Revista Científica Multidisciplinar G-ner@ndo ISSN: 2806-5905

Arquitecturas de Aprendizaje Federado para Detectar Anomalías en Redes de Acceso de Radio Abiertas

Federated Learning Architectures for Detecting Anomalies in Open Radio Access Networks

Jhonatan Leonardo Calderon De La Cruz & David Fernando Zambrano Montenegro

CIENCIA, TECNOLOGÍA Y SOCIEDAD

Julio - diciembre, V°7 - N°2; 2026

Recibido: 03-09-2026
Aceptado: 04-09-2026
Publicado: 09-09-2026

PAIS

- Ecuador, Portoviejo
- Ecuador, Portoviejo

INSTITUCION

- Universidad Técnica de Manabí
- Universidad Técnica de Manabí

CORREO:

✉ jcalderon3399@utm.edu.ec
✉ david.zambrano@utm.edu.ec

ORCID:

https://orcid.org/0009-0003-9285-0642
https://orcid.org/0000-0002-8833-1546

FORMATO DE CITA APA.

Calderon, J. & Zambrano, D. (2026). Arquitecturas de Aprendizaje Federado para Detectar Anomalías en Redes de Acceso de Radio Abiertas. *Revista G-ner@ndo, V°7 (N°2). Pág. 4003 – 4032.*

Resumen

La desagregación funcional, la apertura de interfaces y la programabilidad de las redes de acceso de radio abiertas favorecen la automatización de su gestión y optimización, aunque también amplían la superficie de ataque y multiplican los puntos donde pueden manifestarse comportamientos anómalos. El aprendizaje federado permite entrenar modelos colaborativos sin transferir los datos brutos generados por los componentes de la red. Este estudio analizó las arquitecturas de aprendizaje federado propuestas para detectar anomalías en redes de acceso de radio abiertas, atendiendo a su topología, distribución funcional, mecanismos de agregación, modelos de aprendizaje, medidas de seguridad y condiciones de evaluación. Se realizó una revisión sistemática de la literatura conforme a PRISMA 2020, considerando publicaciones de 2022 a 2026 indexadas en IEEE Xplore, Scopus y Web of Science. De 91 registros identificados, 55 correspondieron a estudios únicos y 15 cumplieron los criterios de elegibilidad. Las propuestas se agruparon en arquitecturas cliente-servidor, jerárquicas, descentralizadas entre pares y conceptuales o híbridas. Las configuraciones centralizadas y jerárquicas fueron predominantes, mientras que los diseños descentralizados eliminaron el agregador único a costa de una mayor sobrecarga de comunicación y complejidad de coordinación. FedAvg fue el algoritmo de agregación más recurrente y las aplicaciones se concentraron en la detección de intrusiones y ataques de denegación de servicio. Los resultados muestran un desempeño elevado en escenarios controlados, pero su generalización continúa limitada por la escasa validación operacional, la heterogeneidad de los conjuntos de datos y la insuficiente evaluación frente a actores maliciosos.

Palabras clave: aprendizaje federado; detección de anomalías; redes de acceso de radio abiertas; seguridad de redes; arquitectura distribuida.

Abstract

The functional disaggregation, open interfaces, and programmability of open radio access networks facilitate automated network management and optimization, while also expanding the attack surface and increasing the number of points at which anomalous behavior may emerge. Federated learning enables collaborative model training without transferring the raw data generated by network components. This study analyzed federated learning architectures proposed for anomaly detection in open radio access networks, focusing on their topology, functional distribution, aggregation mechanisms, learning models, security measures, and evaluation conditions. A systematic literature review was conducted following PRISMA 2020, covering publications from 2022 to 2026 indexed in IEEE Xplore, Scopus, and Web of Science. Of 91 identified records, 55 corresponded to unique studies and 15 met the eligibility criteria. The selected proposals were grouped into client-server, hierarchical, peer-to-peer decentralized, and conceptual or hybrid architectures. Centralized and hierarchical configurations predominated, whereas decentralized designs removed the single aggregator at the cost of higher communication overhead and coordination complexity. FedAvg was the most recurrent aggregation algorithm, and the applications focused mainly on intrusion and denial-of-service attack detection. The findings show high performance under controlled conditions; however, their generalizability remains constrained by limited operational validation, heterogeneous datasets, and insufficient evaluation against malicious actors.

Keywords: federated learning; anomaly detection; open radio access networks; network security; distributed architecture.

Introducción

La evolución de las redes móviles hacia infraestructuras abiertas, virtualizadas y programables ha modificado la organización tradicional de la red de acceso de radio. Las redes de acceso de radio abiertas (Open Radio Access Network, O-RAN) desagregan funciones previamente integradas, incorporan interfaces abiertas y permiten distribuir capacidades de control e inteligencia entre distintos componentes de la infraestructura. Esta organización favorece la interoperabilidad entre proveedores y facilita la incorporación de aplicaciones y modelos de aprendizaje automático para automatizar tareas de gestión y optimización de la red (Polese et al., 2023).

La apertura arquitectónica también modifica el perfil de seguridad. La coexistencia de componentes virtualizados, interfaces estandarizadas, aplicaciones de terceros y funciones distribuidas amplía los puntos susceptibles de compromiso y dificulta mantener una visión global del comportamiento de la red. Las amenazas pueden afectar el tráfico de usuario, las funciones de control, las aplicaciones xApps y rApps, los modelos de aprendizaje o los mecanismos utilizados para coordinar decisiones entre dominios. La detección de anomalías deja así de ser un problema localizado y pasa a depender de información producida en múltiples componentes con capacidades y niveles de confianza diferentes (Polese et al., 2023; Attanayaka et al., 2023).

La centralización de esta información permitiría disponer de una visión conjunta, pero introduce restricciones relacionadas con privacidad, escalabilidad, latencia y disponibilidad. Además, en escenarios con múltiples operadores o dominios administrativos, la transferencia de tráfico o telemetría hacia un repositorio común puede ser técnica u organizativamente inviable. El aprendizaje federado ofrece una alternativa al permitir que los participantes entrenen modelos sobre sus datos locales y compartan únicamente parámetros o actualizaciones destinadas a construir conocimiento global (McMahan et al., 2017; Kairouz et al., 2021). Esta característica resulta especialmente compatible con una arquitectura

distribuida como O-RAN, donde los datos pueden originarse en nodos E2, unidades distribuidas, unidades centralizadas, controladores inteligentes o redes pertenecientes a distintos operadores.

La aplicación del aprendizaje federado en O-RAN ha evolucionado hacia configuraciones cada vez más diversas. Attanayaka et al. (2023) exploraron esquemas entre pares para la detección de anomalías y analizaron variantes orientadas a reducir la dependencia de un agregador central. Benzaïd et al. (2024) incorporaron aprendizaje continuo para adaptar los modelos frente a nuevas amenazas, mientras que Rumesh et al. (2024) integraron una arquitectura federada jerárquica dentro de un gemelo digital. Otros trabajos han combinado aprendizaje federado con aprendizaje por refuerzo, confianza cero, blockchain, colaboración entre redes privadas y mecanismos completamente descentralizados (Abou El Houda et al., 2024; Moudoud et al., 2024; Hung et al., 2025; Spantideas et al., 2026).

Esta diversidad confirma la viabilidad del paradigma, pero dificulta determinar qué configuraciones resultan más apropiadas para cada escenario. Las propuestas difieren en la ubicación de los clientes, el número de niveles de agregación, los modelos de aprendizaje, los algoritmos de combinación de actualizaciones, los mecanismos de protección y las condiciones experimentales. Asimismo, los estudios recurren a conjuntos de datos y métricas heterogéneos, y solo una parte de ellos valida sus arquitecturas mediante bancos de pruebas o entornos próximos a una implementación O-RAN. En consecuencia, comparar únicamente los valores de exactitud o puntuación F1 no permite establecer qué diseño ofrece mejores condiciones de escalabilidad, resiliencia, privacidad o eficiencia operacional.

