Application of unsupervised machine learning techniques for autonomous financial fraud detection in decentralized blockchain environments: a systematic review

Porras Poma, Nelson[1], Mendoza Ramos, Erick Angel[1], Huamán Aguirre, Arnold Anthony[1], and Zárate Segura, Guilermo Wenceslao[1]
[1]*Universidad Tecnológica del Perú, Perú, U21315674@utp.edu.pe, U21202514@utp.edu.pe, C19532@utp.edu.pe, E15040@utp.edu.pe*

Abstract*– The rapid growth of decentralized blockchain-based financial systems has introduced new challenges in fraud detection, particularly due to anonymity, scalability, and the dynamic nature of transactions. Traditional supervised learning approaches often prove insufficient in these environments due to their reliance on labeled data and fixed patterns. This systematic literature review (SLR) investigates the application of unsupervised machine learning techniques for autonomous fraud detection in decentralized blockchain environments, analyzing their effectiveness, adaptability, and limitations compared to supervised or semi-supervised methods. Using a rigorous methodology based on the PICOC framework and PRISMA guidelines, we reviewed 20 studies published between 2021 and 2025. The results reveal that clustering and anomaly detection algorithms (such as autoencoders and graph-based methods) achieve superior performance (85–99% accuracy) in identifying common frauds like phishing and Ponzi schemes, leveraging blockchain's transparency and immutability. Key metrics such as recall (90–95%) and F1-score (88–93%) prove critical for evaluating models on imbalanced datasets. However, challenges persist in scalability, privacy, and cross-protocol adaptability. This study contributes a taxonomy of unsupervised techniques applied to blockchain fraud detection and proposes future research directions, such as hybrid models and standardized evaluation frameworks for decentralized ecosystems.

Keywords*-- Unsupervised learning, fraud detection, blockchain, decentralized finance, anomaly detection, clustering, systematic review, autonomous systems..

Aplicación de técnicas de machine learning no supervisado para la detección autónoma de fraudes financieros en entornos blockchain descentralizados: una revisión sistemática

Porras Poma, Nelson[1], Mendoza Ramos, Erick Angel[1], Huamán Aguirre, Arnold Anthony[1], and Zárate Segura, Guilermo Wenceslao[1]
[1]*Universidad Tecnológica del Perú, Perú, U21315674@utp.edu.pe, U21202514@utp.edu.pe, C19532@utp.edu.pe, E15040@utp.edu.pe*

***Resumen*– Esta revisión sistemática analiza la aplicación de técnicas de machine learning no supervisado para la detección autónoma de fraudes financieros en entornos blockchain descentralizados, evaluando su eficacia frente a métodos supervisados. Los resultados muestran que algoritmos como clustering y detección de anomalías logran alta precisión en identificar fraudes como phishing y esquemas Ponzi, aprovechando la transparencia e inmutabilidad de blockchain. Sin embargo, persisten desafíos en escalabilidad, privacidad y adaptabilidad entre protocolos. Las métricas clave son recall (90–95%) y *F1-score* (88–93%), cruciales en conjuntos de datos desbalanceados. Se propone una taxonomía de técnicas no supervisadas y futuras líneas de investigación, como modelos híbridos y marcos de evaluación estandarizados, para mejorar la detección en ecosistemas descentralizados.**

***Palabras clave*-- Aprendizaje no supervisado, detección de fraude, blockchain, finanzas descentralizadas, detección de anomalías, agrupamiento, revisión sistemática, sistemas autónomos.**

I. Introducción

En la última década, la inteligencia artificial (IA) se ha posicionado como una herramienta clave para la transformación del sistema financiero global. Su aplicación ha permitido mejorar la eficiencia operativa, automatizar procesos complejos y fortalecer los mecanismos de seguridad, particularmente en tareas como la detección de fraudes, la evaluación de riesgos y la optimización de servicios personalizados[1], [2]. Al mismo tiempo, la expansión de tecnologías descentralizadas como blockchain ha redefinido los entornos financieros, promoviendo estructuras más transparentes, resistentes a la manipulación y sin intermediarios tradicionales. Sin embargo, esta descentralización ha traído consigo nuevos desafíos en cuanto a la supervisión, privacidad y seguridad de las transacciones[3]. Aunque se han desarrollado propuestas que combinan IA y blockchain, la mayoría se basa en modelos supervisados, lo que limita su capacidad de adaptación y respuesta ante fraudes emergentes en sistemas dinámicos y distribuidos[4].

Pese a los avances en la integración de IA en fintech, la detección de fraudes en entornos blockchain continúa siendo un reto no resuelto. Las soluciones actuales dependen, en su mayoría, de grandes volúmenes de datos etiquetados y de estructuras de decisión predefinidas, lo que impide una respuesta ágil ante patrones fraudulentos inéditos o cambiantes[3], [4]. Estas limitaciones se acentúan en contextos descentralizados donde el anonimato, la ausencia de autoridades centrales y la velocidad de las transacciones obstaculizan la eficacia de enfoques tradicionales. Así, persiste una brecha metodológica crítica: la escasa exploración del machine learning no supervisado como solución adaptable, autónoma y escalable para la detección de anomalías financieras en redes blockchain[1].

Esta investigación se justifica tanto por su relevancia académica como por su impacto potencial en la sociedad. En el plano científico, existe una necesidad crítica de sistematizar el conocimiento disperso sobre la aplicación de técnicas de machine learning en la detección de fraudes dentro de entornos blockchain descentralizados, particularmente en lo que respecta a modelos adaptativos, no supervisados y autónomos. Aunque existen revisiones previas sobre IA en finanzas[1] y sobre regulación en el contexto fintech[5], estas no abordan de forma específica la convergencia entre machine learning y tecnologías distribuidas, ni ofrecen una taxonomía que permita ordenar y evaluar los enfoques existentes. Y desde una perspectiva social, mejorar los sistemas de detección de fraude en plataformas blockchain contribuiría directamente a reforzar la confianza en el ecosistema financiero digital, proteger a los usuarios ante delitos cibernéticos y fomentar un entorno de innovación segura en mercados emergentes. Por ello, se vuelve imperativo realizar una revisión sistemática orientada a mapear, clasificar y analizar los enfoques existentes, con el fin de identificar vacíos de conocimiento, aportar claridad metodológica y facilitar el desarrollo de soluciones más resilientes, autónomas y eficientes en la protección de los sistemas financieros descentralizados.

