



Impact of machine learning and deep learning algorithms on QoS, QoE, and security optimization in the VOIP ecosystem: a systematic review

Impacto de los algoritmos de machine learning y deep learning en la optimización de la QoS, QoE y seguridad en el ecosistema VOIP: una revisión sistemática

Miguel Angel Rosero Enrriquez¹ 

¹ Universidad de Nariño. Nariño Pasto, Colombia.

Submitted: 26-06-2026 | Revised: 30-08-2026 | Accepted: 04-10-2026 | Published: 05-10-2026

ABSTRACT

Advances in machine learning and deep learning algorithms have driven the evolution of technological infrastructures, transforming Voice over IP (VoIP) through significant improvements in quality of service (QoS) and quality of experience (QoE). Currently, these implementations enable the prediction of spam calls, the prevention of packet loss in real time, and the optimization of uninterrupted communications. This article presents a systematic literature review aimed at synthesizing these findings after consulting five specialized engineering databases and initially identifying more than 60 articles published in the last five years. Following a filtering process, 25 studies were selected that highlight the effectiveness of models such as CNNs, LSTMs, and SDN networks for the proactive management of massive traffic flows, achieving accuracies exceeding 97 %. Finally, the role of Explainable AI (XAI) as a pillar of transparency is analyzed, concluding that network autonomy improves resilience and end-user satisfaction.

Keywords: VoIP; Machine Learning; Deep Learning; Quality of Service (QoS); Quality of Experience (QoE); Systematic Literature Review.

RESUMEN

Los avances en los algoritmos de machine learning y deep learning sin duda han impulsado la evolución de las infraestructuras tecnológicas que nos rodean. Uno de los campos que se ha visto trastocado por estas mejoras es la transmisión de voz sobre IP (VoIP), más específicamente se han visto optimizados procesos que tiene que ver con la calidad del servicio (QoS) y la experiencia del usuario (QoE). Actualmente, es posible predecir llamadas de spam, prevenir la pérdida de paquetes en tiempo real y optimizar las comunicaciones sin interrupciones. El presente artículo expone una revisión sistemática de la literatura con el objetivo de sintetizar estos hallazgos tras consultar cinco bases de datos especializadas en ingeniería e identificar inicialmente más de 60 artículos publicados en los últimos cinco años. Tras un proceso de filtrado, se seleccionaron 25 estudios que destacan la eficacia de modelos como CNN, LSTM y redes SDN para la gestión proactiva de flujos masivos, alcanzando precisiones hasta ahora. Finalmente, se analiza el rol de la IA Explicable (XAI) como pilar de transparencia, concluyendo que la autonomía de red mejora la resiliencia y la satisfacción del usuario final.

Palabras clave: VoIP; Machine Learning; Deep Learning; Calidad de Servicio (QoS); Calidad de Experiencia (QoE); Revisión Sistemática de la Literatura.

INTRODUCTION

The widespread adoption of VoIP technology has transformed global communication owing to its cost-effectiveness; however, it has also increased exposure to cyberthreats such as vishing and data theft [1]. Signaling protocols such as SIP exhibit structural vulnerabilities that allow denial-of-service and billing fraud to be perpetrated [2,4]. Similarly, Spam over Internet Telephony (SPIT) degrades user experience and consumes network resources, surpassing traditional static defenses [3].

Given the inadequacy of the countermeasures known to date, this work focuses on the transition toward autonomous defense systems. The use of intelligent algorithms enables the prediction of attack signatures by learning from previous legitimate dialogues[2].

The technical analysis demonstrates that models such as CNN and LSTM achieve accuracies above 97% in managing massive flows[3,5]. Deep Reinforcement Learning (DRL) frameworks have mitigated data theft attacks with detection latencies of only 2.1 seconds[1]. Furthermore, algorithms such as XGBoost show nearly 100% effectiveness in SDN networks, while Explainable AI (XAI) via SHAP consolidates administrators' trust in the system[5,6].

This document is organized as follows: Section 2 details the methodological search and selection process; Section 3 presents the analytical results organized by research questions; and finally, the conclusions derived from the study are presented.

METHODS

For the systematic literature review presented in this article, a Software Engineering research process approach was used[24]. The main purpose was to provide an overview of the research area and to identify the quantity, type of research, and available results regarding the impact of machine learning and deep learning on the VoIP ecosystem, so as to build a classification scheme and a valid structure.

Figure 1 illustrates the various stages that constitute the method adopted for this systematic review. These phases enable an understanding of the process followed from the formulation of the research questions to the collection, evaluation, and synthesis of information, thus ensuring a rigorous and organized approach to the analysis of the literature.