La brecha identificada radica en la fragmentación de la evidencia disponible sobre la organización arquitectónica y los compromisos de diseño de estas propuestas. Los estudios analizados abordan de manera heterogénea la ubicación de los participantes, los mecanismos

de agregación, las amenazas, los escenarios de evaluación y las medidas de protección, lo que dificulta establecer patrones comparables entre las diferentes configuraciones federadas.

A partir de esta problemática, el objetivo de este estudio fue analizar las arquitecturas de aprendizaje federado propuestas para detectar anomalías en redes de acceso de radio abiertas, considerando su estructura, distribución funcional, mecanismos de agregación, modelos de aprendizaje, medidas de seguridad, amenazas abordadas, conjuntos de datos, condiciones experimentales y limitaciones. Para ello se desarrolló una revisión sistemática de la literatura conforme a PRISMA 2020, considerando publicaciones comprendidas entre 2022 y 2026 e indexadas en IEEE Xplore, Scopus y Web of Science.

El artículo se organiza en cuatro apartados principales. Materiales y Métodos describe el protocolo de revisión, las fuentes consultadas, los criterios de elegibilidad y el procedimiento de extracción y síntesis. Análisis de Resultados presenta el proceso de selección y caracteriza las arquitecturas, modelos, amenazas y condiciones experimentales identificadas. Discusión examina los compromisos arquitectónicos, la madurez de la evidencia y las brechas de investigación. Finalmente, Conclusiones sintetiza las implicaciones derivadas de la revisión.

Métodos y Materiales

Diseño y protocolo de la revisión

Se desarrolló una revisión sistemática de la literatura orientada a identificar, caracterizar y comparar las arquitecturas de aprendizaje federado aplicadas a la detección de anomalías en redes de acceso de radio abiertas. El alcance comprendió propuestas destinadas a reconocer intrusiones, ataques, amenazas, fallas o comportamientos maliciosos mediante esquemas de aprendizaje distribuido integrados funcionalmente con componentes o servicios de O-RAN.

El proceso se estructuró conforme a las directrices PRISMA 2020 (Page et al., 2021). Antes de ejecutar las búsquedas se estableció un protocolo que definió el objetivo, las

preguntas de investigación, las fuentes de información, el periodo de publicación, los criterios de elegibilidad, el procedimiento de identificación y eliminación de registros duplicados, la estrategia de selección, la evaluación de calidad metodológica y las variables de extracción. Esta definición previa permitió aplicar criterios uniformes durante las distintas fases y preservar la trazabilidad de las decisiones adoptadas.

La revisión consideró publicaciones comprendidas entre 2022 y 2026, redactadas en inglés o español y difundidas como artículos de revista o trabajos completos de congresos sometidos a revisión por pares. La búsqueda definitiva se ejecutó el 21 de julio de 2026.

Preguntas de investigación

Las preguntas de investigación se formularon para examinar las dimensiones arquitectónica, funcional, algorítmica, experimental y prospectiva de las soluciones identificadas. Su estructura orientó tanto la extracción de información como la síntesis comparativa posterior.

Tabla 1. *Preguntas de investigación de la revisión sistemática*

Código	Pregunta de investigación
PI1	¿Qué arquitecturas de aprendizaje federado se han propuesto para detectar anomalías en redes de acceso de radio abiertas?
PI2	¿Cómo se estructuran estas arquitecturas en términos de participantes, agregadores, distribución del entrenamiento y mecanismos de comunicación?
PI3	¿Qué modelos de aprendizaje, algoritmos de agregación y mecanismos de seguridad o privacidad se utilizan?
PI4	¿Qué anomalías, ataques, conjuntos de datos, escenarios experimentales y métricas de evaluación se consideran?
PI5	¿Cuáles son las tendencias, limitaciones, desafíos y oportunidades de investigación identificados?

Fuentes de información y estrategia de búsqueda

La búsqueda bibliográfica se efectuó en IEEE Xplore, Scopus y Web of Science Core Collection. Estas fuentes se seleccionaron por su cobertura de publicaciones especializadas en telecomunicaciones, redes móviles, aprendizaje automático y ciberseguridad, así como por la inclusión de literatura revisada por pares procedente de revistas y congresos científicos.

La estrategia se construyó a partir de tres dominios conceptuales: aprendizaje federado, redes de acceso de radio abiertas y detección de anomalías o amenazas. La expresión de referencia fue:

("federated learning" OR "federated deep learning" OR "federated reinforcement learning") AND ("O-RAN" OR "Open RAN" OR "Open Radio Access Network") AND ("anomaly detection" OR "intrusion detection" OR "attack detection" OR "threat detection" OR "fault detection" OR "malicious behavior")

La cadena se adaptó a la sintaxis, los campos de consulta y los operadores disponibles en cada base de datos, manteniendo la equivalencia semántica de los conceptos centrales. De forma complementaria, se realizaron búsquedas dirigidas para recuperar estudios cuya terminología específica pudiera limitar su identificación mediante la consulta principal.

Tabla 2. *Estrategia de búsqueda aplicada en las fuentes consultadas*

Fuente	Campos de búsqueda	Adaptación de la estrategia	Fecha
IEEE Xplore	Metadatos bibliográficos disponibles en la plataforma	Combinación de términos de aprendizaje federado, O-RAN/Open RAN y detección de anomalías o amenazas; búsqueda complementaria dirigida.	21/07/2026
Scopus	Título, resumen y palabras clave	Adaptación de la expresión de referencia a la sintaxis de Scopus y a sus campos bibliográficos.	21/07/2026
Web of Science Core Collection	Tema	Adaptación de la expresión de referencia a la sintaxis de Web of Science Core Collection.	21/07/2026

Google Scholar no se utilizó como fuente primaria debido a las limitaciones para reproducir las consultas y exportar registros bajo condiciones homogéneas. La ACM Digital Library tampoco se incorporó, puesto que no se dispuso de acceso institucional durante la ejecución de la revisión.

Criterios de elegibilidad

Los criterios de selección se definieron antes del cribado para preservar la correspondencia entre el objetivo de la revisión y la evidencia analizada. Se consideraron elegibles los trabajos publicados durante el periodo establecido, disponibles en inglés o

español y presentados como artículos de revista o trabajos completos de congreso revisados por pares. Además, cada publicación debía abordar explícitamente O-RAN u Open RAN y proponer, implementar, evaluar o analizar una arquitectura de aprendizaje federado destinada a detectar anomalías, intrusiones, ataques, amenazas, fallas o comportamientos maliciosos.

La inclusión requirió asimismo información suficiente para examinar los componentes esenciales de la propuesta, entre ellos los participantes, la ubicación del entrenamiento, el mecanismo de agregación, el modelo de aprendizaje, el escenario de aplicación o los resultados de evaluación.

Se excluyeron las investigaciones sin una relación funcional explícita con la arquitectura O-RAN; aquellas que mencionaban el aprendizaje federado únicamente como antecedente o posible trabajo futuro; y los estudios centrados exclusivamente en optimización, asignación de recursos, autenticación, privacidad o control sin incorporar una tarea de detección. También se descartaron revisiones, editoriales, resúmenes extendidos, publicaciones situadas fuera del periodo establecido y documentos cuyo texto completo no permitiera valorar adecuadamente la propuesta.

Cuando se identificaron versiones sucesivas de una misma investigación, se conservó aquella con mayor desarrollo metodológico, experimental o analítico. Este criterio evitó contabilizar como evidencia independiente resultados procedentes del mismo estudio.

Gestión de registros e identificación de duplicados

Los resultados procedentes de las tres fuentes se integraron en una matriz única y se normalizaron los campos bibliográficos necesarios para su comparación. El DOI normalizado se utilizó como identificador principal para detectar coincidencias. Cuando este dato estaba ausente o presentaba variaciones de formato, se contrastaron el título normalizado, los autores, el año de publicación y los demás metadatos disponibles.