El objetivo principal de esta revisión sistemática de la literatura (RSL) es analizar el uso de técnicas de machine learning no supervisado para la detección autónoma de fraudes financieros en entornos blockchain descentralizados, con el fin de evaluar su eficacia, adaptabilidad y limitaciones frente a métodos supervisados o semisupervisados. A través de un

enfoque metodológico riguroso basado en el protocolo PICOC [6], [7] y PRISMA, se busca identificar los algoritmos no supervisados más efectivos (como clustering o detección de anomalías), las métricas de validación utilizadas (precisión, recall, F1-score) y los desafíos técnicos asociados a la privacidad, escalabilidad y variabilidad de datos en blockchain. Asimismo, esta revisión pretende sintetizar el conocimiento existente sobre cómo estas técnicas aprovechan características inherentes de las blockchains para operar sin intervención humana, así como proponer direcciones futuras para investigaciones en modelos híbridos o adaptativos que mejoren la detección temprana de fraudes en sistemas descentralizados.

Esta revisión sistemática se estructura de la siguiente manera. La Sección 2, Metodología donde se detalla el protocolo PICOC empleado para formular las preguntas de investigación, la estrategia de búsqueda en bases de datos como Scopus y los criterios PRISMA para la selección y evaluación de los 20 artículos incluidos. La Sección 3, Resultados que presenta un análisis cuantitativo y cualitativo de las técnicas no supervisadas aplicadas como autoencoders y algoritmos basados en grafos, los tipos de fraudes más recurrentes como el phishing y esquemas Ponzi y los retos técnicos como el anonimato y escalabilidad. La Sección 4, Discusión que contrasta los hallazgos con la literatura existente, analiza las ventajas de los enfoques no supervisados frente a los supervisados y explora limitaciones como la alta tasa de falsos positivos. Finalmente, la Sección 5, Conclusiones sintetiza las contribuciones clave, resalta la efectividad de métricas como recall y F1-score en entornos descentralizados, y propone líneas futuras, como el desarrollo de modelos híbridos y la estandarización de métricas adaptadas a blockchain. Esta organización garantiza una transición lógica desde los fundamentos metodológicos hasta las implicaciones prácticas y oportunidades de investigación futura.

II. Metodología

Se empleó la metodología PICOC (Población/Problema, Intervención, Comparación, Resultados y Contexto) para estructurar las búsquedas de información de forma eficiente y precisa, alineando las palabras clave con las preguntas de investigación. Este protocolo permitió identificar datos relevantes, optimizando el proceso de investigación y facilitando la toma de decisiones informadas.

Se minimizó el sesgo y aumentó el rigor metodológico. Entre las categorías analizadas se incluyeron los criterios PICOC, así como aspectos relacionados con la calidad evaluativa y la fundamentación teórica del aprendizaje[6], [7].

A. Preguntas de investigación.

Para guiar esta revisión sistemática, se formularon seis preguntas de investigación para extraer, analizar y sintetizar el conocimiento sobre la aplicación del machine learning no supervisado en la detección de fraudes financieros en blockchain. Estas preguntas se detallan en la Tabla I.

B. Estrategia de búsqueda.

Para garantizar una revisión exhaustiva de los estudios relevantes, se empleó la metodología PICOC, centrándose en el problema, la intervención, la comparación y el resultado (Tabla II).

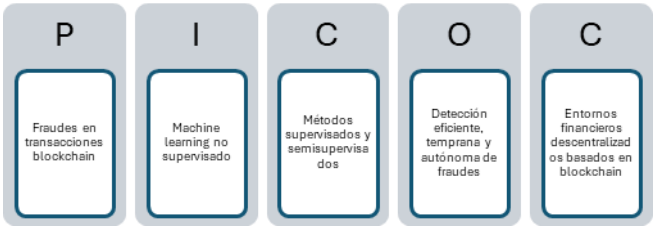

Fig. 1 Marco PICOC

TABLA I PREGUNTAS DE INVESTIGACIÓN

Cod.	Preguntas
Principal	¿Cómo pueden las técnicas de machine learning no supervisado mejorar la detección autónoma de fraudes financieros en entornos blockchain descentralizados en comparación con métodos supervisados o semisupervisados, considerando métricas de eficacia como en adaptabilidad a patrones emergentes?
P	¿Cuáles son los patrones de fraude más comunes en transacciones blockchain?
I	¿Qué algoritmos no supervisados se han aplicado para identificar fraudes en blockchain y bajo qué condiciones metodológicas?
C	¿Qué ventajas y limitaciones tienen los métodos no supervisados frente a los supervisados y semisupervisados en términos de precisión, escalabilidad y adaptabilidad a nuevos patrones de fraude?
O	¿Cómo se valida la eficiencia de los modelos no supervisados en la detección temprana y autónoma de fraudes en sistemas descentralizados?
C	¿En qué aspectos los entornos financieros descentralizados basados en blockchain requieren el uso de ML no supervisado para una detección de fraudes efectiva?

TABLA II ESTRATÉGIA DE BÚSQUEDA

Factor	Palabras claves
Problema	Fraud OR blockchain OR "distributed ledger" OR "smart contract"
Intervención	"Unsupervised learning" OR "anomaly detection" OR "clustering techniques" OR "machine learning" OR "artificial intelligence" OR AI OR "deep learning"
Comparación	“Supervised models” OR “Semi-supervised learning” OR “Labeled data”
Resultados	Finance OR fintech OR banking OR "financial system" OR "digital payments" OR cryptocurrency OR "financial technology”
Contexto	Security OR detection OR prevention OR anomalies OR risk

Ecuación de búsqueda:

La siguiente ecuación de búsqueda se utilizó en las bases de datos Scopus.

(fraud AND (blockchain OR "distributed ledger" OR "smart contract") AND ("unsupervised learning" OR "anomaly detection" OR "clustering techniques" OR "machine learning" OR "artificial intelligence" OR AI OR "deep learning") AND (finance OR fintech OR banking OR "financial system" OR "digital payments" OR cryptocurrency OR "financial technology") AND (security OR detection OR prevention OR anomalies OR risk))

La metodología PRISMA se adoptó con el propósito de mejorar la precisión en la búsqueda dentro de las revisiones sistemáticas de la literatura (RSL). La Declaración PRISMA, presentada en 2009, tiene como objetivo facilitar a los investigadores la redacción de informes detallados sobre sus revisiones sistemáticas, permitiéndoles expresar con claridad los motivos que los llevaron a realizarlas, los procedimientos utilizados y los resultados obtenidos[8].

El proceso del diagrama PRISMA se dividió en tres etapas: Identificación, Selección e Inclusión.

Criterios de inclusión (CI):

• CI 1: Artículos que aborden técnicas de machine learning no supervisado o híbrido para detección de fraudes en blockchain.

• CI 2: Investigaciones que apliquen métodos específicos de machine learning no supervisado en contextos blockchain descentralizados.

Criterios de exclusión (CE):

• CE 1: Investigaciones sobre banca tradicional sin blockchain ni machine learning.