Research questions

This study presents the findings of a systematic review on the use of machine learning (ML) and deep learning (DL) for optimizing user experience in VoIP calls. In order to structure the state of the art, the following questions were posed:

- RQ1: How is machine learning transforming defense strategies against fraud and security threats in IP telephony services?
- RQ2: What are the most effective artificial intelligence models for predicting user satisfaction (QoE) based on VoIP network behavior?
- RQ3: How does AI enable the identification and prioritization of VoIP traffic when it is concealed by encryption and anonymity protocols?
- RQ4: How does the integration of artificial intelligence impact massive traffic management to ensure the quality of voice services in SDN networks?
- RQ5: How do reinforcement learning and network autonomy contribute to routing optimization and resilience in VoIP systems?
- RQ6: What is the role of explainable AI (XAI) in creating VoIP network management systems that are transparent and trustworthy for administrators?

These questions were structured using the PICO model with its PIPOH variant[25], which allows delimiting the essential components of the research to ensure that the findings have technical and professional relevance. Table 1 details these concepts as applied to the study of VoIP network optimization related to machine learning (ML) and deep learning (DL).

Search

The search strategy consisted of exploring technical terms in specialized databases (IEEE Xplorer, EBSCOhost, DOAJ, ResearchRabbit, OpenKnowledgeMaps). Boolean operators were applied to combine concepts related to IP telephony, learning algorithms, and quality of service and quality of experience. Table 2 describes the terms with their respective synonyms and filters.

Table 1. Definition of general concepts using PIPOH.

Criterion	Description
Population including signaling protocols (SIP) and voice transport (RTP).	Network infrastructures and IP telephony (VoIP) services,
Intervention	Implementation of machine learning algorithms, deep learning, SDN networks, and reinforcement learning models.
Professionals	Systems engineers, network administrators, telecommunications solution developers, and researchers.
Results	Quality of Service (QoS) optimization, improvement of search

Source: own elaboration.

Table 3 presents the main topic and the databases that were investigated; in addition, Table 4 includes the string or token constructed from the main topic and synonyms. This token was the fundamental basis for the search and compilation of all the articles that constitute this systematic review.

Selection

All studies identified through the search string were subjected to an exhaustive filtering process. Inclusion and exclusion criteria were defined to ensure that the collected evidence was relevant to the impact of ML/DL on the VoIP ecosystem.

Inclusion Criterion (IC): Engineering-focused document that presents the application of artificial intelligence algorithms (ML, DL) for the optimization of security, QoS, or QoE in VoIP networks, published between 2021 and 2026.

It should be noted that, although the chronological criterion prioritizes publications between 2021 and 2026, 8 previous studies (2016–2020) were exceptionally included. These sources were classified as methodological literature, as they were indispensable for defining the software engineering framework[24]and the clinical question structure[25]that govern this review.

Table 2. Terms, synonyms, and filters used to compose the search string.

Main Term	Dimensions / Synonyms (Keywords)	Additional Filters
Machine learning, deep learning, artificial intelligence, neural networks, reinforcement learning.	AI-based optimization of VoIP ecosystems, Voice over IP, IP telephony, QoS, QoE in VoIP, SIP signaling, RTP streams. Engineering, telecommunications, computer science, educational technology. 2021–2026.	Engineering, telecommunication, computer science, educational technology. 2021–2026. Journals, systematic reviews.

Source: authors' own elaboration.

Table 3. Main term and databases.

Main Term	Information Sources (Databases)
AI-based optimization of ecosystems	IEEE Xplore: Main repository for technical articles and engineering conferences on QoS and QoE in VoIP. Open Knowledge Maps: A visual discovery tool for identifying thematic clusters. ResearchRabbit: A citation mapping platform for identifying seminal and recent literature. EBSCOhost: Multidisciplinary database for peer-reviewed articles (Peer-reviewed). DOAJ (Directory of Open Access Journals): Directory of high-quality open access journals.

Source: authors' own elaboration.

Table 4. Search string.

Main Term	Search String
AI-Based Optimization of QoS, QoE in VoIP Ecosystems	("VoIP" OR "Voice over IP" OR "IP Telephony") AND ("Machine Learning" OR "Deep Learning" OR ("VoIP" OR "Voice over IP" OR "IP Telephony") AND ("Machine Learning" OR "Deep Learning" OR

Source: authors' own elaboration.

Exclusion criteria (EC): Documents not available for download, or research focused on traditional analog telephony without machine learning components.