Ante registros coincidentes entre distintas bases de datos se conservó una sola referencia, priorizando aquella que proporcionara la información bibliográfica más completa. La fuente de procedencia se mantuvo registrada para preservar la trazabilidad del proceso.

Selección de los estudios

La selección se desarrolló en dos etapas. Inicialmente se examinaron títulos y resúmenes para determinar su correspondencia con el objetivo y los criterios de elegibilidad. Los registros cuya información resultaba insuficiente para adoptar una decisión fundada se conservaron para la siguiente fase, evitando exclusiones sustentadas únicamente en resúmenes incompletos.

Posteriormente se realizó la lectura integral de las publicaciones potencialmente elegibles. Esta evaluación verificó la existencia de una relación explícita entre aprendizaje federado, detección de anomalías y arquitectura O-RAN, así como la disponibilidad de información suficiente sobre organización funcional, metodología o evaluación. Las razones de exclusión se documentaron individualmente para mantener la trazabilidad y representar posteriormente el proceso mediante el diagrama de flujo PRISMA 2020.

Evaluación de la calidad metodológica

La solidez de los estudios incluidos se examinó mediante diez criterios: claridad del objetivo o problema de investigación; relación explícita con O-RAN; descripción de la arquitectura de aprendizaje federado; identificación de participantes, localización del entrenamiento y agregación; definición de la anomalía o amenaza; descripción del método, conjunto de datos y escenario de evaluación; pertinencia de las métricas; correspondencia entre resultados y objetivo; análisis de limitaciones o amenazas a la validez; y disponibilidad de información suficiente para favorecer la reproducibilidad.

Cada criterio recibió una puntuación de 1 cuando se encontraba claramente satisfecho, 0,5 cuando la información era parcial y 0 cuando estaba ausente o no podía verificarse. La

puntuación acumulada permitió clasificar los estudios en tres niveles: calidad alta, entre 8 y 10 puntos; calidad media, entre 6 y 7,5; y calidad baja, para valores inferiores a 6.

La evaluación metodológica no funcionó como criterio automático de exclusión. Su finalidad fue contextualizar la fortaleza de la evidencia y distinguir entre estudios con validación experimental, propuestas conceptuales y trabajos con diferentes niveles de detalle y reproducibilidad.

Extracción de los datos

La extracción se realizó mediante una matriz estructurada vinculada con las cinco preguntas de investigación. Para cada publicación se registraron el año, la arquitectura federada, la topología, los participantes, la ubicación del agregador, la distribución funcional del entrenamiento, el algoritmo de agregación, el modelo de aprendizaje, la anomalía o amenaza analizada, el conjunto de datos, la distribución de los datos, el entorno experimental, el número de participantes, las rondas o épocas de entrenamiento, las métricas, los principales resultados, los mecanismos de seguridad y privacidad, las ventajas, las limitaciones y el nivel de calidad metodológica.

Se distinguió entre información declarada explícitamente por los autores y datos no especificados. Cuando una publicación no proporcionaba información sobre distribución de datos, mecanismo de agregación, número de rondas u otros parámetros relevantes, el elemento se registró como no informado, sin inferirlo a partir de configuraciones habituales del aprendizaje federado.

Clasificación y síntesis de la evidencia

Las propuestas se clasificaron según la relación funcional entre los participantes y las entidades responsables de la agregación. Esta organización permitió distinguir arquitecturas cliente-servidor, jerárquicas, descentralizadas entre pares y conceptuales o híbridas. La

clasificación facilitó la comparación de estudios que empleaban denominaciones distintas, pero compartían patrones equivalentes de entrenamiento y coordinación.

Las amenazas se organizaron según la tarea principal abordada: detección de intrusiones y tráfico malicioso multiclase, ataques de denegación de servicio, interferencia maliciosa, anomalías de tráfico o carga y anomalías relacionadas con políticas u orquestación. Los escenarios de evaluación se diferenciaron entre simulación, emulación, banco de pruebas, utilización de datos reales y propuestas conceptuales.

La síntesis adoptó un enfoque narrativo y comparativo. No se realizó un metaanálisis debido a la heterogeneidad de los conjuntos de datos, las clases de ataque, las distribuciones entre clientes, los modelos de aprendizaje, las topologías federadas y las métricas empleadas. Por esta razón, el análisis no estableció una jerarquía basada exclusivamente en valores de exactitud o puntuación F1, sino que examinó de manera conjunta la estructura arquitectónica, la distribución funcional, los mecanismos de agregación y protección, el entorno experimental y las condiciones bajo las cuales se obtuvieron los resultados.

Análisis de resultados

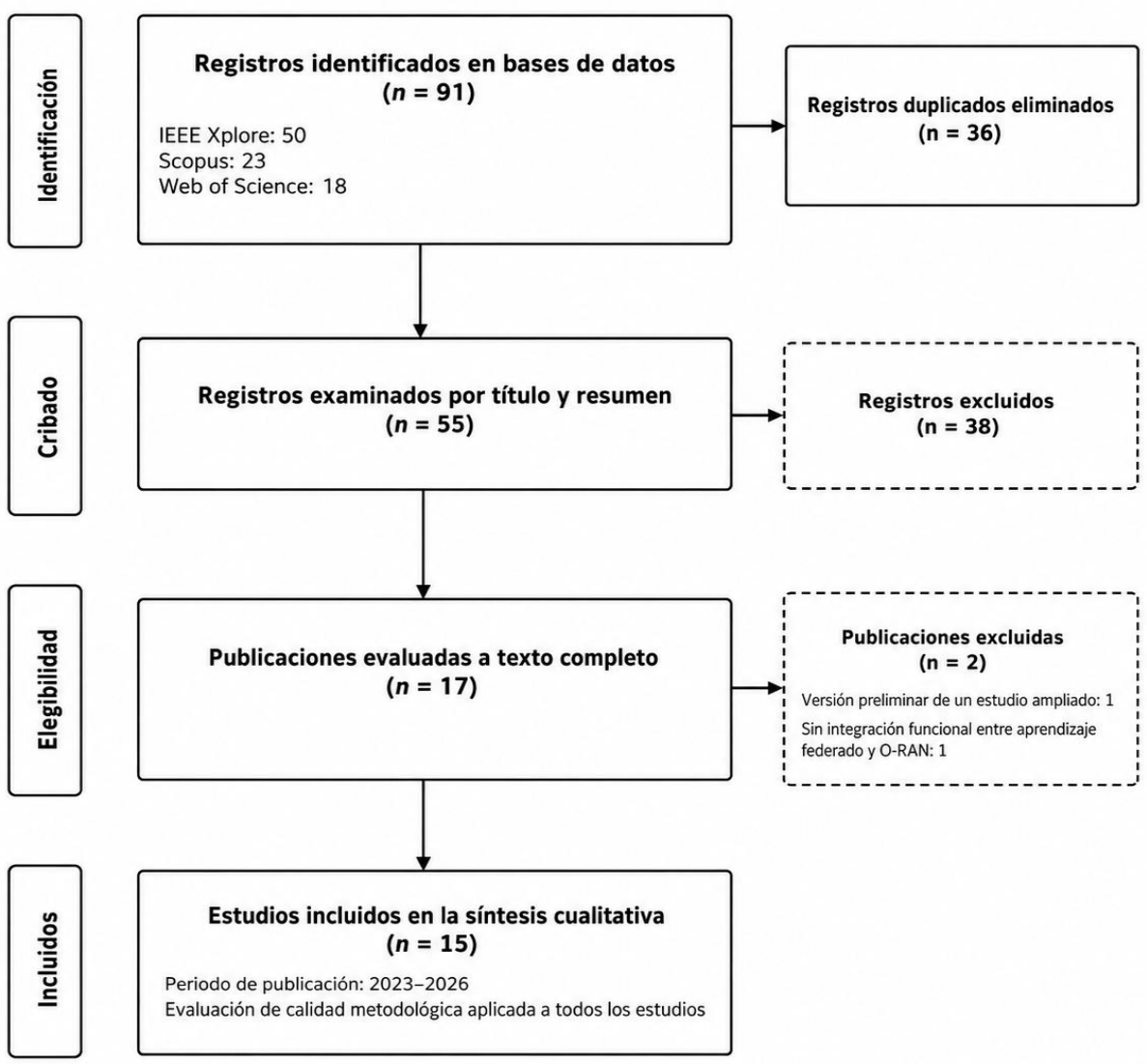
Selección y caracterización de los estudios

La búsqueda en IEEE Xplore, Scopus y Web of Science identificó 91 registros. IEEE Xplore aportó 50 referencias, incluidas 49 recuperadas mediante la consulta principal y una obtenida mediante la búsqueda complementaria; Scopus proporcionó 23 registros y Web of Science, 18. Tras la identificación y eliminación de 36 apariciones duplicadas, 55 publicaciones únicas fueron sometidas al cribado por título y resumen.

