120–129, 2021. https://doi.org/10.1080/17483107.2019.1642394

[20] G. H. Guyatt *et al.*, "GRADE: An emerging consensus on rating quality of evidence and strength of recommendations," *BMJ*, vol. 336, no. 7650, pp. 924–926, 2008. https://doi.org/10.1136/bmj.39489.470347.AD

[21] E. Healy, M. Delfarah, E. Johnson, and D. Wang, "A deep learning algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker and reverberation," *The Journal of the Acoustical Society of America*, vol. 145, no. 3, p. 1378, 2019. https://doi.org/10.1121/1.5093547

[22] Y.-H. Lai, W.-Z. Zheng, S.-T. Tang, S.-H. Fang, W.-H. Liao, and Y. Tsao, "Improving the performance of hearing aids in noisy environments based on deep learning technology," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 404–408. https://doi.org/10.1109/EMBC.2018.8512277

[23] N. Mamun and J. H. L. Hansen, "Speech enhancement for cochlear implant recipients using deep complex convolution transformer with frequency transformation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2616–2629, 2024. https://doi.org/10.1109/TASLP.2024.3366760

[24] M. Keshavarzi, T. Reichenbach, and B. C. J. Moore, "Transient noise reduction using a deep recurrent neural network: Effects on subjective speech intelligibility and listening comfort," *Trends in Hearing*, vol. 25, p. 23312165211041475, 2021. https://doi.org/10.1177/23312165211041475

[25] A. Ni, E. Lobarinas, and N. Kehtarnavaz, "Personalization of hearing AID DSLV5 prescription amplification in the field via a real-time smartphone APP," in *2023 24th International Conference on Digital Signal Processing (DSP)*, 2023, pp. 1–5. https://doi.org/10.1109/DSP58604.2023.10167984

[26] A. Ni, E. Lobarinas, and N. Kehtarnavaz, "Efficient personalization of amplification in hearing aids via multi-band bayesian machine learning," *IEEE Access*, vol. 12, pp. 112116–112123, 2024. https://doi.org/10.1109/ACCESS.2024.3441762

[27] L. Sahoo, U. Patnaik, N. Singh, G. Dwivedi, G. Nagre, and K. S. Sahoo, "Comparing audiological outcomes of conventional and AI-upgraded cochlear implant speech processors," *Indian Journal of Otolaryngology and Head and Neck Surgery: Official Publication of the Association of Otolaryngologists of India*, vol. 76, pp. 4356–4364, 2024. https://doi.org/10.1007/s12070-024-04860-z

[28] P. Gonzalez, Z.-H. Tan, J. Østergaard, J. Jensen, T. S. Alstrøm, and T. May, "Investigating the design space of diffusion models for speech enhancement," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 32, pp. 4486–4500, 2024. https://doi.org/10.1109/TASLP.2024.3473319

[29] R. Hinrichs and J. Ostermann, "Pruning-aware loss functions for STOI-optimized pruned recurrent autoencoders for the compression of the stimulation patterns of cochlear implants at zero delay," in *2024 58th Asilomar Conference on Signals, Systems, and Computers*, 2024, pp. 1427–1432. https://doi.org/10.1109/IEEECONF60004.2024.10943066

[30] C. Jehn, A. Kossmann, N. K. Vavatzanidis, A. Hahne, and T. Reichenbach, "CNNs improve decoding of selective attention to speech in cochlear implant users," *Journal of Neural Engineering*, vol. 22, no. 3, 2025. https://doi.org/10.1088/1741-2552/addb7b

[31] N. Y.-H. Wang *et al.*, "Improving the intelligibility of speech for simulated electric and acoustic stimulation using fully convolutional neural networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 184–195, 2021. https://doi.org/10.1109/TNSRE.2020.3042655

[32] J. O'Sullivan, Z. Chen, S. A. Sheth, G. McKhann, A. D. Mehta, and N. Mesgarani, "Neural decoding of attentional selection in multi-speaker environments without access to separated sources," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 1644–1647. https://doi.org/10.1109/EMBC.2017.8037155

[33] S. Rajkumar, S. Muttan, V. Sapthagirivasan, V. Jaya, and S. S. Vignesh, "Development of improved software intelligent system for audiological solutions," *J. Med. Syst.*, vol. 42, no. 7, pp. 1–12, 2018. https://doi.org/10.1007/s10916-018-0978-6

[34] A. M. Raghavan, N. Lipschitz, J. T. Breen, R. N. Samy, and G. D. Kohlberg, "Visual speech recognition: Improving speech perception in noise through artificial intelligence," *Otolaryngol. Head Neck Surg.*, vol. 163, no. 4, pp. 771–777, 2020. https://doi.org/10.1177/0194599820924331