The selection was conducted in 3 phases:

- Phase 1: Elimination of duplicate articles among the 5 databases.
- Phase 2: Discarding non-downloadable articles.
- Phase 3: Application of the inclusion/exclusion criteria by reading titles and abstracts.

It was necessary to consult the articles on different dates, as described in Table 5.

Quality assessment

To ensure the validity of the 25 final documents, 7 quality criteria were applied based on source provenance, relevance of the content regarding VoIP, and the use of machine learning and deep learning. These criteria made it possible to review the titles and abstracts with technical rigor, ensuring that only related research was included.

Table 5. Dates of consultation and file downloads.

Information Sources	Search Date
IEEE Xplore	18 February 2026
EBSCOhost	18 February 2026
DOAJ	18 February 2026
Open Knowledge Maps	25 February 2026
ResearchRabbit	25 February 2026

Source: authors' own elaboration.

Data extraction

After the execution of the search strategy, the results were filtered to ensure the quality of the evidence collected. This selection process allowed us to go from 60 initial records to a select sample of 25 relevant studies. The results of this discard flow and the eligibility criteria applied are summarized in Table 6.

Table 6. Quality assessment of the search and selection process.

Source of Information	Search Result (Initial)	Duplicate Files	Excluded Files	Relevant Studies
IEEE Xplore	32	0	18	14
ResearchRabbit	10	4	2	4
EBSCOhost	8	3	2	3
Open Knowledge Maps	7	3	1	3
DOAJ	3	1	1	1
Total	60	11	24	25

Source: Authors' own elaboration.

Table 7. Studies included in the systematic review.

Year	Number of Articles	Referencias
2016	1	[25]
2017	1	[23]
2018	1	[21]
2019	2	[18][24]
2020	2	[15][22]
2021	6	[4][5][9][13][20][12]
2022	2	[2][19]
2023	3	[14][16][17]
2024	3	[3][6][10]
2025	2	[7][11]
2026	2	[1][8]
Total	25	

Source: Authors' own elaboration.

RESULTS

Transformation of defense strategies against fraud and security threats in IP telephony using machine learning

Regarding proactive defense, the implementation of techniques such as Deep Reinforcement Learning (DRL) enables systems to learn evolving attack patterns. This approach facilitates the mitigation of vishing and data theft in real time through constant monitoring of transitions between periods of voice activity and silence[1].

To combat Telephone Spam (SPIT), nonlinear models using Artificial Neural Networks (ANN) have been developed, which analyze Call Detail Records (CDR) to differentiate between legitimate users and spammers, thereby achieving accuracies above 97% by modeling complex associations among input parameters[2]. Regarding Network and User Behavior Analysis, trust evaluation has been strengthened through the use of social graphs via systems such as ROBO-SPOT, which analyze caller reputation based on the reciprocity and duration of interactions to achieve success rates of 97% in detecting malicious calls[3].

Likewise, the Adaptive Prediction of SIP Signaling allows anticipating attacks through the use of Hidden Markov Models (HMM) and Bayesian inference to detect anomalous message sequences and block malicious signatures even before the call is formally established[4]. Regarding Security in Encrypted Environments and Optimization, among the greatest challenges is visibility in VPN or TOR channels. Deep Learning, through CNN and LSTM models, enables classifying threats by analyzing spatio-temporal characteristics of the flow without the need to decrypt the content[5]. To ensure that these processes do not affect quality of service (QoS), techniques such as State Transition Limitation (LNST) are applied, which reduces significantly, the computational load[6].

Additionally, the use of Bootstrap Forest Models and multilayer classification systems has proven to be effective for identifying voice services within HTTPS channels with an F1-Score of 0.90, thus allowing for dispensing with deep packet inspection (DPI) or the use of specific IP addresses and ports for detection[7,9].

Identification and Prioritization of VoIP Traffic in Encryption and Anonymity Environments Using AI

Regarding Identification in Encrypted and Anonymous Environments, visibility in complex channels such as VPN or TOR represents one of the greatest challenges; however, the use of Deep Learning models such as CNN and LSTM allows threats and traffic types to be classified by analyzing the spatio-temporal characteristics of the flow (FSTFs) without the need to decrypt the content[5,9].

This proactive approach allows systems such as the MA-PLE identification engine, based on the learning of hidden representations, to detect VoIP data streams (RTPs) with greater than 99% accuracy from the payload structure, allowing their detection even before signaling information arrives[16].

Likewise, Bootstrap Forest models have proven highly effective for identifying voice services within encrypted HTTPS channels, with an F1-Score of 0.90, which allows administrators to forgo deep packet inspection (DPI) or dependence on specific ports and IP addresses that are usually masked[7,9]. For prioritization and dynamic management, frameworks such as RNN-BLSTM are used, which predict traffic trends for resource allocation[14].