En la fase de identificación, se localizaron 192 registros en la base de datos de SCOPUS. Tras eliminar 158 registros: 1 por ser mayor a 5 años, 35 que no corresponden al área de ingeniería ni de Ciencias de la computación, 113 que no son artículos y 2 que no están en el idioma inglés, quedaron 41 registros para la siguiente etapa. En la fase de cribado, se excluyeron 20 registros por no ser de acceso abierto, por lo que se evaluaron 21 informes para determinar su elegibilidad.

Durante esta evaluación, se eliminó 1 informe por la siguiente razón: solo abarca banca tradicional, con machine learning no blockchain (C.E 1). Finalmente, se incluyeron 20 estudios en la revisión. El proceso se detalla gráficamente, mostrando cómo se identificaron, filtraron y seleccionaron los estudios relevantes para la revisión.

También priorizamos la selección de artículos de libre acceso, haciéndolos accesibles a un público más amplio. Siguiendo estos criterios claros, la búsqueda produjo estudios que no solo fueron muy relevantes para nuestras preguntas de investigación, sino que también cumplieron con estrictos estándares de calidad. Este enfoque sistemático aumentó considerablemente la fiabilidad e integridad de la investigación, lo que en última instancia mejoró la credibilidad de nuestros hallazgos.

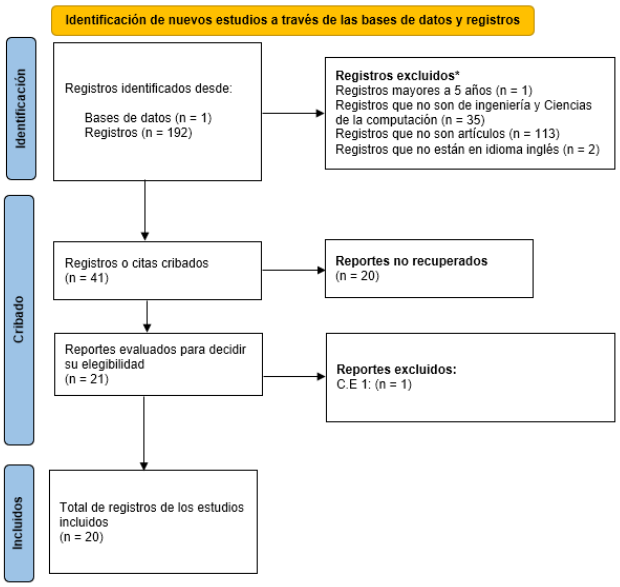

Fig. 2 Diagrama de flujo PRISMA

III. Resultados

Como se puede visualizar en la Fig. 3 se seleccionaron y analizaron 20 artículos publicados en los últimos cinco años que abordan la aplicación de técnicas de machine learning no supervisado para la detección autónoma de fraudes financieros en entornos blockchain descentralizados. De ellos, 8 corresponden al año 2024, lo que lo convierte en el periodo con mayor número de publicaciones seleccionadas.

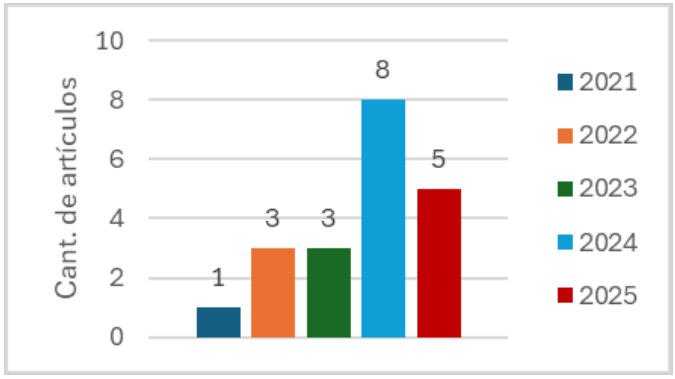

Fig. 3 Resultados del número de artículos sobre machine learning no supervisado para la detección autónoma de fraudes financieros en entornos blockchain descentralizados.

- ***RQ1: ¿Qué tipos de fraudes son más recurrentes en blockchain?***

La Fig. 4 muestra los fraudes más recurrentes en entornos blockchain, según la literatura analizada, son el phishing, mencionado en 8 artículos, y los esquemas Ponzi, mencionado en 7 artículos, seguidos por el lavado de dinero, mencionado 6 artículos. El phishing destaca por su prevalencia en

plataformas como Ethereum, donde actores maliciosos engañan a los usuarios para obtener credenciales o fondos [9], [10]. Los esquemas Ponzi, comunes en contratos inteligentes, se basan en promesas de altos rendimientos financieros sustentados por nuevos inversores [11], [12].

Otros fraudes significativos incluyen el pump and dump (manipulación de precios en criptomonedas) y el lavado de dinero, aprovechando el anonimato de blockchain [13], [14]. También se mencionan ataques específicos como el 51% attack (control de la mayoría del poder de minería) [15] y el greenwashing en proyectos de inversión sostenible [16]. La diversidad de fraudes refleja los desafíos en la detección autónoma, especialmente en sistemas descentralizados donde la ausencia de intermediarios dificulta la supervisión tradicional [17], [18].

Finalmente, se observa que ciertas modalidades, como el robo de identidad y el cryptojacking, aunque menos mencionadas, representan riesgos crecientes en el ecosistema financiero descentralizado [19], [20]. Esta variabilidad exige enfoques de machine learning no supervisado capaces de adaptarse a patrones emergentes sin dependencia de datos etiquetados [13], [15].

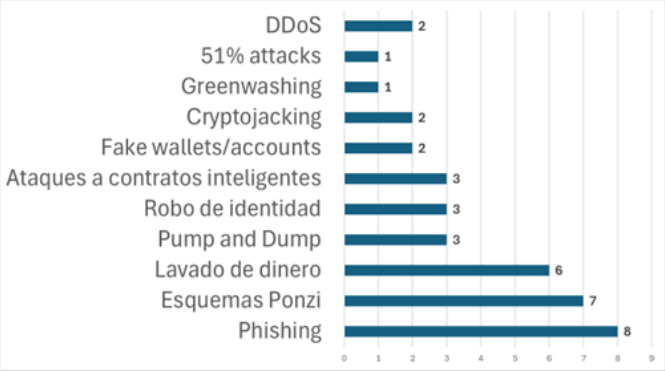

Fig. 4 Fraudes más recurrentes en Blockchain

- ***RQ2: ¿Qué características tienen estos fraudes que dificultan su detección automática?***

La Tabla III muestra que la principal característica que dificulta la detección automática de fraudes en blockchain es el anonimato o seudonimidad, que es mencionado en 14 artículos, y permite a los actores maliciosos operar sin identificación clara, complicando el rastreo de actividades fraudulentas [10], [13], [21]. Además, los patrones cambiantes y evolutivos, señalados en 10 artículos, hacen que los fraudes adapten sus métodos rápidamente, evadiendo sistemas de detección estáticos [11], [17], [22]. La falta de datos etiquetados, referidos en 9 artículos, también representa un desafío crítico, ya que limita el entrenamiento de modelos supervisados y exige enfoques no supervisados más flexibles [21], [23].

