Digital Object Identifier: 10.24054/rcta.v1i47.4356

Systematic review on identifying anomalies in the consumption of massive web services using machine learning techniques

Revisión sistemática sobre identificación de anomalías en el consumo de servicios web masivos mediante técnicas de machine learning

Ing. Jesús Andrés Cruz Sanabria[1], **Ing. Joaquín Iván Barrera Lozada**[1],
MSc. Karla Yohana Sánchez Mojica[1]

[1] ***Universidad de La Salle,*** *Facultad de Ingeniería, Maestría en Inteligencia Artificial, Bogotá, Cundinamarca, Colombia.*

Correspondence: *{jcruz47, jbarrera17, kasanchez}@unisalle.edu.co*

Received: *September 20, 2025.* ***Accepted:*** *December 18, 2025.* ***Published:*** *February 09, 2026.*

How to cite: *J. A. Cruz Sanabria, J. I. Barrera Lozada y K. Y. Sánchez Mojica, "Systematic review on identifying anomalies in the consumption of massive web services using machine learning techniques", RCTA, vol. 1, n.º. 47, pp. 209-218, feb. 2026. Recovered from https://ojs.unipamplona.edu.co/index.php/rcta/article/view/4356*

Abstract: This paper provides a systematic literature review on the application of machine learning techniques for anomaly detection within web services and distributed systems. Our methodology involved structured queries across major academic databases, including IEEE Xplore, Scopus, ScienceDirect, and the ACM Digital Library, covering research published between 2021 and 2025. Following the application of rigorous inclusion and exclusion criteria, a final cohort of 50 relevant articles was selected for detailed analysis. These studies were categorized based on data types, learning paradigms, application domains, and evaluation metrics to pinpoint current trends, strengths, and inherent limitations in the field. Our findings highlight a clear shift toward hybrid models and deep learning architectures, alongside a growing emphasis on explainability and scalability in distributed environments.

Keywords: literature review, anomaly detection, machine learning, QoS, web services.

Resumen: Este artículo presenta una revisión sistemática de la literatura sobre el uso de técnicas de Machine Learning aplicadas a la identificación de anomalías en servicios web y sistemas distribuidos. El proceso de revisión se desarrolló a partir de búsquedas estructuradas en bases de datos académicas reconocidas, incluyendo IEEE Xplore, Scopus, ScienceDirect y ACM Digital Library, considerando publicaciones entre 2021 y 2025. Se aplicaron criterios explícitos de inclusión y exclusión, lo que permitió seleccionar un conjunto final de cincuenta artículos relevantes. Los estudios analizados se organizaron según tipo de dato, enfoque de aprendizaje, dominio de aplicación y métricas empleadas, con el fin de identificar tendencias, fortalezas y limitaciones del estado del arte. Los resultados evidencian una creciente adopción de modelos híbridos y arquitecturas profundas, así como un interés sostenido por la explicabilidad y la escalabilidad en entornos distribuidos.

Palabras clave: revisión sistemática, detección de anomalías, machine learning, QoS, servicios web.

1. INTRODUCTION

Today, digital infrastructure faces unprecedented pressure due to the massification of network access, which by 2025 has reached 74% of the world's population [1], [2], equivalent to 6 billion individuals, as illustrated in Figure 1.

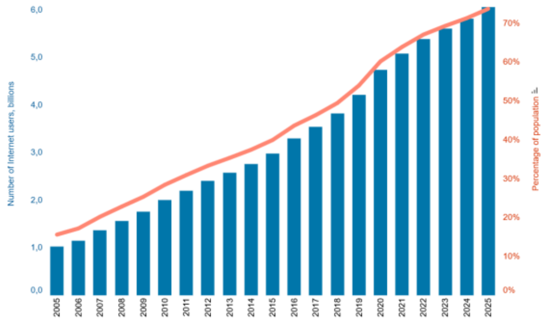

Fig. 1. *Internet user number*
Fountain: *International Telecommunication Union (ITU)*

This growth is closely linked to the adoption of mobile devices. By the end of 2024, approximately 4.4 billion people had a smartphone to access the internet, representing 54% of the world's population. Furthermore, over 80% of these connections are made via 4G and 5G networks [3]. This is further compounded by the fact that, by the beginning of 2025, there were 5.56 billion internet users, with 58% of the global population accessing the internet via their own devices. This has allowed for the consolidation of a massive web services ecosystem where video traffic accounts for 75.9% of the total volume of cellular data [1]. This volume of users operates on a large-scale network architecture, in which data traffic has grown exponentially. A consumption of approximately 1.5 zettabytes is projected, which has motivated the adoption of Machine Learning- based architectures for the proactive detection of anomalies [2] that guarantee the correct consumption of services.