Impact of Artificial Intelligence on Massive Traffic Management to Ensure Quality of Service in SDN Networks

In software-defined networks (SDN), the integration of AI enables proactive management by identifying so-called “Elephant Flows”

(large, long-duration flows), which are the main causes of congestion and quality of service degradation[17].

Deep learning algorithms such as XGBoost, GBM, and GDF implemented in the SDN controller achieve validation accuracies of up to 100 % for predicting these massive flows. Similarly, this capability enables the automatic reservation of optimal routes based on the load of each link in real time[17,18]. This dynamic prediction capability ensures that latency- and bandwidth-sensitive applications, such as VoIP and videoconferencing, maintain guaranteed QoS even under critical fluctuations in massive traffic[17,22].

Contribution of Reinforcement Learning and Network Autonomy to Routing Optimization and Resilience in VoIP Systems

Reinforcement Learning (RL) allows network systems to continuously learn from their event history, surpassing conventional protocols (such as OLSR or BATMAN) that cannot avoid routes that have exhibited recurring failures or overloads in the past[19]. Improved protocols based on this technique, such as e-RLRP, contribute to resilience through reward mechanisms that optimize the flow path and substantially reduce packet loss and delay[19,20]. Furthermore, network autonomy facilitated by RL enables dynamic adjustments, such as modifying the “Hello” message interval, to compensate for the overhead generated by intelligent management itself. Moreover, the implementation of these models achieves improvements greater than 90% in the voice quality perceived by the end user[19,21].

The Role of Explainable AI (XAI) in the Creation of Transparent and Trustworthy VoIP Network Management Systems

The fundamental role of Explainable AI (XAI) is to mitigate the “black box” nature of complex deep learning models, making management decisions interpretable and reliable for human operators[17].

Tools such as SHAP (Shapley Additive Explanations) make it possible to identify and quantify which specific attributes (such as packet or byte size) had the greatest influence on congestion prediction or anomaly detection. Consequently, this allows administrators to validate the system's diagnoses before applying large-scale routing policies[17,22]. Integrating XAI fosters a symbiosis between algorithmic efficiency and human expert judgment, resulting in more transparent and secure VoIP management systems aligned with the organization's operational intentions[23].

CONCLUSIONS

With the advancement of the Internet, VoIP infrastructure has established itself as a primary solution for worldwide connectivity owing to its scalability and ability to integrate with modern networks. Nevertheless, this transition requires a higher level of intelligence in network management than conventional methods can provide; therefore, it is essential that experts in the field leverage advanced machine learning and deep learning tools to ensure that the user and environmental experience in communication is optimal and capable of addressing medium-complexity problems without constant human supervision.

In this work, a systematic literature review was conducted (see Section 2), following a methodology focused on research processes in software engineering.

Through this methodology, various studies on the use of machine learning and deep learning algorithms to improve service quality in VoIP calls were compiled, organized, and analyzed. Studies from 5 specialized databases (IEEE Xplore, EBSCOhost, DOAJ, ResearchRabbit, and OpenKnowledgeMaps) were identified, all published within the last 5 years. After applying quality criteria and removing duplicates, 17 were discarded, leaving 25 relevant studies that served to answer the 6 research questions posed.

Section 3 presents the 10 main findings of the review. Regarding RQ1, the review highlights that machine learning is transforming how fraud is combated. Models such as deep reinforcement learning (DRL) and artificial neural networks (ANN) make it possible to move from reactive defenses to more proactive strategies, achieving detection rates that were previously unusual or unthinkable in manual human work. Regarding user satisfaction (RQ2), the results show that combined models such as QoEPredict (based on XGBoost and Random Forest), together with fuzzy logic approaches (F-QoS), achieve high accuracies with extremely low execution times. This positions them as more effective alternatives to traditional statistical methods for predicting quality of experience (QoE).

On the other hand, continuing in numerical order, regarding RQ3 and RQ4, the review notes the importance of CNN and LSTM for detecting VoIP traffic even when it is concealed through encryption, including VPN or Tor, without the need to decrypt the content. Furthermore, in software-defined networking (SDN) environments, algorithms such as XGBoost and GBM enable the efficient management of large volumes of traffic, identifying high-impact flows ("elephant flows") with high levels of accuracy.

On the other hand, regarding autonomy and resilience (RQ5), reinforcement learning-based protocols, such as e-RLRP, are observed to outperform traditional approaches by dynamically adapting routes according to the network history.