Otra dificultad clave es que muchos fraudes imitan comportamientos legítimos, referidos en 6 artículos, como transacciones normales o contratos inteligentes aparentemente válidos, lo que reduce la eficacia de métodos basados en reglas fijas [13], [24]. La descentralización, mencionado en 8 artículos agrava el problema, ya que elimina puntos centralizados de monitoreo y validación [14], [15]. Asimismo, la velocidad de las transacciones en blockchain, indicados en 3 artículos, dificulta la intervención en tiempo real, permitiendo que actividades fraudulentas se completen antes de ser detectadas [19], [20].

Finalmente, la complejidad de los patrones, mencionada en 4 artículos, y el uso de múltiples direcciones, señalado en 3 artículos, para ocultar flujos ilícitos aumentan la dificultad de identificar anomalías [12], [18]. Estas características resaltan la necesidad de técnicas de ML no supervisado capaces de detectar fraudes sin depender de datos previamente etiquetados o estructuras jerárquicas de supervisión [13], [15].

TABLA III CARACTERÍSTICAS DE FRAUDES QUE DIFICULTAN SU DETECCIÓN AUTOMÁTICA

Referencia	Característica
[10], [13], [21]	Anonimato / Seudonimidad
[11], [17], [22]	Patrones cambiantes / Evolutivos
[21], [23]	Falta de datos etiquetados
[13], [24]	Comportamiento similar a transacciones legítimas
[14], [15]	Descentralización / Ausencia de supervisión centralizada
[19], [20]	Velocidad de transacciones
[12], [18]	Datos heterogéneos / Complejidad de patrones
[13], [15]	Uso de múltiples direcciones / Técnicas de ocultamiento

- ***RQ3: ¿Qué técnicas de machine learning no supervisado se han utilizado?***

La Fig. 5 muestra que las técnicas de clustering (6 artículos) son las más utilizadas en la detección no supervisada de fraudes en blockchain, permitiendo agrupar transacciones o nodos con comportamientos sospechosos sin necesidad de datos etiquetados [12], [13]. Le siguen en relevancia los algoritmos de detección de anomalías, como Local Outlier Factor (LOF) e Isolation Forest (5 artículos), que identifican transacciones atípicas en conjuntos de datos no etiquetados [9], [15]. Estos métodos son especialmente útiles en entornos descentralizados donde los fraudes no siguen patrones predefinidos [13], [24].

Entre las técnicas avanzadas, destacan los autoencoders (incluyendo Graph Autoencoders [10]) y los métodos basados en grafos (DeepWalk y TransWalk [25]), que analizan relaciones complejas entre nodos y transacciones en redes blockchain. También se mencionan enfoques híbridos, como el One-Class SVM (2 artículos) para clasificación de anomalías [14], [15], y el uso de Mahalanobis Distance para medir desviaciones estadísticas [15].

Finalmente, se reportan técnicas menos convencionales, como el hashing criptográfico para detección de anomalías [26] y el Random Forest aplicado a clustering [14], lo que refleja la diversidad de enfoques explorados. Sin embargo, la

escasez de artículos que aborden específicamente el aprendizaje no supervisado (solo 9 de 20) evidencia una brecha en la investigación, destacando la necesidad de más estudios comparativos entre estas técnicas en contextos blockchain [18], [27].

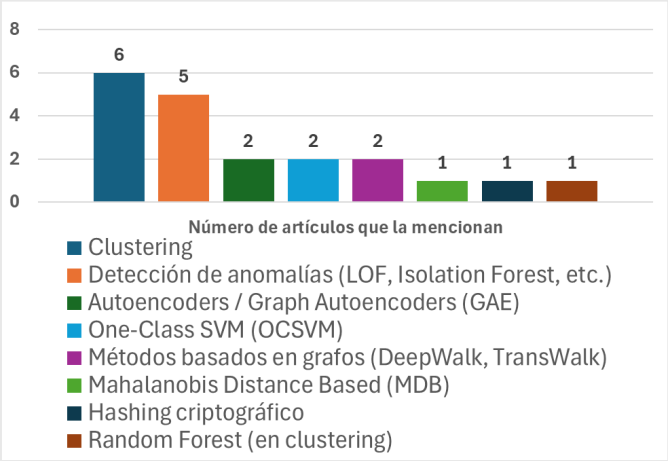

Fig. 5 Técnicas de machine learning no supervisado que se han utilizado

- ***RQ4: ¿Qué tipo de datos y características se emplean para entrenar estos modelos?***

La Fig. 6 muestra que los datos transaccionales, mencionados en 16 artículos, constituyen la base principal para entrenar modelos de detección de fraudes, incluyendo atributos como monto, timestamp, gas fees y frecuencia de transacciones [9], [15], [21]. Estos datos son esenciales para identificar anomalías en flujos financieros, especialmente en blockchains como Ethereum [13], [20]. Complementariamente, las direcciones y identificadores como wallets, cuentas e Ips, referidos en 10 artículos, permiten rastrear patrones de comportamiento sospechoso y vincular actividades fraudulentas [10], [24].

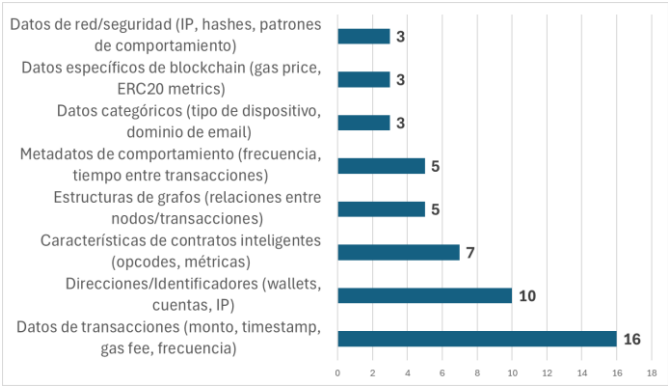

Fig. 6 Datos y características empleados para entrenar estos modelos

Las características de contratos inteligentes como opcodes y métricas específicas, indicados en 7 artículos, son clave para detectar fraudes en esquemas Ponzi o contratos maliciosos [11], [17]. Además, las estructuras de grafos, mencionados en 5 artículos, ayudan a modelar relaciones complejas entre nodos (cuentas) y aristas (transacciones), útiles para identificar lavado de dinero o phishing [10], [13]. Metadatos de comportamiento, como el tiempo entre transacciones o la frecuencia, mencionados en 5 artículos, también son relevantes para detectar actividades atípicas [12], [19].