The infrastructure for this demand is approximately 6,111 public data centers worldwide by the end of 2025 [4], whose operational capacity faces critical technical pressure due to the integration of generative Artificial Intelligence (AI) applications that require extreme cloud processing power compared to other popular services [4].

Although mobile broadband coverage currently reaches 96% of the global population [2], the sophistication of 5G networks, which cover 55.1% of the world's population, adds variables with a high technical complexity that make it difficult to monitor and detect atypical traffic patterns [2], [3].

In this context, anomaly detection plays a key role in preserving the operational stability of web service providers, especially in scenarios characterized by high demand variability. The central problem lies in the fact that traditional methods, based on static rules and fixed thresholds, may be insufficient to manage the dynamic nature and demand spikes generated by high-concurrency events on streaming platforms, e -commerce sites, financial services, and other platforms.

Technical complexity is pressured by the diversity of devices and the transition to 5G, as mentioned above, which already covers 55.1% of the world's population, but coexists with 16% of users who still depend on 3G technologies or basic devices [2], which can be seen in Figure 2, generating critical disparities in Quality of Service (QoS) and Quality of Experience (QoE).

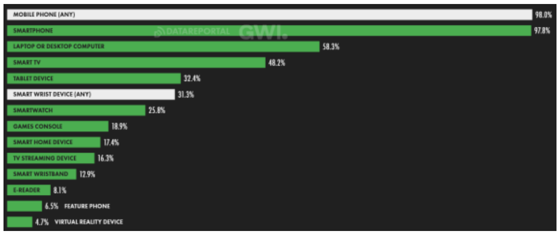

Fig. 2. *Trend of devices used to access the internet.*
Source: *Digital 2025 Global Overview Report.*

The objective of this article is to conduct a systematic literature review on methods for identifying anomalies in different web fields such as streaming services and IoT, among others, and to establish a framework of variables that can be used for future studies.

2. METHODOLOGY

This study is a systematic literature review, seeking to classify and synthesize, in a structured manner, the existing knowledge on how anomaly identification has been addressed in the field of Machine Learning. Its approach is mixed, combining qualitative analysis—examining approaches, trends, and conceptual frameworks—with quantitative elements, such as classifying articles by data type, learning methods, and objectives.

It should be clarified that this study does not aim to experimentally validate hypotheses, but rather to analyze how the scientific community has addressed the problem of anomaly detection, what solutions have been proposed, and what gaps still remain, offering a view of the state of the art that serves as a basis for future research and development.

2.2 Methodological design

The review focused on scientific articles published between 2021 and 2025 that addressed the detection or prediction of anomalies using machine learning techniques applied to web services, distributed systems, cloud infrastructures, telecommunications networks, and the Internet of Things (IoT). Both experimental research and reviews and surveys were considered.

The methodological process relied on predefined inclusion and exclusion criteria, a structured search strategy, and an analysis method that allowed for consistent comparison of studies.

2.3 Inclusion and exclusion criteria

2.3.1 Inclusion criteria

Articles that met the following criteria were included in the review:

- Publications related to the detection or prediction of anomalies using Machine Learning.
- Studies applied to web services, cloud systems, microservices, networks, IoT or large-scale digital environments.
- Literature reviews or surveys that analyze trends, challenges, or comparisons of techniques.
- Articles published between 2021 and 2025.
- Articles published in journals whose quartile is 1 or 2. Exceptions are accepted if the article is published in reliable sources.

2.3.2 Exclusion criteria

Studies that presented any of the following characteristics were excluded:

- Articles whose publication date is prior to 2021.
- Articles with a focus on industries other than those mentioned in the inclusion criteria.
- Articles that are not related to the use of Machine Learning.
- Articles published in journals with quartiles below 2.