[35] A. A. Saoji *et al.*, "How does deep neural network-based noise reduction in hearing aids impact cochlear implant candidacy?," *Audiology Research*, vol. 14, no. 6, pp. 1114–1125, 2024. https://doi.org/10.3390/audiolres14060092

[36] N. Westhausen, H. Kayser, T. Jansen, and B. Meyer, "Real-time multichannel deep speech enhancement in hearing aids: Comparing monaural and binaural processing in complex acoustic scenarios," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 4596–4606, 2024. https://doi.org/10.1109/TASLP.2024.3473315

[37] B. Essaid, H. Kheddar, N. Batel, and M. Chowdhury, "Deep learning-based coding strategy for improved cochlear implant speech perception in noisy environments," *IEEE Access*, vol. 13, pp. 35707–35732, 2025. https://doi.org/10.1109/ACCESS.2025.3542953

[38] T. Gajecki, Y. Zhang, and W. Nogueira, "A deep denoising sound coding strategy for cochlear implants," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 9, pp. 2700–2709, 2023. https://doi.org/10.1109/TBME.2023.3262677

[39] F. Hajiaghababa, H. R. Marateb, and S. Kermani, "The design and validation of a hybrid digital-signal-processing plug-in for traditional cochlear implant speech processors," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 103–109, 2018. https://doi.org/10.1016/j.cmpb.2018.03.003

[40] Md. I. Hossain, Md. S. Islam, Mst. T. Khatun, R. Ullah, A. Masood, and Z. Ye, "Dual-transform source separation using sparse nonnegative matrix factorization," *Circuits Syst. Signal Process.*, vol. 40, no. 4, pp. 1868–1891, 2021. https://doi.org/10.1007/s00034-020-01564-x

[41] L. Bramsløw, G. Naithani, A. Hafez, T. Barker, N. H. Pontoppidan, and T. Virtanen, "Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm," *The Journal of the Acoustical Society of America*, vol. 144, no. 1, p. 172, 2018. https://doi.org/10.1121/1.5045322

[42] S. Klein, L. Rossol, F. Venema, S. Schönewald, J. Karrenbauer, and H. Blume, "Noise reduction in hearing-aid processors: Traditional methods vs. Neural networks," in *2025 IEEE 36th International Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, 2025, pp. 172–173. https://doi.org/10.1109/ASAP65064.2025.00037

[43] T. Goehring, X. Yang, J. J. M. Monaghan, and S. Bleeck, "Speech enhancement for hearing-impaired listeners using deep neural networks with auditory-model based features," in *2016 24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 2300–2304. https://doi.org/10.1109/EUSIPCO.2016.7760659

[44] H. Schröter, T. Rosenkranz, A.-N. Escalante-B., and A. Maier, "Low latency speech enhancement for hearing aids using deep filtering," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 30, pp. 2716–2728, 2022. https://doi.org/10.1109/TASLP.2022.3198548

[45] J. Wathour, P. J. Govaerts, and N. Deggouj, "From manual to artificial intelligence fitting: Two cochlear implant case studies," *Cochlear Implants International*, vol. 21, no. 5, pp. 299–305, 2020. https://doi.org/10.1080/14670100.2019.1667574

[46] P. A. L. Bysted, "Hearing loss compensation using computational models of hearing and deep learning," 2024. https://vbn.aau.dk/files/770648921/PHD_PALB_ONLINE.pdf

[47] P. Leer, J. Jensen, Z.-H. Tan, J. Østergaard, and L. Bramsløw, "How to train your ears: Auditory-model emulation for large-dynamic-range inputs and mild-to-severe hearing losses," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 32, pp. 2006–2020, 2024. https://doi.org/10.1109/TASLP.2024.3378099

[48] I. López-Espejo, A. Edraki, W. Chan, Z. Tan, and J. Jensen, "On the deficiency of intelligibility metrics as proxies for subjective intelligibility," *Speech Commun.*, vol. 150, pp. 9–22, 2023. https://doi.org/10.1016/j.specom.2023.04.001

[49] T. J. Cox, J. Barker, W. Bailey, S. Graetzer, M. A. Akeroyd, and J. F. Culling, "Overview of the 2023 ICASSP SP Clarity challenge: Speech enhancement for hearing aids," in *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2. https://doi.org/10.1109/ICASSP49357.2023.10433922

[50] C. Wen, G. Torfs, and S. Verhulst, "Artifact-free sound quality in DNN-based closed-loop systems for audio processing," *arXiv preprint arXiv:2501.04116*, 2025.