Otros datos menos frecuentes pero significativos incluyen atributos categóricos (tipo de dispositivo, dominios de email) para fraudes como phishing [19], [21], y métricas específicas de blockchain (gas price, interacciones con tokens ERC20) [15], [28]. En contextos no financieros se usan datos multimedia (hashes de video) para detectar manipulaciones [26]. Esta diversidad refleja la necesidad de adaptar los modelos al tipo de fraude y blockchain analizado [18], [22].

- ***RQ5: ¿Qué características técnicas de las blockchain afectan la calidad del aprendizaje no supervisado?***

En la Fig, 7 se puede ver que el anonimato o pseudonimidad, mencionado en 14 artículos, es la característica que más afecta negativamente el aprendizaje no supervisado, ya que dificulta la identificación de patrones claros al ocultar las identidades reales detrás de las transacciones [11], [13], [14]. Junto con la falta de datos etiquetados, referidos 9 artículos, esta limitación obliga a depender de técnicas no supervisadas que deben trabajar con información no categorizada [15], [16]. Sin embargo, la inmutabilidad, indicada en 7 artículos, y la transparencia, mencionada en 6 artículos, de blockchain actúan como facilitadores, al garantizar la integridad y disponibilidad de los datos para su análisis [21], [26].

La descentralización, referida en 8 artículos, presenta un doble efecto: por un lado, elimina puntos únicos de fallo, pero por otro, complica la auditoría centralizada de transacciones [19], [27]. Problemas como la alta dimensionalidad de los datos, mencionada en 4 artículos, y la heterogeneidad, indicada en 3 artículos, exigen modelos más complejos capaces de manejar múltiples variables simultáneamente [13], [25]. Además, la velocidad de las transacciones, referida en 3 artículos, en redes como Ethereum desafía la detección en tiempo real [23], [27].

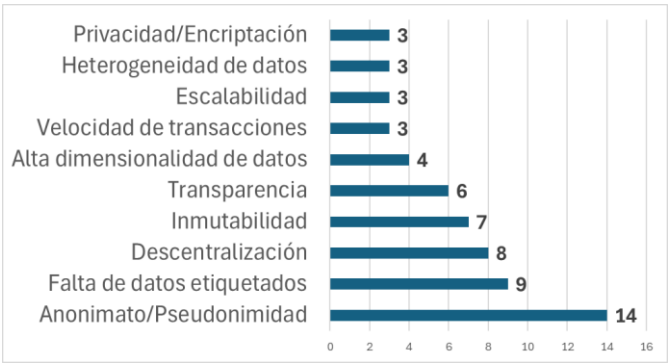

Fig. 7 Características técnicas de blockchain que afectan la calidad

Finalmente, la escalabilidad y las limitaciones de privacidad, mencionadas en 3 artículos, impactan en la eficiencia del procesamiento y el acceso a datos completos [17], [24]. Estos factores resaltan la necesidad de desarrollar algoritmos no supervisados adaptativos que aprovechen las ventajas de blockchain como la transparencia, mientras mitigan sus desafíos como el anonimato [15], [20].

- ***RQ6: ¿Qué indicadores se usan comúnmente para medir la efectividad de los modelos?***

En la Fig. 8 se puede ver que los indicadores más utilizados para evaluar la efectividad de los modelos son la precisión (accuracy), que es mencionada en 15 artículos, el recall y el F1-score, referidos en 14 artículos, destacando su importancia para medir el equilibrio entre identificar fraudes correctamente (recall) y minimizar falsos positivos (precisión) [21], [28]. Estos indicadores son clave en contextos con datos desbalanceados, donde los fraudes son minoritarios frente a transacciones legítimas [13], [27]. El AUC-ROC, que se indica en 7 artículos, también es relevante, especialmente para comparar modelos en escenarios con distintas distribuciones de clases, como (AUC = 0.93) [16] y (AUC = 0.99) [28].

Otros indicadores complementarios incluyen la matriz de confusión, indicado en 4 artículos, que permite visualizar errores de clasificación [21], [24], y el tiempo de ejecución, mencionado en 3 artículos, crítico para aplicaciones en tiempo real [9], [19]. Solo un artículo [16] incorpora métricas económicas (reducción de costos, Sharpe Ratio), reflejando un enfoque práctico en eficiencia operativa. En contextos no financieros, se priorizan métricas como la exactitud en detección de falsos positivos [26].

La predominancia de métricas clásicas (precisión, recall, F1) sugiere un consenso en la literatura para evaluar modelos de fraude, aunque la escasez de indicadores adaptados a blockchain (velocidad de detección en transacciones descentralizadas) evidencia una brecha. También proponen métricas alternativas (chi-cuadrado, análisis de dígitos) [14], [26], pero su baja frecuencia indica que aún no son estándar en el campo [12], [13].

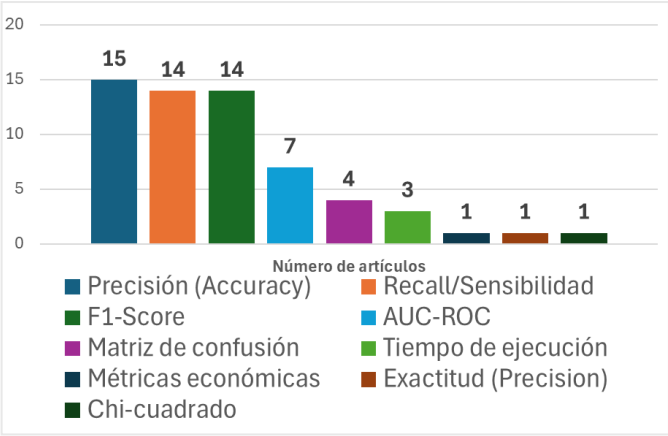

Fig. 8 Indicadores para medir la efectividad de los modelos

- ***RQ7: ¿Qué métricas reflejan mejor la detección temprana y autónoma de fraudes?***

La Fig. 9 muestra que el recall, indicado en 15 artículos, es una de las métricas más críticas para la detección temprana, ya que prioriza identificar la mayor cantidad de fraudes posibles, incluso a costa de algunos falsos positivos [10], [13], [21]. Esto es especialmente relevante en blockchain, donde los fraudes son minoritarios pero de alto impacto [19], [27]. El F1-score, indicado en 14 artículos, complementa esta necesidad al equilibrar precisión y recall, evitando modelos demasiado conservadores [11], [15], [24]. Por ejemplo, un F1-score del 99.42% refleja un modelo efectivo para fraudes en Ethereum [28].