Durante esta fase se excluyeron 38 registros por no satisfacer los criterios temáticos o metodológicos establecidos. Las 17 publicaciones restantes fueron evaluadas mediante lectura de texto completo. Dos estudios se excluyeron en esta etapa: uno correspondía a una versión preliminar posteriormente ampliada mediante una publicación con mayor desarrollo

metodológico y experimental, mientras que el otro no establecía una integración funcional entre el aprendizaje federado y los componentes, funciones o interfaces de O-RAN. La síntesis quedó conformada por 15 estudios. La Figura 1 presenta el flujo completo de selección.

Figura 1. *Diagrama de flujo del proceso de selección de los estudios conforme a PRISMA 2020.*



Nota. No se registraron publicaciones cuyo texto completo no pudiera recuperarse.

La distribución temporal se concentró entre 2024 y 2026. Se identificó una publicación correspondiente a 2023, cinco a 2024, cinco a 2025 y cuatro a 2026. En conjunto, 14 de los 15 estudios, equivalentes al 93,3 %, fueron publicados a partir de 2024.

Los entornos de evaluación también presentaron diferencias. Ocho investigaciones utilizaron exclusivamente simulaciones, cuatro incorporaron bancos de pruebas o plataformas de emulación, una combinó datos reales con un entorno simulado y dos desarrollaron propuestas conceptuales sin validación experimental completa.

Calidad metodológica

La puntuación media obtenida mediante la rúbrica de evaluación fue de 8,67 sobre 10. Doce estudios alcanzaron una valoración alta, dos fueron clasificados con calidad media y uno obtuvo una valoración baja.

Tabla 3. *Distribución de los estudios según la calidad metodológica*

Nivel de calidad	Intervalo	Número de estudios	Porcentaje
Alta	8,0–10,0	12	80,0 %
Media	6,0–7,5	2	13,3 %
Baja	< 6,0	1	6,7 %

La claridad del objetivo alcanzó una puntuación media de 1,00. La relación explícita con O-RAN y la definición de la anomalía o amenaza obtuvieron 0,97; la descripción de la arquitectura federada y de la distribución funcional de sus componentes alcanzaron 0,90; y la claridad de las métricas y la correspondencia entre resultados y objetivo obtuvieron 0,93. Los valores inferiores correspondieron a la descripción del método y escenario experimental, con 0,80, y al análisis de limitaciones o amenazas a la validez y la reproducibilidad, ambos con 0,63.

Las puntuaciones individuales oscilaron entre 5,5 y 10 puntos. Dos estudios alcanzaron la puntuación máxima, mientras que el valor mínimo correspondió a una propuesta predominantemente conceptual sin validación experimental completa.

Tipologías arquitectónicas

Las arquitecturas identificadas se distribuyeron en cuatro categorías: cliente-servidor, jerárquicas, descentralizadas entre pares y conceptuales o híbridas.

Tabla 4. *Características principales de los estudios incluidos*

Estudio	Arquitectura	Distribución funcional y agregación	Modelo	Amenaza	Datos / entorno
Attanayaka et al. (2023)	Descentralizada entre pares	xApps en Near-RT RIC; agregación segura entre pares	MLP	Intrusiones multiclase	UNSW-NB15; simulación
Benzaïd et al. (2024)	Jerárquica	xApps en Near-RT RIC; servidor rApp en Non-RT RIC; FedAvg/FedProx	MLP	Denegación de servicio	5G-NIDD
Abou El Houda et al. (2024)	Jerárquica multiagente	Agentes en Near-RT RIC; coordinación en Non-RT RIC	DQN / FMARL	Interferencia maliciosa	WSN-DS; simulación
Rumesh et al. (2024)	Jerárquica de dos niveles	Nodos E2; agregación Near-RT RIC y consolidación Non-RT RIC	LSTM	DDoS y consumo abusivo de ancho de banda	Colosseum; emulación
Ahakonye et al. (2024)	Cliente-servidor	Sensores distribuidos en componentes O-RAN; agregación no especificada	Clasificador no especificado	Intrusiones multiclase	Edge-IIoTset; simulación
Moudoud et al. (2024)	Jerárquica con confianza cero	Agentes Near-RT RIC; coordinación Non-RT RIC	FMARL	Intrusiones	WSN-DS; simulación
Mehrban et al. (2025)	Cliente-servidor multicapa	Dispositivos/componentes O-RAN; coordinación federada	RF, KNN, SVM, LightGBM y NN	Tráfico malicioso	TON_IoT; simulación
Albaseer et al. (2025)	Cliente-servidor	Agentes RIC y servidor federado; FedAvg	TabTransformer	Intrusiones multiclase	CIC-IDS2018; simulación
Shekhar et al. (2025)	Conceptual o híbrida	Componentes 5G/O-RAN descritos de forma general	FL, GNN, modelos temporales y DRL	Amenazas de seguridad 5G	Propuesta conceptual
Hung et al. (2025)	Cliente-servidor entre dominios	Near-RT RIC de dos redes privadas; consumidor Y1; FedAvg	Regresión logística	DoS de alta y baja tasa	OpenAirInterface y FlexRIC; dos redes privadas
Alalyan et al. (2025)	Descentralizada entre pares	Agentes/RIC distribuidos; promedio seguro y selección	Aprendizaje y transferencia	Ciberataques multiclase	UNSW-NB15 y banco de pruebas O-RAN
Chouikhi et al. (2026)	Cliente-servidor multi-RIC	RIC locales y coordinador federado; promedio ponderado	Detector ligero	DoS, xApp maliciosa y envenenamiento de políticas	Telemetría multi-RIC; simulación
Spantideas et al. (2026)	Totalmente descentralizada	dApps en O-DU/O-CU; cada nodo entrena y agrega FedAvg	DNN y LSTM	Anomalías de tráfico o carga	Datos sintéticos y tráfico móvil real
Ratnayake et al. (2026)	Cliente-servidor con confianza cero	O-CU/O-DU/Near-RT RIC; agregación y políticas en Non-RT RIC	Transformer	Intrusiones y tráfico malicioso	Cuatro conjuntos de datos; simulación
Dua et al. (2026)	Conceptual o híbrida	Capas de coordinación y orquestación; coordinación federada de confianza	Modelos temporales y XAI	Anomalías de políticas u orquestación	Arquitectura conceptual

Tabla 5. *Distribución de las arquitecturas de aprendizaje federado*

Arquitectura	Número de estudios	Porcentaje
Cliente-servidor	6	40,0 %
Jerárquica	4	26,7 %
Descentralizada entre pares	3	20,0 %
Conceptual o híbrida	2	13,3 %

Las configuraciones cliente-servidor concentraron seis estudios y utilizaron diferentes entidades de coordinación, entre ellas servidores federados, aplicaciones de gestión, consumidores Y1 y coordinadores multi-RIC. Las arquitecturas jerárquicas distribuyeron el aprendizaje principalmente entre controladores inteligentes de tiempo cercano al real y de tiempo no real, mientras que los tres diseños descentralizados prescindieron de un agregador único y trasladaron la combinación de actualizaciones hacia los propios participantes. Los dos marcos restantes presentaron configuraciones conceptuales o híbridas vinculadas con confianza cero, explicabilidad y orquestación autónoma.