Finally, regarding RQ6, explainable artificial intelligence is crucial for building trust in these systems. It enables administrators to understand how and why AI makes certain decisions, reducing reliance on models that do not fully explain their reasoning process.

REFERENCES

1. S. M. Ekolama et al., "A Deep-Reinforcement Learning Approach to Miti-gate Vishing Attacks and Data Theft in VoIP Calls," *J. High-Freq. Comm. Tech.*, vol. 4, no. 1, pp. 416–428, 2026.
2. L. Behan et al., "Efficient Detection of Spam Over Internet Telephony by Machine Learning Algorithms," *IEEE Access*, vol. 10, 2022.
3. M. A. Azad et al., "ROBO-SPOT: Detecting Robocalls by Understanding User Engagement and Connectivity Graph," *Big Data Mining and Anal.*, vol. 7, no. 2, 2024.
4. D. Pereira et al., "A Machine Learning Approach for Prediction of Signaling SIP Dialogs," *IEEE Access*, vol. 9, 2021.
5. F. U. Islam et al., "VoIP Traffic Detection in Tunneled and Anonymous Networks Using Deep Learning," *IEEE Access*, vol. 9, 2021.
6. M. A. Ramírez-Reyna et al., "Joint Connection-Level and Packet-Level Per-formance Evaluation of VoIP," *IEEE Access*, vol. 12, 2024.
7. S. Durairaju y A. K. Cherukuri, "Bootstrap Forest based method for Encry-pted Network Traffic Analysis," *J. Inf. Org. Sci.*, vol. 49, no. 2, 2025.
8. R. A. Rafi et al., "Multilingual Low-Latency Emergency VoIP System Using LLM," *IEEE Access*, vol. 13, 2025.
9. A. S. Khatouni et al., "Machine Learning Based Classification Accuracy of Encrypted Service Channels," *J. Netw. Syst. Manag.*, vol. 29, no. 8, 2021.
10. R. Sangeetha et al., "Sketching of interactive VoIP traffic with multivariate statistical learning," *J. Intell. Fuzzy Syst.*, vol. 46, 2024.
11. K. M. T. Osee et al., "A Hybrid Ensemble-Based QoE Prediction From QoS and User Satisfaction Data," *IEEE Access*, vol. 13, 2025.
12. D. K. Altmemi et al., "A New Approach Based on an Intelligent Method to Classify Quality of Service," *Informatica*, vol. 46, no. 9, 2022.
13. A. H. Mohseni et al., "Toward a comprehensive subjective evaluation of VoIP users' quality of experience (QoE)," *Multimedia Tools and Applications*, vol. 80, 2021.
14. W. A. Aziz et al., "Content-Aware Network Traffic Prediction Framework for QoS-Aware Dynamic Manage-ment," *IEEE Access*, vol. 11, 2023.
15. B. Schuetz y N. Aschenbruck, "SPQER: Speech Quality Evaluation Using Word Recognition for VoIP," *IEEE Open Journal*, vol. 1, 2020.
16. M. Kapoor et al., "Detecting VoIP Data Streams: Approaches Using Hidden Representation Learning," *AAAI*, vol. 37, no. 13, 2023.
17. G. Wassie et al., "Traffic prediction in SDN for explainable QoS using deep learning approach," *Scientific Reports*, vol. 13, no. 1, 2023.
18. J. Xie et al., "A survey of machine learning techniques applied to SDN," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, 2019.
19. P. Larrenie et al., "Low complexity adaptive machine learning approaches for latency prediction," *Proc. MLN*, 2022.
20. D. R. Militani et al., "Enhanced Routing Algorithm Based on Reinforcement ML: A Case of VoIP," *Sensors*, vol. 21, no. 2, 2021.
21. D. Dugaev et al., "Adaptive Reinforcement Learning-Based Routing Proto-col," *Proc. APEIE*, 2018.
22. E. Gelenbe et al., "Self-Aware Networks That Optimize Security, QoS, and Energy," *Proc. IEEE*, vol. 108, no. 7, 2020.
23. S. Lundberg y S. I. Lee, "A unified approach to interpreting model predic-tions," *Adv. Neural Inform. Pro-cess. Syst.*, 2017.
24. J. A. Jiménez-Toledo et al., "Consideraciones en los procesos de enseñanza-aprendizaje... revisión sistemá-tica," *TecnoLógicas*, vol. 22, 2019.
25. J. D. Martínez Díaz et al., "El diseño de preguntas clínicas en la práctica basada en la evidencia," *Enfermería Glob.*, vol. 15, 2016.

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Miguel Angel Rosero Enrriquez.

Writing–review and editing: Miguel Angel Rosero Enrriquez.