2.4 Sample

In this research, 50 scientific articles were found, comprising 60% experimental studies that propose or evaluate machine learning models for anomaly detection, and 40% literature reviews and surveys that summarize and analyze the state of the art. This is represented in Figure 3. The aim was to combine practical evidence with theoretical analyses, facilitating a more comprehensive understanding of the topic.

2.5 Instruments for collecting and analyzing information

To determine which resources were useful for this study, and following the inclusion and exclusion criteria, an analysis matrix was created, designed to systematically collect the most relevant information from each selected article. This tool allowed for consistent comparison of the studies and reduced potential biases during the analysis. The matrix included the following aspects:

- Year of publication
- Type of study
- Application domain
- Type of data used
- Learning approach (supervised, unsupervised, or blended)
- Machine Learning techniques used
- Main objective of the study
- Main results and limitations

The above was established based on practices found in previous systematic reviews, some of which are part of this study.

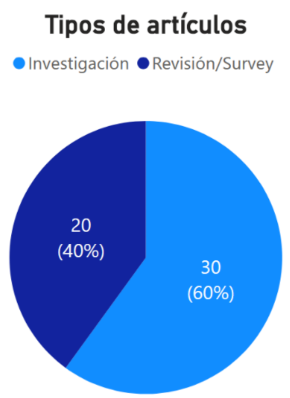

Fig. 3. *Types of articles used.*
Source: *own elaboration.*

2.6 Procedure

The research was conducted in several consecutive stages:

- Phase 1: The research topic was delimited and the objectives were defined, focused on analyzing the use of Machine Learning for the detection of anomalies in massive web services.
- Phase 2: A systematic search was conducted in recognized academic databases, using combinations of keywords related to anomaly detection, Machine Learning, web services, cloud computing and IoT.
- Phase 3: The articles found were initially evaluated by title and abstract. Then, the full text of the preselected studies was reviewed to verify compliance with the established criteria.
- Phase 4: Relevant information from each article was recorded in the analysis matrix. From this data, descriptive analyses were performed to identify trends, predominant approaches, and areas of application.
- Phase 5: The results were organized, building the current research and defining the basis for future lines of work.

3. RESULTS

The results presented in this section correspond to an analytical synthesis of the findings reported in the reviewed studies, and not to any experimentation carried out by the authors.

3.1. Quantitative evaluation of models and architectures

The analysis of studies reporting quantitative validations, summarized in Table 1, suggests that domain specialization is a recurring factor associated with the performance of the evaluated models [5], [6], [7] . In the reviewed studies on high-speed environments, such as IoT and 6G networks, outstanding performance is reported for techniques based on LSH and hybrid approaches, with accuracies close to 100% in certain datasets [8], [9]. Conversely, in critical security or medical domains, graph and attention architectures (GNN, Transformers) dominate the capture of nonlinear dependencies [10], [11].

Table 1: *Analysis on evaluation, proposal or comparison of models with performance metrics*

Model / Algorithm	Reference Dataset	Key Metric
LSH + Random Forest [8]	ToN-IoT, MQTT- IoT	Accuracy: 99.82%
HABBAs (AdaBoost + Bagging) [12]	CICDDOS2019	Accuracy: 99.95%
AE (Vector Reconstruction Error) [13]	CIDDS-001	F1 score: 100%
WDLog (Wide & Deep Learning) [14]	HDFS, BGL	F1-score > 90%
CCTAK (TCN + KAN + VAE) [15]	SKAB, SWaT	AUC-ROC: 0.8191
GIN (Graph Att. + Informer) [10]	MSL, SMAP	F1 score: 0.9604
SMOTETomek + ML Models [16]	WSN-DS	Accuracy: 99.92%
Diner (Memory AE + SNR) [17]	GAIA, NAB	F1 score: 0.706
AnyLog (BERT + SOM + AE) [18]	HDFS, BGL	Accuracy: 95.0%
AADS (Online Clustering) [19]	Breast Cancer, Ionosphere	F1-score: 76.13%
DGMM + ML Ensemble [20]	Cellular Traffic	RMSE: 0.026
XMLAD (Decision Tree Logic) [21]	NokiaFL (Real)	Recall : 100%
MDI vs Deep Learning [22]	UCR Archive	AUC-ROC: 0.66
Deep Isolation Forest (DIF)	Tabular / Graph / TS	AUC-PR: +144% vs iF [23], [24]
C-LSTM-AE [25]	Yahoo Webscope S5	Best F1 vs CNN/RNN