Distribución funcional de participantes y agregadores

Los participantes federados se localizaron en diferentes componentes de la infraestructura. Entre las ubicaciones identificadas se encontraron controladores inteligentes de tiempo cercano al real, aplicaciones xApps, nodos E2, estaciones base, dispositivos conectados y aplicaciones descentralizadas desplegadas en unidades distribuidas o centralizadas.

En las arquitecturas jerárquicas, el entrenamiento local se situó principalmente en componentes próximos a la red de acceso, mientras que la agregación global se asignó al controlador inteligente de tiempo no real. Rumesh et al. (2024) incorporaron un nivel adicional de agregación en los controladores inteligentes de tiempo cercano al real antes de consolidar el modelo en el nivel superior.

En otras configuraciones, el agregador adoptó funciones específicas. Benzaïd et al. (2024) ubicaron el servidor federado como una rApp en el controlador inteligente de tiempo no real. Hung et al. (2025) emplearon un consumidor Y1 para coordinar modelos procedentes de dos redes privadas, mientras que Chouikhi et al. (2026) utilizaron un coordinador federado para integrar información procedente de múltiples controladores inteligentes. Ratnayake et al. (2026) asociaron la agregación y la decisión de políticas con el controlador inteligente de tiempo no real.

Attanayaka et al. (2023) y Alalyan et al. (2025) distribuyeron la combinación de actualizaciones entre los propios participantes mediante mecanismos de promedio seguro. Spantideas et al. (2026) extendieron esta distribución hacia aplicaciones descentralizadas alojadas en unidades distribuidas y centralizadas, de modo que cada nodo participó tanto en el entrenamiento como en la agregación.

Modelos de aprendizaje y algoritmos de agregación

Los modelos utilizados abarcaron aprendizaje automático convencional, aprendizaje profundo y aprendizaje por refuerzo. Attanayaka et al. (2023) y Benzaïd et al. (2024) emplearon perceptrones multicapa; Rumesh et al. (2024) utilizó redes de memoria a corto y largo plazo; Hung et al. (2025) recurrió a regresión logística; y Mehrban et al. (2025) evaluó bosques aleatorios, vecinos más cercanos, máquinas de vectores de soporte, LightGBM y redes neuronales.

Abou El Houda et al. (2024) y Moudoud et al. (2024) incorporaron redes Q profundas dentro de esquemas federados multiagente. Albaseer et al. (2025) utilizó TabTransformer junto con técnicas de aprendizaje con pocos ejemplos y verificación mediante un modelo de lenguaje. Ratnayake et al. (2026) empleó una arquitectura Transformer para la clasificación del tráfico.

FedAvg fue el algoritmo de agregación explícito más recurrente, con presencia en cinco estudios. Cuatro trabajos utilizaron promedios simples, ponderados o jerárquicos; dos aplicaron mecanismos de promedio seguro entre pares; y cuatro describieron procedimientos generales, parciales o conceptuales.

Tabla 6. *Mecanismos de agregación identificados*

Mecanismo de agregación	Número de estudios	Porcentaje
FedAvg	5	33,3 %
Promedio simple, ponderado o jerárquico	4	26,7 %
Promedio seguro entre pares	2	13,3 %
General, no especificado o conceptual	4	26,7 %

Benzaïd et al. (2024) compararon FedAvg y FedProx dentro de un esquema de aprendizaje federado continuo. Attanayaka et al. (2023) incorporaron cálculo seguro de promedios y una variante basada en cifrado homomórfico, mientras que Alalyan et al. (2025) combinaron promedio seguro, selección de participantes y transferencia de aprendizaje.

Mecanismos de seguridad y privacidad

La conservación de los datos en los participantes locales constituyó el mecanismo común a las implementaciones federadas. Sobre esta base se identificaron medidas adicionales orientadas a proteger la comunicación, la agregación, el acceso o la confianza entre componentes.

Attanayaka et al. (2023) evaluaron cálculo seguro de promedios y cifrado homomórfico. Moudoud et al. (2024) combinaron aprendizaje federado multiagente con una arquitectura de confianza cero y blockchain. Mehrban et al. (2025) integraron aprendizaje federado y transferencia de aprendizaje dentro de un esquema multicapa de confianza cero respaldado por blockchain. Spantideas et al. (2026) incorporaron canales cifrados, autenticación mutua, reputación y filtrado de actualizaciones.

Albaseer et al. (2025) añadieron verificación selectiva de pseudoetiquetas mediante un modelo de lenguaje, mientras que Ratnayake et al. (2026) combinaron aprendizaje federado, clasificación mediante Transformers y control de acceso basado en confianza cero. La privacidad diferencial fue mencionada como mecanismo complementario en algunas propuestas recientes, aunque su evaluación experimental no fue homogénea.

Anomalías y amenazas analizadas

La categoría predominante correspondió a la detección de intrusiones y tráfico malicioso multiclase, presente en ocho estudios. Cuatro investigaciones se orientaron principalmente a ataques de denegación de servicio. La interferencia maliciosa, las anomalías

de tráfico o carga y las anomalías relacionadas con políticas u orquestación aparecieron en un estudio cada una.

Tabla 7. *Distribución de las amenazas analizadas*

Categoría	Número de estudios	Porcentaje
Intrusiones y tráfico malicioso multiclase	8	53,3 %
Ataques de denegación de servicio	4	26,7 %
Interferencia maliciosa	1	6,7 %
Anomalías de tráfico o carga	1	6,7 %
Anomalías de políticas u orquestación	1	6,7 %

Las amenazas específicas incluyeron inundaciones ICMP, SYN y UDP, Slowloris, ataques distribuidos de denegación de servicio, escaneo de puertos, interferencia, comportamiento malicioso de aplicaciones xApps y manipulación de políticas. Algunos estudios abordaron múltiples categorías de ataques dentro de un mismo clasificador.

Conjuntos de datos y escenarios de evaluación

UNSW-NB15 y WSN-DS fueron los conjuntos de datos reutilizados con mayor frecuencia, con presencia en dos estudios cada uno. También se emplearon 5G-NIDD, Edge-IIoTset, TON_IoT y CIC-IDS 2018. Otros trabajos generaron información mediante infraestructuras o plataformas relacionadas con redes móviles.

Benzaïd et al. (2024) utilizaron 5G-NIDD para evaluar nueve tareas sucesivas de detección. Rumesh et al. (2024) emplearon trazas procesadas mediante Colosseum y tráfico de ataque generado para el escenario experimental. Hung et al. (2025) construyeron un conjunto de datos mediante dos redes privadas 5G implementadas con OpenAirInterface y FlexRIC. Alalyan et al. (2025) combinaron UNSW-NB15 con tráfico obtenido en un banco de pruebas 5G O-RAN.

Las distribuciones de datos variaron entre los estudios. Algunas evaluaciones consideraron particiones independientes e idénticamente distribuidas (IID), mientras que otras introdujeron condiciones no independientes e idénticamente distribuidas (no IID). Hung et al.

(2025), por ejemplo, distribuyeron los ataques de forma diferenciada entre dos redes privadas, de modo que cada dominio entrenó inicialmente con amenazas distintas.

Métricas y desempeño reportado

La exactitud, la precisión, la sensibilidad y la puntuación F1 fueron las métricas de clasificación más utilizadas. También se identificaron matrices de confusión, curvas ROC, tasas de falsos positivos, pérdida de entrenamiento y medidas relacionadas con transferencia de conocimiento.

Las métricas operativas estuvieron presentes en un número menor de trabajos e incluyeron tráfico de comunicación, latencia, consumo energético, tiempo de ejecución, convergencia, utilización de memoria y carga de procesamiento.