[51] S. Drgas, L. Bramsløw, A. Politis, G. Naithani, and T. Virtanen, "Dynamic processing neural network architecture for hearing loss compensation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 203–214, 2024. https://doi.org/10.1109/TASLP.2023.3328285

[52] M. Meeuws, D. Pascoal, I. Bermejo, M. Artaso, G. De Ceulaer, and P. J. Govaerts, "Computer-assisted CI fitting: Is the learning capacity of the intelligent agent FOX beneficial for speech understanding?" *Cochlear Implants International*, vol. 18, no. 4, pp. 198–206, 2017. https://doi.org/10.1080/14670100.2017.1325093

[53] P. C. Groot, T. Heskes, T. M. H. Dijkstra, and J. M. Kates, "Predicting preference judgments of individual normal and hearing-impaired listeners with Gaussian Processes," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 811–821, 2011. https://doi.org/10.1109/TASL.2010.2064311

[54] Z. Sun, Y. Li, H. Jiang, F. Chen, X. Xie, and Z. Wang, "A supervised speech enhancement method for smartphone-based binaural hearing aids," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 5, pp. 951–960, 2020. https://doi.org/10.1109/TBCAS.2020.2988121

[55] E. Ceolini *et al.*, "Brain-informed speech separation (BISS) for enhancement of target speaker in multitalker speech perception," *NeuroImage*, vol. 223, p. 117282, 2020. https://doi.org/10.1016/j.neuroimage.2020.117282

[56] V. Choudhari *et al.*, "Brain-controlled augmented hearing for spatially moving conversations in multi-talker environments," *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, vol. 11, no. 41, p. e2401379, 2024. https://doi.org/10.1002/advs.202401379

[57] J. Higgins and J. Thomas, Eds., *Cochrane Handbook for Systematic Reviews of Interventions (Current Version)*. Cochrane, 2024. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current

[58] M. Campbell *et al.*, "Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline," *BMJ*, vol. 368, p. l6890, 2020. https://doi.org/10.1136/bmj.l6890

8 AUTHORS

Jaime Raymundo Martínez Solís is with the Universidad Autónoma de Zacatecas, Zacatecas, Zac., México. He holds a Master of Science in Computer Science and a Master's degree in Software Engineering. His research interests center on software development and artificial intelligence applications applied to computational audiology, digital signal processing (DSP), and speech intelligibility optimization within hearing assistance systems (E-mail: mtzsol@uaz.edu.mx).

Juvenal Villanueva Maldonado is with the Universidad Autónoma de Zacatecas, Zacatecas, Zac., México. He is a Mechanical and Electrical Engineer from UNAM, and holds a Master's and Ph.D. in Engineering with a specialization in Control from the Institute of Engineering of the same university. He has carried out teaching, research, and technology transfer activities in various academic institutions and technology-based companies. He served as a CONACYT Research Fellow at CIDTE of the Autonomous University of Zacatecas, where he participated in technological innovation projects for both the public and private sectors. He is currently a Research Professor at UAZ, responsible for the Automotive Electronic Engineering and Intelligent Systems and Video Games programs, and a member of the National System of Researchers (SNI), Level I. His research interests include mobile robotics, control and automation, intelligent systems, Internet of Things, and dynamic systems modeling (E-mail: juvenal.villanueva@uaz.edu.mx).

Carlos Eric Galván Tejada earned his Bachelor of Science in Computer Science and Master of Engineering degrees from the Autonomous University of Zacatecas. He holds a PhD in Information and Communication Technologies with a specialization in Ambient Intelligence from the Monterrey Institute of Technology and Higher Education (ITESM), Monterrey Campus. His research interests include Biomedical Engineering, Artificial Intelligence applied to Medicine, Biomedical Signal Processing, Health Data Mining, and the Generation of Synthetic Data for clinical analysis. He also develops science outreach projects for children, focusing on medicine and technology. He is currently a member of the National System of Researchers (SNI Level 3) in Area III: Medicine and Health Sciences (E-mail: ericgalvan@uaz.edu.mx).

Gloria Viviana Cerrillo Rojas completed her postgraduate studies at the Escuela Nacional de Ciencias Biológicas (ENCB) and earned her Ph.D. from the Universidad Autónoma de Aguascalientes (UAA). She is a professor and researcher at the Academic Unit of Chemical Sciences of the Universidad Autónoma de Zacatecas (UAZ). She has authored numerous scientific articles and book chapters. Her research interests focus on Bioinformatics in environment and health (E-mail: anaroce267@gmail.com).