LSTM-Markov Hybrid [26]	Smart City Sensors	AD Efficiency: 96.03%
HGN2HIA (Heterog. Graph Att.) [27]	QoS Web Services	p- value : 0.005
DDQN-PER (RL) [28]	Occupancy Detection	Recall : 97.10%
Ttrees (CIAN Methodology) [29]	MONROE, Nokia	Accuracy: 99.6%
Generalized iForest (GIF) [30]	Benchmarks (Aloi)	Superior AUC-PR
CNN (1D, 2D, 3D) Models [6]	IoT-DS-2	Accuracy : >99.7%
Multi-method TS Evaluation [31]	SWAT, SMD	LSTM AUC: 0.863
Feedback K- means [32]	Web Service QoS	RMSE: 0.051
CAWAL Framework [7]	Web Portal Logs	Accuracy: 92.5%
XGBoost (African J.) [33]	CICIDS2017	F1 score: 0.987
VAE Stability Analysis [34]	Wireless Comm.	134 anomalies detected.
POT Threshold Selection [35]	Firewall logs	70% of Opt. MCC

Source: *own elaboration.*

3.2. Theoretical and taxonomic synthesis of the literature

Table 2 shows an analysis of the literature whose results focus on identifying relevant findings or challenges, such as the evolution from statistical filtering to explainable anomaly detection (XAD) methods [36], [37].

Some reviewed works point to the phenomenon known as “Clever Hans”, in which complex models achieve high precision by relying on spurious correlations of the data set [36].

Table 2: *Analysis of review articles, taxonomy, and theoretical frameworks*

Focus	**Datasets / Benchmarks Analyzed**	**Outstanding Discovery or Challenge**
Taxonomy of 52 algorithms [38], [39]	KDD Cup 99/98, NSL-KDD, UCI Repository	Classification into 7 mechanisms; isolation methods are the most scalable.
XAD (Explainable Detection) [36], [40]	Enron, Yelp, Amazon, Twitter Sybil, Elliptic	Identifying the "Clever Hans" effect: accuracy based on noise rather than causality.
Deep Transfer Learning (DTL) [41]	CWRU, IMS, PHM 2012 (Industrial Series)	Risk of negative transfer if the domain gap is excessive.
Graph Anomaly Detection (GAD) [40]	DBLP, Wikipedia, Reddit, Amazon, Enron	Systematization of structural anomalies in nodes, edges and subgraphs.
GNN in IIoT environments [11]	SWaT, WADI, BATADAL, Xcos, epanetCPA	Need to model evolutionary relational interdependence in cyber-physical systems.
Knowledge-based Systems (KBS) [37]	NSL-KDD, UNSW-NB15, DS2OS	Semantic systems offer greater interpretability, but face updating challenges.
DL for Log Detection [42]	HDFS, BGL, Thunderbird, Spirit, OpenStack	Challenge of template instability and massive volume of unstructured data.
MTSAD (Multivariate Series) [43]	CHB-MIT (EEG), Gas Pipeline, Yahoo Webscope	Granularity classification: point, interval, and full series anomalies.
Encrypted Network Traffic (SSL/TLS) [44]	CTU-13, CIC-IDS-2017, UNSW-NB15, MTA	Deep inspection infeasibility (DPI); dependence on statistical flow characteristics.
AutoML in Anomaly Detection [45]	General benchmarks of Outlier Detection	The CASH problem: difficulty of automating selection and adjustment without fundamental truth labels.
AD in Smart Environments [46]	Yahoo! S5, NAB, UCR Archive, Space Shuttle	Challenges of label scarcity and contextualization in smart environments.
Microservices and RCA [47]	KPIs, traces and logs of "Sock Shop"	Root cause analysis (RCA) requires correlation between KPIs and service logs.
Federated Machine Learning [48], [49]	datasets from networks and mobile devices	Balancing data privacy and statistical efficiency in decentralized training.