Rumesh et al. (2024) reportaron una exactitud máxima de 99,87 % en la configuración evaluada. Benzaïd et al. (2024) registraron una mejora de al menos 14,4 % respecto del aprendizaje federado convencional en su escenario de aprendizaje continuo y una conservación superior al 98,8 % del conocimiento adquirido previamente. Hung et al. (2025) obtuvieron aproximadamente 97,97 % y documentaron la detección de amenazas que no habían sido observadas durante el entrenamiento local de cada red.

Albaseer et al. (2025) reportaron una mejora de seis puntos porcentuales en exactitud, una reducción del 20 % en la propagación de errores y una disminución del 40 % en el consumo energético dentro del escenario con menor disponibilidad de etiquetas. Alalyan et al. (2025) mantuvieron valores de detección próximos al 99 % y registraron reducciones aproximadas del 74 % en el costo de comunicación durante la simulación y del 55 % en el banco de pruebas.

Spantideas et al. (2026) obtuvieron pérdidas de entrenamiento y prueba comparables a las de una configuración cliente-servidor. En el escenario de tres nodos, el esquema

descentralizado generó aproximadamente 48,77 MB de tráfico frente a 17,24 MB del diseño cliente-servidor.

Debido a las diferencias entre conjuntos de datos, modelos, número de participantes, clases de ataque y distribuciones de entrenamiento, no se calculó una medida agregada del desempeño predictivo.

Evolución temática de la literatura

El estudio de 2023 se centró en arquitecturas entre pares y mecanismos de agregación segura. Las publicaciones de 2024 incorporaron aprendizaje continuo, estructuras jerárquicas, gemelos digitales, aprendizaje federado multiagente y mecanismos de confianza cero.

Durante 2025 aparecieron propuestas que integraron aprendizaje con pocos ejemplos, verificación mediante modelos de lenguaje, colaboración entre redes privadas y arquitecturas multicapa de seguridad. Los trabajos publicados en 2026 incorporaron coordinación multi-RIC, aprendizaje completamente descentralizado entre aplicaciones, modelos Transformer y arquitecturas orientadas a la orquestación autónoma.

Discusión

Correspondencia entre las arquitecturas federadas y la organización funcional de O-RAN

Los resultados muestran que la aplicación del aprendizaje federado a la detección de anomalías en redes de acceso de radio abiertas no converge hacia una única configuración arquitectónica. Las propuestas se distribuyen entre modelos cliente-servidor, esquemas jerárquicos, diseños descentralizados entre pares y marcos híbridos, lo que evidencia que el problema no consiste únicamente en entrenar un modelo distribuido, sino en decidir dónde ubicar el aprendizaje, cómo coordinarlo y qué nivel de confianza asignar a cada participante.

La predominancia de configuraciones cliente-servidor y jerárquicas guarda una relación directa con la estructura funcional de O-RAN. La existencia de nodos E2, controladores inteligentes de tiempo cercano al real y controladores inteligentes de tiempo no real proporciona puntos definidos para distribuir entrenamiento, inferencia y agregación. Benzaïd et al. (2024), Rumesh et al. (2024) y Moudoud et al. (2024) aprovechan esta separación para mantener el procesamiento próximo a la fuente de datos y reservar la coordinación global para componentes con una visión más amplia del estado de la red. Esta organización resulta coherente con el funcionamiento multinivel de O-RAN, aunque introduce dependencia de los elementos que concentran la agregación.

Las arquitecturas descentralizadas modifican ese equilibrio. Attanayaka et al. (2023), Alalyan et al. (2025) y Spantideas et al. (2026) prescinden de un agregador único y distribuyen la combinación de modelos entre los participantes. El beneficio principal reside en reducir la dependencia de una entidad central y mejorar la continuidad del entrenamiento ante fallas individuales. Sin embargo, esta decisión incrementa la cantidad de intercambios y exige mecanismos adicionales para sincronizar actualizaciones, validar participantes y mantener coherencia entre versiones del modelo.

El comportamiento observado en FedRA ilustra con claridad este compromiso. Spantideas et al. (2026) obtuvieron pérdidas comparables a una arquitectura cliente-servidor, pero con un incremento considerable del volumen de comunicación. La descentralización, por tanto, no representa una mejora absoluta, sino una redistribución de los costos arquitectónicos: se reduce la dependencia central a cambio de mayores requerimientos de comunicación y coordinación.

En términos de diseño, ninguna de las familias identificadas puede considerarse óptima para todos los escenarios. Las arquitecturas centralizadas favorecen la simplicidad de coordinación; las jerárquicas aprovechan la estructura multinivel de O-RAN; y las descentralizadas priorizan autonomía y tolerancia a fallas. La elección debe responder al

número de dominios participantes, las restricciones temporales, la capacidad computacional disponible, el volumen de actualizaciones y el nivel de confianza existente entre las entidades que intervienen en el entrenamiento.

Distribución funcional del entrenamiento y la agregación

La ubicación de los participantes federados constituye una decisión determinante porque condiciona simultáneamente el acceso a los datos, la latencia y el costo de comunicación. La presencia recurrente de clientes en controladores inteligentes de tiempo cercano al real, aplicaciones xApps y nodos E2 indica una tendencia a mantener el procesamiento cerca de la fuente de telemetría. Esta distribución reduce la necesidad de transferir datos brutos y permite que la detección opere con información generada directamente dentro del dominio administrado por cada participante.

No obstante, entrenamiento, agregación e inferencia responden a escalas temporales distintas. La inferencia asociada con eventos sensibles a la latencia puede ejecutarse próxima al plano de acceso, mientras que la agregación global y la actualización de modelos pueden desplazarse hacia niveles con mayor capacidad computacional y restricciones temporales menos estrictas. Rumesh et al. (2024) materializan esta separación mediante una agregación intermedia en controladores de tiempo cercano al real y una consolidación posterior en el controlador de tiempo no real.

La propuesta de Hung et al. (2025) amplía este esquema hacia la colaboración entre dominios. El uso de un consumidor Y1 permitió coordinar dos redes privadas y detectar ataques que no habían sido observados durante el entrenamiento local. El valor de este resultado trasciende la mejora predictiva, ya que demuestra que el aprendizaje federado puede utilizarse para compartir conocimiento de seguridad entre redes sin intercambiar directamente sus flujos de tráfico.

A pesar de estos avances, el entrenamiento síncrono continúa dominando las evaluaciones. Este supuesto simplifica la coordinación experimental, pero resulta restrictivo para entornos donde los participantes presentan diferencias de capacidad, carga y conectividad. La incorporación de aprendizaje asíncrono, selección adaptativa de clientes y mecanismos tolerantes a actualizaciones retrasadas constituye, por ello, una extensión necesaria para aproximar las evaluaciones a condiciones operacionales.

Capacidad de detección y condiciones de validación

Los valores elevados de exactitud, precisión, sensibilidad y puntuación F1 reportados en varios estudios respaldan la capacidad del aprendizaje federado para mantener un desempeño competitivo sin centralizar los datos de entrenamiento. Sin embargo, estas métricas deben interpretarse dentro de las condiciones específicas de cada experimento.

Las diferencias en conjuntos de datos, número de clases, balance de muestras, distribución entre participantes y modelos utilizados impiden establecer comparaciones directas. Una arquitectura evaluada con clasificación binaria y datos IID no enfrenta las mismas condiciones que otra sometida a múltiples clases de ataque y particiones no IID. En consecuencia, las cifras de exactitud no pueden emplearse como criterio suficiente para ordenar las propuestas.

También existe una diferencia sustancial entre demostrar la eficacia del clasificador y validar la arquitectura O-RAN donde se pretende desplegar. UNSW-NB15, WSN-DS, TON_IoT y CIC-IDS 2018 proporcionan escenarios útiles para evaluar modelos de detección, pero no reproducen necesariamente la telemetría, las interfaces y las relaciones funcionales propias de O-RAN. Cuando estos conjuntos se emplean sin una infraestructura representativa, la evidencia obtenida sustenta principalmente el comportamiento del aprendizaje federado, no su viabilidad operacional dentro de la red de acceso.