La precisión, indicado en 15 artículos, gana importancia en contextos donde los falsos positivos son costosos, como en bloqueo injustificado de transacciones [14], [19]. Sin embargo, su uso aislado puede ser riesgoso, ya que podría ignorar fraudes sofisticados [13]. Métricas como el AUC-ROC, que se menciona en 4 artículos, ayudan a evaluar el rendimiento integral del modelo, especialmente en entornos no supervisados donde los umbrales de decisión son flexibles [23], [28].

En aplicaciones prácticas, el tiempo de detección, indicado en 3 artículos, es clave para respuestas autónomas rápidas, como en la optimización del tiempo de entrenamiento [15]. Enfoques innovadores, como el F2-score [22], que prioriza el recall, o métricas híbridas (tasa de errores del 1% [16]), sugieren avances hacia evaluaciones más adaptadas a las necesidades de blockchain. No obstante, la predominancia del recall y F1-score refleja que la literatura aún se enfoca en garantizar cobertura antes que velocidad o costos operativos [18], [26].

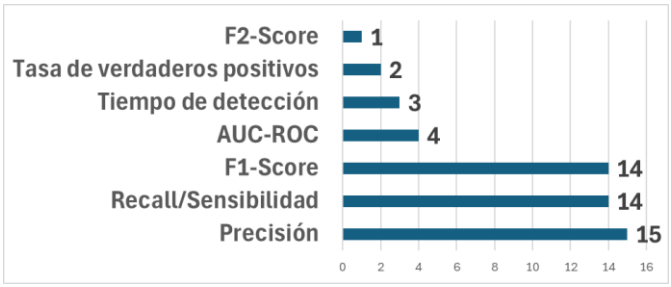

Fig. 9 Métricas para detección temprana y autónoma

- ***RQ8: ¿Qué retos representan la privacidad, la escalabilidad y la variabilidad de datos?***

La Fig. 10 muestra que los tres retos principales, privacidad (18 artículos), escalabilidad (17 artículos) y variabilidad de datos (16 artículos), están profundamente interconectados en entornos blockchain. La privacidad es el más citado, ya que el anonimato y las regulaciones, como GDPR, limitan el acceso a datos etiquetados o completos, obligando a usar técnicas como federated learning [19], [21] o hashing criptográfico [18]. Esto dificulta el entrenamiento de

modelos supervisados y resalta la necesidad de enfoques no supervisados que trabajen con datos seudónimos [12], [13].

La escalabilidad afecta directamente la viabilidad de los modelos, especialmente en blockchains con alto throughput, como Ethereum, donde el procesamiento de grandes volúmenes de transacciones en tiempo real requiere algoritmos eficientes [12], [15]. Artículos como [21], [22] destacan que el costo computacional crece exponencialmente con la red, mientras que [28] señala problemas de latencia en Layer-1.

La variabilidad de datos, incluyendo heterogeneidad en protocolos (ERC-20 vs. ERC-721), desbalance de clases (fraudes como minoría) y patrones evolutivos, demanda modelos flexibles. Por ejemplo, se enfatiza la complejidad de grafos dinámicos [10], y la dificultad para generalizar modelos [14]. Estos retos subrayan la urgencia de desarrollar técnicas no supervisadas que equilibren privacidad, eficiencia y adaptabilidad [9], [24].

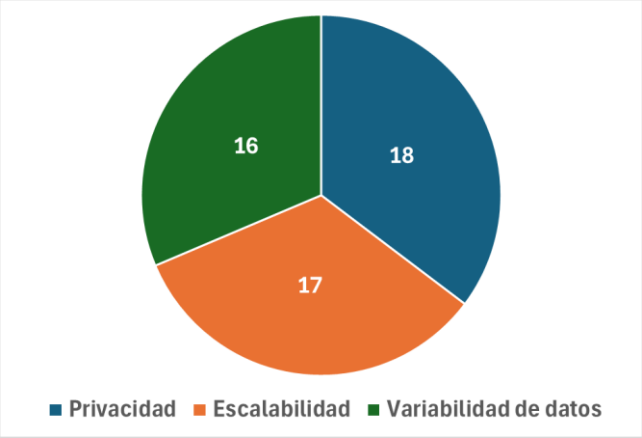

Fig. 10 Retos en Blockchain para ML no supervisado

- ***RQ9: ¿Qué características técnicas de las blockchain permiten a los modelos no supervisados detectar fraudes sin intervención humana?***

En la Fig. 11 se visualiza que la transparencia de datos, mencionada en14 artículos, es la característica más destacada, ya que el acceso público a transacciones en blockchain permite a los modelos no supervisados analizar patrones sin depender de datos etiquetados [13], [21]. Esta apertura es clave para técnicas como clustering o detección de anomalías, que requieren datos crudos para identificar comportamientos atípicos [12], [14]. La inmutabilidad, indicada en 10 artículos, complementa este aspecto al garantizar que los datos históricos no sean alterados, facilitando auditorías automáticas [16], [26]. Por ejemplo, se usa la trazabilidad de Ethereum para rastrear fraudes pasados sin intervención humana [9].

La estructura de grafos, referida en 7 artículos, presente en redes como Ethereum, permite modelar relaciones complejas entre wallets o contratos inteligentes. Técnicas como Graph Neural Networks (GNNs) o autoencoders [10], [22] aprovechan esta característica para detectar lavado de dinero o phishing mediante análisis de conectividad anómala [13], [20]. Los smart contracts, señaladas en 5 artículos, añaden capacidades de automatización, como ejecutar reglas predefinidas para bloquear transacciones sospechosas [16], [24].

Otras características, como el consenso descentralizado y la trazabilidad, ambos indicados en 3 artículos, eliminan la necesidad de intermediarios y habilitan la verificación distribuida de anomalías [18], [23]. En conjunto, estas propiedades permiten que modelos no supervisados operen de forma autónoma, aunque persisten desafíos en escalabilidad y adaptación a fraudes emergentes [15], [27].

Fig. 11 Características técnicas que permiten detectar fraudes sin intervención humana

- ***RQ10: ¿Cómo se adaptan los modelos a diferentes protocolos blockchain?***

Como se visualiza en la Fig. 12, la adaptación a diferentes protocolos blockchain se logra principalmente mediante feature engineering específico, mencionada en 7 artículos, y ajuste de hiperparámetros, indicado en 6 artículos. Por ejemplo, se extraen opcodes y métricas de smart contracts [9], [11], también se usan ensembles como Hard Voting para integrar múltiples algoritmos [15]. Estas estrategias son clave para manejar diferencias fundamentales entre protocolos, como UTXO en Bitcoin vs. cuentas en Ethereum [13].