Federated Learning for IoT [48], [49]	CIFAR-10, MNIST, Imagenet (used as proxies)	Weight divergence in non-IID data reduces accuracy by up to 55%.
Machine Learning SLR (General) [39]	KDD Cup 99, NSL-KDD, UCI, Real- life datasets	Only 27% of the registered studies use purely unsupervised methods.
Sensor Systems (Multi-pers.) [50]	Bot- IoT, ODDS, NAB, Yahoo Webscope, ELKI	Importance of statistical and deep hybridization for data streaming.
Multimedia Streaming (User-centric) [51]	LIVE Netflix, CSIQ, LFOVIA, MCQoE, FCC	Modeling Quality of Experience (QoE) using subjective and objective metrics.
AD for WCN Network Failures [34]	GuifiSants (Actual Production), Linux Kernel features	The inclusion of hardware metrics (CPU/RAM) increases the ability to detect network faults.
LSTM in Technical Systems [23], [52]	Amazon, Wireless Sensor Network, Electricity consumption	Encoder-decoder architectures are superior for learning stationary and non-stationary time relationships.
Time-Series DL (Guidelines)	SWAT, SMAP, WADI, SMD, MSL	Noise is the critical factor that complicates detection in industrial control systems (CPS).

Source: *own elaboration.*

3.3. Real-time operational analysis

The third focus of this analysis is on the critical transition of detection models from laboratory environments to real production infrastructures, where operational sustainability and time accuracy are mandatory requirements.

In this context, the research of [9] is disruptive in proposing the StreamWNNov algorithm, which introduces a vital semantic distinction between the concept of "novelty", understood as an emerging pattern that must be taught as a model, and that of "anomaly", which is an exception that triggers immediate alerts.

The results obtained on Spanish electricity demand validate that this incremental learning approach allows reducing the prediction error (MAPE) to 2.07% in continuous flows, while maintaining a computational complexity of O(1) in its online phase [53]. This self-updating capability guarantees the viability of real-time processing without incurring the prohibitive costs of massive retraining, a scalability factor that resonates with the warnings of [53]. According to the trend analysis of the latter, there is a critical gap in the current state of the art: most intrusion detection models (IDS) for IoT networks are evaluated statically or offline, ignoring that the massive attack surface and computational load of Deep Learning can invalidate the immediate response needed to "zero-day" threats at the network edge.

Finally, the technical convergence between artificial intelligence and real-time systems (RTS) is examined by [54] , who emphasize that in safety-critical domains, such as drones or industrial robotics, adherence to worst-case execution time (WCET) and strict time constraints are as vital as the accuracy of the algorithms themselves. They conclude that the design of high-impact systems must be based on a systemic balance where the sophistication of the architecture is harmonized with the constraints of heterogeneous hardware and the deterministic latency required by the monitored physical environment.

3.4 Discussion of Trends and Results

Analysis of the reviewed studies reveals several recurring patterns, although not all are equally prevalent in every domain. The patterns that appear most frequently in recent literature are discussed below:

- Although multiple studies report better performance of Deep Learning in large-scale, multivariate scenarios, this dominance is not absolute. Some work shows that, under data or latency constraints, simpler approaches can offer comparable results.
- Historical transition, from 2021 to 2025, where a shift can be identified from proximity methods (k-NN, SVM) focused on point anomalies, towards dynamic graph architectures that model the interdependence of sensors in 6G networks.
- Several authors agree that accuracy, considered in isolation, is insufficient to evaluate the actual performance of anomaly detection systems. Current scientific validation requires post-hoc techniques such as SHAP to break down the contribution of each sensor,

transforming the "black box" into a diagnostic tool.

- Isolation-based models (Isolation Forest) offer the greatest scalability, being the only viable ones for real-time deployments on edge devices with limited power.
- The results reported for algorithms such as StreamWNNov suggest that the ability to distinguish between “novelty” and “anomaly” is a relevant aspect for the sustainability of systems.

4. CONCLUSIONS

Anomaly detection has evolved from a peripheral statistical task to a cornerstone of resilience in hyperconnected digital ecosystems. After evaluating 50 recent articles, it is concluded that there is no one-size-fits-all solution; the effectiveness of a system depends on the balance between the nature of the data, latency requirements, and the need for transparency in decision-making. While deep learning dominates in identifying irregularities in complex multivariate flows, classical and knowledge-based methods remain undeniably relevant due to their computational efficiency and native interpretability.