Los experimentos desarrollados con OpenAirInterface, FlexRIC, Colosseum o bancos de pruebas específicos aportan una validación más cercana al escenario de interés. Hung et al. (2025), Rumesh et al. (2024) y Alalyan et al. (2025) incorporan elementos de este tipo y, por ello, ofrecen mayor información sobre la interacción entre distribución de datos, componentes O-RAN y coordinación federada.

La evaluación futura debería ampliar este enfoque mediante métricas de sistema. La latencia de extremo a extremo, el volumen de datos transmitidos, el tiempo de convergencia, la utilización de procesamiento, el consumo energético y la disponibilidad resultan tan relevantes como la capacidad de clasificación. Un modelo con elevada exactitud puede perder utilidad práctica si introduce una sobrecarga incompatible con los requisitos operacionales de la red.

Seguridad del proceso federado

La permanencia de los datos en los participantes reduce la exposición asociada con su centralización, pero no elimina los riesgos de seguridad. Las actualizaciones intercambiadas durante el entrenamiento pueden ser manipuladas o utilizadas para inferir información sobre los datos locales. De igual manera, un participante legítimo pero comprometido puede introducir actualizaciones destinadas a degradar el modelo global.

Los mecanismos identificados muestran una transición desde la simple localización de los datos hacia esquemas de protección más completos. Attanayaka et al. (2023) incorporaron cálculo seguro de promedios y cifrado homomórfico; Moudoud et al. (2024) y Mehrban et al. (2025) combinaron aprendizaje federado con confianza cero y blockchain; Spantideas et al. (2026) añadieron autenticación mutua, reputación y filtrado de actualizaciones. Estas aproximaciones cubren distintos componentes de la seguridad, aunque rara vez son evaluadas de manera conjunta.

La autenticación permite verificar la identidad de los participantes, pero no garantiza la legitimidad de sus actualizaciones. El cifrado protege la confidencialidad durante la transmisión, aunque no impide incorporar un modelo previamente manipulado. Del mismo modo, la descentralización elimina el riesgo asociado con un único agregador, pero distribuye la capacidad de influencia entre un mayor número de nodos.

Una arquitectura federada robusta requiere, por tanto, combinar mecanismos de autenticación, agregación resistente a valores atípicos, reputación dinámica, detección de comportamiento anómalo y procedimientos de recuperación. La escasa evaluación frente a envenenamiento de datos, manipulación de modelos, clientes bizantinos y colusión constituye una de las principales brechas identificadas.

La incorporación de modelos de lenguaje para verificar pseudoetiquetas, como en Albaseer et al. (2025), introduce una línea distinta de protección orientada a reducir errores derivados del etiquetado limitado. Sin embargo, esta estrategia incorpora nuevas dependencias relacionadas con latencia, consumo de recursos y confiabilidad del verificador. Su utilidad debe evaluarse junto con estos costos y no únicamente mediante la mejora de la clasificación.

Madurez de la evidencia y reproducibilidad

La calidad metodológica obtenida por la mayoría de los estudios demuestra que las propuestas suelen presentar objetivos claros y una relación coherente entre diseño, experimentación y resultados. Este nivel de calidad, sin embargo, no equivale a madurez tecnológica. La predominancia de simulaciones indica que buena parte de la evidencia permanece situada en una fase de validación experimental.

Los despliegues reales incorporan condiciones que rara vez aparecen de manera simultánea en los estudios revisados: heterogeneidad de hardware, variaciones de carga, pérdidas de conectividad, tráfico cifrado, múltiples dominios administrativos y restricciones de

recursos. La respuesta de una arquitectura federada frente a estos factores puede diferir de la observada en entornos controlados.

La reproducibilidad constituye una limitación adicional. Los criterios relacionados con amenazas a la validez y disponibilidad de información reproducible obtuvieron las puntuaciones más bajas de la evaluación metodológica. La ausencia de código, configuraciones, semillas, esquemas de partición o parámetros de comunicación dificulta comprobar los resultados y determinar qué parte de la mejora deriva del modelo, del algoritmo de agregación o de la propia arquitectura.

La disponibilidad de artefactos experimentales permitiría establecer comparaciones más consistentes. Un entorno reproducible debería incluir modelos base, configuraciones de participantes, particiones de datos, scripts de ataque, parámetros de entrenamiento y perfiles de comunicación. Esta práctica facilitaría distinguir los efectos atribuibles a la topología federada de aquellos asociados con decisiones experimentales particulares.

Evolución de las propuestas y brechas de investigación

La evolución temporal muestra una ampliación progresiva del alcance de las arquitecturas. Las primeras propuestas se concentraron en el intercambio entre pares y la protección de la agregación. Posteriormente aparecieron aprendizaje continuo, gemelos digitales, aprendizaje federado multiagente y confianza cero. Los trabajos más recientes incorporan aprendizaje con pocos ejemplos, modelos de lenguaje, coordinación multi-RIC, Transformers y esquemas completamente descentralizados.

Esta tendencia indica que el aprendizaje federado comienza a integrarse como parte de sistemas de seguridad y automatización más amplios. No obstante, el aumento de complejidad incrementa también el número de componentes susceptibles de falla o manipulación. La incorporación de modelos más avanzados debe acompañarse de

evaluaciones que consideren no solo su capacidad predictiva, sino también su impacto sobre la operación y la seguridad global del sistema.

La robustez adversarial emerge como una prioridad inmediata. La comparación de FedAvg con mecanismos resistentes a valores extremos o clientes maliciosos permitiría determinar hasta qué punto la precisión puede mantenerse frente a ataques al proceso de agregación. Del mismo modo, los escenarios futuros deberían incluir colusión, envenenamiento de modelos y compromiso del coordinador, en lugar de asumir participantes completamente confiables.

La heterogeneidad entre dominios plantea una segunda prioridad. Los participantes pueden administrar celdas, usuarios y servicios con distribuciones distintas, por lo que un único modelo global puede no representar adecuadamente las condiciones locales. El aprendizaje personalizado, continuo y asíncrono ofrece alternativas para adaptar el conocimiento global sin eliminar las particularidades de cada dominio.

Una tercera brecha corresponde a la ausencia de un benchmark específico para O-RAN. La utilización de conjuntos heterogéneos impide comparar las propuestas bajo condiciones equivalentes. Un entorno de referencia debería integrar telemetría de red, indicadores de desempeño, eventos de control, tráfico de usuario y múltiples categorías de amenaza, además de proporcionar particiones IID y no IID.

La evaluación debería incorporar simultáneamente métricas predictivas, operativas y de seguridad. Exactitud, sensibilidad y puntuación F1 tendrían que analizarse junto con latencia, tráfico de comunicación, consumo energético, robustez adversarial y pérdida de privacidad. Este enfoque permitiría valorar las arquitecturas como sistemas de seguridad distribuidos y no únicamente como clasificadores.

Los hallazgos permiten establecer que el aprendizaje federado se adapta conceptualmente a la distribución funcional de O-RAN, pero su integración introduce

compromisos que no pueden resolverse mediante una única configuración. La centralización simplifica la coordinación, la jerarquización aprovecha la organización multinivel de la arquitectura y la descentralización incrementa la autonomía, aunque ninguna alternativa satisface simultáneamente todos los requisitos de desempeño, seguridad y eficiencia.

La literatura actual demuestra principalmente viabilidad experimental. La transición hacia soluciones operacionales exige validar las arquitecturas bajo condiciones multioperador, distribuciones heterogéneas, participantes adversariales y restricciones reales de comunicación. Asimismo, será necesario disponer de benchmarks y artefactos reproducibles que permitan comparar las propuestas bajo condiciones equivalentes.