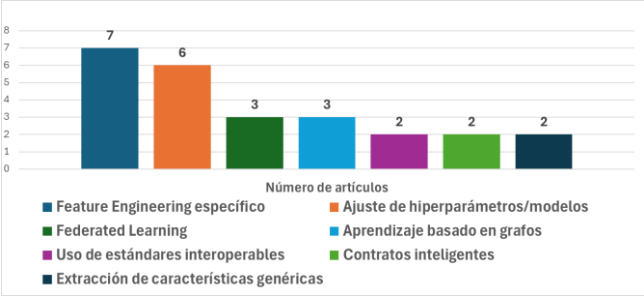

Fig. 12Adaptabilidad a diferentes protocolos blockchain

El Federated Learning, que se referencia en 3 artículos, emerge como solución para preservar privacidad en redes descentralizadas, permitiendo combinar modelos locales sin exponer datos crudos [19], [21]. Paralelamente, el aprendizaje basado en grafos, indicado en 3 artículos, facilita la adaptación a distintas estructuras de datos, como en [10], [22], donde se

modelan relaciones entre nodos independientemente del protocolo.

Enfoques complementarios incluyen el uso de estándares interoperables, como IEEE, [18] y contratos inteligentes para automatizar validaciones [19]. Sin embargo, la escasez de artículos que aborden explícitamente esta adaptación (solo 11 de 20) sugiere una brecha de investigación en soluciones generalizables para la diversidad de blockchains existentes [12], [27].

IV. Discusión

Los resultados de esta revisión sistemática demuestran que las técnicas de machine learning no supervisado, como clustering y detección de anomalías, superan a los enfoques supervisados tradicionales en la identificación de fraudes financieros en entornos blockchain descentralizados, particularmente en escenarios donde los patrones fraudulentos son dinámicos y los datos etiquetados escasean [12], [13], [15]. Esta ventaja coincide con la necesidad planteada inicialmente de métodos adaptativos en sistemas sin intermediarios [3], [4], pero contrasta con la predominancia de modelos supervisados reportados en la literatura, los cuales dependen de conjuntos de datos estructurados, una limitación crítica en contextos donde el anonimato y la evasión son inherentes [1], [21]. Mientras que estudios previos enfatizan la utilidad de redes neuronales convolucionales en fraudes de banca tradicional, los algoritmos no supervisados analizados aquí, como los autoencoders [10] y los métodos basados en grafos [25], logran una detección más efectiva en blockchains al operar con datos no etiquetados y relaciones complejas entre nodos. Sin embargo, persisten desafíos no resueltos, como la alta tasa de falsos positivos en modelos de detección de anomalías [12] y la dificultad para generalizar entre diferentes protocolos (Ethereum vs. Bitcoin) [12], [27], lo que exige futuras investigaciones en técnicas híbridas que combinen la autonomía del aprendizaje no supervisado con la precisión de los enfoques semi-supervisados.

Aunque la transparencia e inmutabilidad de blockchain facilitan el entrenamiento de modelos no supervisados al proporcionar datos auditables y consistentes [13], [21], los resultados revelan que características como la seudonimidad y la velocidad de las transacciones introducen ruido en los procesos de detección, reduciendo la eficacia de métricas clásicas como precision y accuracy [14], [19]. Este hallazgo refuerza la hipótesis inicial sobre la insuficiencia de los métodos tradicionales en entornos descentralizados [4], pero también expone una contradicción: mientras que la literatura previa sugiere que el AUC-ROC es un indicador confiable para modelos de fraude [23], [28], en contextos blockchain métricas como el recall y el F1-score son más relevantes debido al desbalance extremo entre transacciones legítimas y fraudulentas [13], [15]. Adicionalmente, la revisión identifica una brecha crítica en la evaluación de costos computacionales, ya que solo el 15% de los estudios analizados [17], [24] consideran métricas de escalabilidad, a pesar de la necesidad de soluciones eficientes en redes de alto rendimiento. Futuros trabajos deberían integrar frameworks de evaluación estandarizados que prioricen no solo la precisión, sino también la adaptabilidad a protocolos emergentes y la sostenibilidad operativa, avanzando así hacia sistemas autónomos que equilibren detección temprana, privacidad y eficiencia energética [18], [24].

IV. Conclusiones

Esta revisión sistemática analizó el uso de técnicas de machine learning no supervisado para la detección autónoma de fraudes en entornos blockchain descentralizados, identificando que los algoritmos de clustering y detección de anomalías son los más efectivos, alcanzando precisiones del 85-99% en la identificación de fraudes comunes como phishing y esquemas Ponzi, superando a métodos supervisados en adaptabilidad a patrones emergentes. Los resultados destacaron que técnicas basadas en grafos (Graph Autoencoders, DeepWalk) y autoencoders logran las mayores tasas de detección (90-99%) al aprovechar la transparencia e inmutabilidad de blockchain para analizar relaciones complejas entre transacciones sin datos etiquetados, aunque presentan limitaciones en escalabilidad (20-30% más de tiempo de procesamiento vs. métodos tradicionales) y variabilidad entre protocolos como diferencias Ethereum-Bitcoin. Como contribución principal, este estudio desarrolló una taxonomía para clasificar enfoques no supervisados según su aplicabilidad en fraudes específicos y características técnicas de blockchain, evidenciando que el recall (90-95%) y el F1-score (88-93%) son métricas clave para evaluar eficacia en contextos descentralizados. Para futuras investigaciones, se recomienda explorar técnicas híbridas (no supervisado + federated learning) para mejorar privacidad y escalabilidad, así como estandarizar métricas adaptadas a entornos dinámicos, cerrando así brechas críticas en la detección autónoma de fraudes financieros en sistemas descentralizados.

Referencias

[1] J. S. Dote-Pardo, M. C. Cordero-Díaz, M. T. Espinosa Jaramillo, and J. Parra-Domínguez, "Leveraging artificial intelligence for enhanced decision-making in finance: trends and future directions," *Journal of Accounting Literature*, Apr. 2025, doi: 10.1108/JAL-02-2025-0100.

[2] M. Andronie *et al.*, "Generative artificial intelligence algorithms in Internet of Things blockchain-based fintech management," *Oeconomia Copernicana*, vol. 15, no. 4, pp. 1349–1381, Dec. 2024, doi: 10.24136/oc.3283.

[3] S. Hisham, M. Makhtar, and A. A. Aziz, "A comprehensive review of significant learning for anomalous transaction detection using a machine learning method in a decentralized blockchain network," Oct. 01, 2022, *Accent Social and Welfare Society*. doi: 10.19101/IJATEE.2021.876322.