Several authors suggest that the incorporation of explainability, federated learning, and autonomous mechanisms could become a dominant line of research in the coming years, especially in regulated or distributed environments.

The ability to provide tangible reasoning behind each alert not only increases user confidence but also facilitates proactive prevention of systemic failures. Finally, the integration of incremental learning frameworks will allow AI to dynamically adapt to a volatile world, transforming simple alerts into actionable knowledge for the stability of critical global infrastructure.

REFERENCES

[1] Meltwater, “Digital 2025 Global Overview Report,” Meltwater, 1, Feb. 2025. [Online]. Available: https://drive.google.com/file/d/1fonmdNLLMbVFB9f-azux0ZTJPsw1CeZ8/view

[2] “Measuring digital development: Facts and Figures 2025,” ITU. Accessed: Feb. 04, 2026. [Online]. Available: https://www.itu.int/hub/publication/d-ind-ict_mdd-2025-3/

[3] “The State of Mobile Internet Connectivity 2025: Overview Report | GSMA Intelligence.” Accessed: Feb. 05, 2026. [Online]. Available: https://www.gsmaintelligence.com/research/the-state-of-mobile-internet-connectivity-2025-overview-report

[4] “How Many Data Centers Are There and Where Are They Being Built?” Accessed: Feb. 05, 2026. [Online]. Available: https://www.abiresearch.com/blog/data-centers-by-region-size-company

[5] K. Choi, J. Yi, C. Park, and S. Yoon, “Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines,” *IEEE Access*, vol. 9, pp. 120043–120065, 2021, doi: 10.1109/ACCESS.2021.3107975.

[6] I. Ullah and Q. H. Mahmoud, “Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks,” *IEEE Access*, vol. 9, pp. 103906–103926, 2021, doi: 10.1109/ACCESS.2021.3094024.

[7] Ö. Canay and Ü. Kocabıçak, “Predictive modeling and anomaly detection in large-scale web portals through the CAWAL framework,” *Knowledge-Based Systems*, vol. 306, p. 112710, Dec. 2024, doi: 10.1016/j.knosys.2024.112710.

[8] M. L. Hernandez-Jaimes, A. Martinez-Cruz, and K. A. Ramírez-Gutiérrez, “A Machine Learning approach for anomaly detection on the Internet of Things based on Locality-Sensitive Hashing,” *Integration*, vol. 96, p. 102159, May 2024, doi: 10.1016/j.vlsi.2024.102159.

[9] S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, “Machine Learning and Deep Learning Techniques for Internet of Things Network Anomaly Detection—Current Research Trends,” *Sensors*, vol. 24, no. 6, 2024, doi: 10.3390/s24061968.

[10] C. Wang and G. Liu, “From anomaly detection to classification with graph attention and transformer for multivariate time series,” *Advanced Engineering Informatics*, vol. 60, p. 102357, Apr. 2024, doi: 10.1016/j.aei.2024.102357.

[11] Y. Wu, H. -N. Dai, and H. Tang, “Graph Neural Networks for Anomaly Detection in Industrial Internet of Things,” *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214–9231, Jun. 2022, doi: 10.1109/JIOT.2021.3094295.

[12] M. M. Saeed, R. A. Saeed, M. Abdelhaq, R. Alsaqour, M. K. Hasan, and R. A. Mokhtar, “Anomaly Detection in 6G Networks Using

Machine Learning Methods," *Electronics*, vol. 12, no. 15, 2023, doi: 10.3390/electronics12153300.

[13] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, p. 1, Jan. 2023, doi: 10.1186/s42400-022-00134-9.

[14] W. Niu, X. Liao, S. Huang, Y. Li, X. Zhang, and B. Li, "A robust Wide & Deep learning framework for log-based anomaly detection," *Applied Soft Computing*, vol. 153, 2024, doi: 10.1016/j.asoc.2024.111314.

[15] Y. Abudurexiti, G. Han, F. Zhang, and L. Liu, "An explainable unsupervised anomaly detection framework for Industrial Internet of Things," *Computers & Security*, vol. 148, p. 104130, Jan. 2025, doi: 10.1016/j.cose.2024.104130.

[16] Md. A. Talukder, S. Sharmin, M. A. Uddin, M. M. Islam, and S. Aryal, "MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs," *Int. J. Inf. Secur.*, vol. 23, no. 3, pp. 2139–2158, Jun. 2024, doi: 10.1007/s10207-024-00833-z.