Conclusiones

La revisión evidencia que el aprendizaje federado constituye una alternativa arquitectónicamente compatible con las redes de acceso de radio abiertas, debido a que permite distribuir el entrenamiento entre componentes próximos a las fuentes de datos y compartir conocimiento sin centralizar los datos brutos. No obstante, su incorporación no responde a una configuración única. Las arquitecturas cliente-servidor, jerárquicas y descentralizadas representan diferentes compromisos entre coordinación, autonomía, resiliencia y costo de comunicación; por ello, su pertinencia debe establecerse en función de las condiciones del despliegue y no únicamente del desempeño predictivo alcanzado.

La correspondencia entre las funciones federadas y la estructura multinivel de O-RAN favorece especialmente las configuraciones jerárquicas, al permitir separar el procesamiento local de la coordinación global. Los esquemas descentralizados ofrecen una vía para reducir la dependencia de un agregador único, aunque trasladan parte de la complejidad hacia la sincronización, el consenso y la validación de las actualizaciones. En consecuencia, la descentralización no constituye por sí misma un criterio de superioridad arquitectónica; su conveniencia depende del equilibrio requerido entre disponibilidad, latencia, escalabilidad y sobrecarga de comunicación.

Los resultados también muestran que mantener los datos en los participantes es insuficiente para garantizar la seguridad del proceso federado. La integridad de las actualizaciones, la resistencia frente a participantes comprometidos y la protección contra ataques de inferencia o envenenamiento deben incorporarse como requisitos de diseño. Esto adquiere especial relevancia en O-RAN, donde pueden coexistir componentes, aplicaciones y dominios administrativos pertenecientes a diferentes actores. Una arquitectura desplegable deberá combinar mecanismos de autenticación, agregación robusta, protección criptográfica, evaluación de confianza y recuperación ante comportamientos adversariales.

La principal limitación para trasladar estas propuestas hacia entornos operacionales no reside en la capacidad de los modelos para alcanzar altos niveles de detección, sino en la limitada representatividad de las condiciones bajo las cuales han sido evaluados. La dependencia de simulaciones, la utilización de conjuntos de datos no específicos de O-RAN, la escasez de escenarios multioperador y la evaluación insuficiente frente a participantes maliciosos restringen la generalización de los resultados. Asimismo, la ausencia de configuraciones experimentales comunes dificulta determinar qué mejoras corresponden realmente a la topología federada, al algoritmo de agregación o al modelo de aprendizaje utilizado.

El desarrollo futuro requiere bancos de pruebas reproducibles y conjuntos de datos que representen telemetría, eventos de control, tráfico y amenazas propias de O-RAN, junto con distribuciones heterogéneas entre participantes. Las evaluaciones deberían integrar métricas predictivas con latencia, comunicación, consumo de recursos, privacidad y robustez adversarial. Bajo estas condiciones será posible establecer con mayor precisión qué arquitecturas de aprendizaje federado proporcionan un equilibrio operacionalmente viable entre inteligencia distribuida, seguridad, eficiencia y resiliencia.

Referencias bibliográficas

Abou El Houda, Z., Moudoud, H., & Brik, B. (2024). Federated deep reinforcement learning for efficient jamming attack mitigation in O-RAN. IEEE Transactions on Vehicular Technology, 73(7), 9334–9343. https://doi.org/10.1109/TVT.2024.3359998

Ahakonye, L. A. C., Nwakanma, C. I., & Kim, D.-S. (2024). Roadmap for integrating Open RAN and AI for intrusion detection systems in 6G networks. In 2024 15th International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1586–1591). IEEE. https://doi.org/10.1109/ICTC62082.2024.10826894

Alalyan, F., Awad, M., Jaafar, W., & Langar, R. (2025). Secure distributed federated learning for cyberattacks detection in B5G open radio access networks. IEEE Open Journal of the Communications Society, 6, 3067–3081. https://doi.org/10.1109/OJCOMS.2024.3523468

Albaseer, A., Hamood, M., Al-Sabri, R., Abdallah, M., & Al-Fuqaha, A. (2025). A zero-touch O-RAN framework for federated few-shot IDS with LLM-oracle verification. In GLOBECOM 2025 – 2025 IEEE Global Communications Conference (pp. 3170–3175). IEEE. https://doi.org/10.1109/GLOBECOM59602.2025.11432014

Attanayaka, D., Porambage, P., Liyanage, M., & Ylianttila, M. (2023). Peer-to-peer federated learning based anomaly detection for open radio access networks. In ICC 2023 – IEEE International Conference on Communications (pp. 5464–5470). IEEE. https://doi.org/10.1109/ICC45041.2023.10278993

Benzaïd, C., Hossain, F. M., Taleb, T., Gómez, P. M., & Dieudonne, M. (2024). A federated continual learning framework for sustainable network anomaly detection in O-RAN. In 2024 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1–6). IEEE. https://doi.org/10.1109/WCNC57260.2024.10570951

Chouikhi, S., Khoukhi, L., Abou El Houda, Z., & Descamps, P. (2026). Federated multi-RIC threat intelligence for secure and scalable open 6G radio access networks. In 2026 IEEE Wireless Communications and Networking Conference Workshops (WCNCW) (pp. 1–5). IEEE. https://doi.org/10.1109/WCNCW67598.2026.11555534

Dua, M., Kundu, T., & Gupta, A. (2026). Trustworthy autonomous RAN orchestration: An architectural framework for zero-trust intent-based control with explainable AI. In 2026 35th Wireless and Optical Communications Conference (WOCC) (pp. 1–6). IEEE. https://doi.org/10.1109/WOCC69802.2026.11556178

Hung, C.-F., Kuo, C.-C., & Cheng, S.-M. (2025). Low-rate denial-of-service attack detection in open radio access network: Integrating federated learning. IEEE Vehicular Technology Magazine, 20(4), 75–83. https://doi.org/10.1109/MVT.2025.3623298

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K. A., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (Vol. 54, pp. 1273–1282). Proceedings of Machine Learning Research. https://proceedings.mlr.press/v54/mcmahan17a.html

Mehrban, A., Abou El Houda, Z., Moudoud, H., Brik, B., & Khoukhi, L. (2025). A blockchain-enabled multi-layered zero-trust security framework for O-RAN. In 2025 International Wireless Communications and Mobile Computing (IWCMC) (pp. 1564–1569). IEEE. https://doi.org/10.1109/IWCMC65282.2025.11059720

Moudoud, H., Abou El Houda, Z., & Brik, B. (2024). Zero trust security architecture for 6G open radio access networks (ORAN). IEEE Networking Letters, 6(4), 272–275. https://doi.org/10.1109/LNET.2024.3514357

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Polese, M., Bonati, L., D'Oro, S., Basagni, S., & Melodia, T. (2023). Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges. IEEE Communications Surveys & Tutorials, 25(2), 1376–1411. https://doi.org/10.1109/COMST.2023.3239220

Ratnayake, H., Siriwardhana, Y., Wijewardhana, U., & Liyanage, M. (2026). FORTRESS: A federated learning and transformer-based zero trust framework for O-RAN security. In 2026 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit) (pp. 1091–1098). IEEE. https://doi.org/10.1109/EuCNC/6GSummit68295.2026.11577317

Rumesh, Y., Attanayaka, D., Porambage, P., Pinola, J., Groen, J., & Chowdhury, K. (2024). Federated learning for anomaly detection in Open RAN: Security architecture within a digital twin. In 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit) (pp. 877–882). IEEE. https://doi.org/10.1109/EuCNC/6GSummit60053.2024.10597083

Shekhar, Bhardwaj, V., Singh, R., Essa, I. M., Gharban, H. A. J., & Shukla, P. K. (2025). AI-driven security frameworks for 5G wireless communication networks. In 2025 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN) (pp. 293–298). IEEE. https://doi.org/10.1109/EmergIN67762.2025.11450600

Spantideas, S. T., Giannopoulos, A. E., Levis, G. A., & Trakadas, P. (2026). FedRA: A fully decentralized federated learning framework for robust intelligence sharing across O-RAN dApps. IEEE Communications Magazine. Advance online publication. https://doi.org/10.1109/MCOM.001.2500368.