[17] Y. Jing *et al.*, "Diner: Interpretable Anomaly Detection for Seasonal Time Series in Web Services," *IEEE Transactions on Services Computing*, vol. 17, no. 5, pp. 2248–2260, Oct. 2024, doi: 10.1109/TSC.2024.3422894.

[18] A. Aziz and K. Munir, "Anomaly Detection in Logs Using Deep Learning," *IEEE Access*, vol. 12, pp. 176124–176135, 2024, doi: 10.1109/ACCESS.2024.3506332.

[19] M. Y. Iqbal Basheer *et al.*, "Autonomous anomaly detection for streaming data," *Knowledge-Based Systems*, vol. 284, p. 111235, Jan. 2024, doi: 10.1016/j.knosys.2023.111235.

[20] A. Gorshenin, A. Kozlovskaya, S. Gorbunov, and I. Kochetkova, "Mobile network traffic analysis based on probability-informed machine learning approach," *Computer Networks*, vol. 247, p. 110433, Jun. 2024, doi: 10.1016/j.comnet.2024.110433.

[21] J. M. Ramírez, F. Díez, P. Rojo, V. Mancuso, and A. Fernández-Anta, "Explainable machine learning for performance anomaly detection and classification in mobile networks," *Computer Communications*, vol. 200, pp. 113–131, Feb. 2023, doi: 10.1016/j.comcom.2023.01.003.

[22] F. Rewicki, J. Denzler, and J. Niebling, "Is It Worth It? Comparing Six Deep and Classical Methods for Unsupervised Anomaly Detection in Time Series," *Applied Sciences*, vol. 13, no. 3, 2023, doi: 10.3390/app13031778.

[23] U. A. Usmani, I. Abdul Aziz, J. Jaafar, and J. Watada, "Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research," *IEEE Access*, vol. 12, pp. 174564–174590, 2024, doi: 10.1109/ACCESS.2024.3495819.

[24] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep Isolation Forest for Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, Dec. 2023, doi: 10.1109/TKDE.2023.3270293.

[25] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 112–122, Jan. 2022, doi: 10.1109/TSMC.2020.2968516.

[26] S. V. and S. A., "LSTM-Markov based efficient anomaly detection algorithm for IoT environment," *Applied Soft Computing*, vol. 136, p. 110054, Mar. 2023, doi: 10.1016/j.asoc.2023.110054.

[27] S. Lv, F. Yi, P. He, and C. Zeng, "QoS Prediction of Web Services Based on a Two-Level Heterogeneous Graph Attention Network," *IEEE Access*, vol. 10, pp. 1871–1880, 2022, doi: 10.1109/ACCESS.2021.3138127.

[28] D. Fährmann, N. Jorek, N. Damer, F. Kirchbuchner, and A. Kuijper, "Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments," *IEEE Access*, vol. 10, pp. 60836–60848, 2022, doi: 10.1109/ACCESS.2022.3179720.

[29] M. Moulay, R. G. Leiva, P. J. Rojo Maroni, F. Diez, V. Mancuso, and A. Fernández Anta, "Automated identification of network anomalies and their causes with interpretable machine learning: The CIAN methodology and TTrees implementation," *Computer Communications*, vol. 191, pp. 327–348, Jul. 2022, doi: 10.1016/j.comcom.2022.05.013.

[30] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tourneret, "Generalized isolation forest for anomaly detection," *Pattern Recognition Letters*, vol. 149, pp. 109–119, Sep. 2021, doi: 10.1016/j.patrec.2021.05.022.

[31] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. Salvo Rossi, "Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance

Analysis and Future Directions," *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052844.

[32] Y. Song, "Web service reliability prediction based on machine learning," *Computer Standards & Interfaces*, vol. 73, p. 103466, Jan. 2021, doi: 10.1016/j.csi.2020.103466.

[33] V. R. Gudelli, "Anomaly Detection in Cloud Networks Using Machine Learning Algorithms," Jun. 2024, doi: 10.5281/ZENODO.15271016.

[34] L. Cerdà-Alabern, G. Iuhasz, and G. Gemmi, "Anomaly detection for fault detection in wireless community networks using machine learning," *Computer Communications*, vol. 202, pp. 191–203, Mar. 2023, doi: 10.1016/j.comcom.2023.02.019.

[35] A. Komadina, M. Martinić, S. Groš, and Ž. Mihajlović, "Comparing Threshold Selection Methods for Network Anomaly Detection," *IEEE Access*, vol. 12, pp. 124943–124973, 2024, doi: 10.1109/ACCESS.2024.3452168.

[36] Z. Li, Y. Zhu, and M. Van Leeuwen, "A Survey on Explainable Anomaly Detection," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 1, pp. 1–54, Jan. 2024, doi: 10.1145/3609333.

[37] A. Q. Khan, S. El Jaouhari, N. Tamani, and L. Mroueh, "Knowledge-based anomaly detection: Survey, challenges, and future directions," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108996, Oct. 2024, doi: 10.1016/j.engappai.2024.108996.

[38] D. Samariya and A. Thakkar, "A Comprehensive Survey of Anomaly Detection Algorithms," *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, Jun. 2023, doi: 10.1007/s40745-021-00362-9.

[39] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[40] X. Ma *et al.*, "A Comprehensive Survey on Graph Anomaly Detection With Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12012–12038, Dec. 2023, doi: 10.1109/TKDE.2021.3118815.

[41] P. Yan *et al.*, "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE Access*, vol. 12, pp. 3768–3789, 2024, doi: 10.1109/ACCESS.2023.3349132.

[42] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, "Deep learning for anomaly detection in log data: A survey," *Machine Learning with Applications*, vol. 12, p. 100470, Jun. 2023, doi: 10.1016/j.mlwa.2023.100470.

[43] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion*, vol. 91, pp. 93–102, Mar. 2023, doi: 10.1016/j.inffus.2022.10.008.

[44] I. H. Ji, J. H. Lee, M. J. Kang, W. J. Park, S. H. Jeon, and J. T. Seo, "Artificial Intelligence-Based Anomaly Detection Technology over Encrypted Traffic: A Systematic Literature Review," *Sensors*, vol. 24, no. 3, p. 898, Jan. 2024, doi: 10.3390/s24030898.

[45] M. Bahri, F. Salutari, A. Putina, and M. Sozio, "AutoML: state of the art with a focus on anomaly detection, challenges, and research directions," *International Journal of Data Science and Analytics*, vol. 14, no. 2, pp. 113–126, Aug. 2022, doi: 10.1007/s41060-022-00309-0.

[46] D. Fährmann, L. Martín, L. Sánchez, and N. Damer, "Anomaly Detection in Smart Environments: A Comprehensive Survey," *IEEE Access*, vol. 12, pp. 64006–64049, 2024, doi: 10.1109/ACCESS.2024.3395051.

[47] J. Soldani and A. Brogi, "Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–39, Mar. 2023, doi: 10.1145/3501297.

[48] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated Learning for Internet of Things: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.

[49] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated Machine Learning: Survey, Multi-Level Classification, Desirable Criteria and Future Directions in Communication and Networking Systems," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1342–1397, 2021, doi: 10.1109/COMST.2021.3058573.

[50] L. Erhan *et al.*, "Smart anomaly detection in sensor systems: A multi-perspective review," *Information Fusion*, vol. 67, pp. 64–79, Mar. 2021, doi: 10.1016/j.inffus.2020.10.001.

[51] M. Ghosh and C. Singhal, "A review on machine learning based user-centric multimedia streaming techniques," *Computer Communications*, vol. 231, p. 108011, Feb. 2025, doi: 10.1016/j.comcom.2024.108011.

[52] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, “A survey on anomaly detection for technical systems using LSTM networks,” *Computers in Industry*, vol. 131, p. 103498, Oct. 2021, doi: 10.1016/j.compind.2021.103498.

[53] L. Melgar-García, D. Gutiérrez-Avilés, C. Rubio-Escudero, and A. Troncoso, “Identifying novelties and anomalies for incremental learning in streaming time series forecasting,” *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106326, Aug. 2023, doi: 10.1016/j.engappai.2023.106326.

[54] J. Bian *et al.*, “Machine Learning in Real-Time Internet of Things (IoT) Systems: A Survey,” *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8364–8386, Jun. 2022, doi: 10.1109/JIOT.2022.3161050.