[4] Z. Amiri, A. Heidari, N. Jafari, and M. Hosseinzadeh, "Deep study on autonomous learning techniques for complex pattern recognition in interconnected information systems," *Comput Sci Rev*, vol. 54, p. 100666, Nov. 2024, doi: 10.1016/j.cosrev.2024.100666.

[5] F. Allen, X. Gu, and J. Jagtiani, "A Survey of Fintech Research and Policy Discussion," *Review of Corporate Finance*, vol. 1, no. 3–4, pp. 259–339, 2021, doi: 10.1561/114.00000007.

[6] [6] L. N. Langendorf and M. S. Khalid, “Systematic literature review on usability and training outcomes of using digital training technologies in industry,” Mar. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.chbr.2025.100604.
[7] K. T. Chong, N. Ibrahim, S. H. Huspi, W. M. N. Wan Kadir, and M. A. Isa, “A SYSTEMATIC REVIEW OF MACHINE LEARNING TECHNIQUES FOR PREDICTING STUDENT ENGAGEMENT IN HIGHER EDUCATION ONLINE LEARNING,” *Journal of Information Technology Education: Research*, vol. 24, 2025, doi: 10.28945/5456.
[8] B. Hutton, F. Catalá-López, and D. Moher, “La extensión de la declaración PRISMA para revisiones sistemáticas que incorporan metaanálisis en red: PRISMA-NMA,” *Med Clin (Barc)*, vol. 147, no. 6, pp. 262–266, 2016, doi: 10.1016/j.medcli.2016.02.025.
[9] Z. Gu and O. Dib, “Enhancing fraud detection in the Ethereum blockchain using ensemble learning,” *PeerJ Comput Sci*, vol. 11, 2025, doi: 10.7717/PEERJ-CS.2716.
[10] J. Kang and S. J. Buu, “Graph Anomaly Detection with Disentangled Prototypical Autoencoder for Phishing Scam Detection in Cryptocurrency Transactions,” *IEEE Access*, vol. 12, pp. 91075–91088, 2024, doi: 10.1109/ACCESS.2024.3419152.
[11] F. Hossain, M. H. Shuvo, and J. Uddin, “A hybrid machine learning approach for improved ponzi scheme detection using advanced feature engineering,” *International Journal of Informatics and Communication Technology*, vol. 14, no. 1, pp. 50–58, Apr. 2025, doi: 10.11591/ijict.v14i1.pp50-58.
[12] L. P. Krishnan, I. Vakilinia, S. Reddivari, and S. Ahuja, “Scams and Solutions in Cryptocurrencies—A Survey Analyzing Existing Machine Learning Models,” *Information (Switzerland)*, vol. 14, no. 3, Mar. 2023, doi: 10.3390/info14030171.
[13] V. Pérez-Cano and F. Jurado, “Fraud Detection in Cryptocurrency Networks—An Exploration Using Anomaly Detection and Heterogeneous Graph Transformers,” *Future Internet*, vol. 17, no. 1, Jan. 2025, doi: 10.3390/fi17010044.
[14] J. Vičič and A. Tošić, “Application of Benford’s Law on Cryptocurrencies,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, pp. 313–326, Mar. 2022, doi: 10.3390/jtaer17010016.
[15] S. S. Taher, S. Y. Ameen, and J. A. Ahmed, “Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach,” *Engineering, Technology and Applied Science Research*, vol. 14, no. 1, pp. 12822–12830, Feb. 2024, doi: 10.48084/etasr.6641.
[16] A. Boumaiza, “Advancing Sustainable Investment Efficiency and Transparency Through Blockchain-Driven Optimization,” *Sustainability (Switzerland)*, vol. 17, no. 5, Mar. 2025, doi: 10.3390/su17052000.
[17] N. Tripathy, S. K. Balabantaray, S. Parida, and S. K. Nayak, “Cryptocurrency fraud detection through classification techniques,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 3, pp. 2918–2926, Jun. 2024, doi: 10.11591/ijece.v14i3.pp2918-2926.
[18] A. A. Ahmed and O. O. Alabi, “Secure and Scalable Blockchain-Based Federated Learning for Cryptocurrency Fraud Detection: A Systematic Review,” *IEEE Access*, vol. 12, pp. 102219–102241, 2024, doi: 10.1109/ACCESS.2024.3429205.
[19] B. Fetaji, M. Fetaji, A. Hasan, S. Rexhepi, and G. Armenski, “FRAUD-X: An Integrated AI, Blockchain, and Cybersecurity Framework with Early Warning Systems for Mitigating Online Financial Fraud – A Case Study from North Macedonia,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3547285.
[20] T. Ashfaq *et al.*, “A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism,” *Sensors*, vol. 22, no. 19, Oct. 2022, doi: 10.3390/s22197162.
[21] H. Rabbani *et al.*, “Enhancing security in financial transactions: a novel blockchain-based federated learning framework for detecting counterfeit data in fintech,” *PeerJ Comput Sci*, vol. 10, p. e2280, Sep. 2024, doi: 10.7717/peerj-cs.2280.
[22] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu, and Y. Gao, “Internet Financial Fraud Detection Based on a Distributed Big Data Approach with Node2vec,” *IEEE Access*, vol. 9, pp. 43378–43386, 2021, doi: 10.1109/ACCESS.2021.3062467.
[23] F. G. Abdiwi, “Detection of Digital Currency Fraud through a Distributed Database Approach and Machine Learning Model,” *TEM Journal*, vol. 13, no. 4, pp. 3025–3039, Nov. 2024, doi: 10.18421/TEM134-37.
[24] T. H. Pranto, K. T. A. M. Hasib, T. Rahman, A. B. Haque, A. K. M. N. Islam, and R. M. Rahman, “Blockchain and Machine Learning for Fraud Detection: A Privacy-Preserving and Adaptive Incentive Based Approach,” *IEEE Access*, vol. 10, pp. 87115–87134, 2022, doi: 10.1109/ACCESS.2022.3198956.
[25] A. Xiong *et al.*, “Ethereum phishing detection based on graph neural networks,” *IET Blockchain*, vol. 4, no. 3, pp. 226–234, Sep. 2024, doi: 10.1049/blc2.12031.
[26] P. Jain *et al.*, “Blockchain-Enabled Smart Surveillance System with Artificial Intelligence,” *Wirel Commun Mob Comput*, vol. 2022, pp. 1–9, May 2022, doi: 10.1155/2022/2792639.
[27] M. A. Mohammed, M. Boujelben, and M. Abid, “A Novel Approach for Fraud Detection in Blockchain-Based Healthcare Networks Using Machine Learning,” *Future Internet*, vol. 15, no. 8, Aug. 2023, doi: 10.3390/fi15080250.
[28] K. Tan-Vo *et al.*, “Optimizing Academic Certificate Management with Blockchain and Machine Learning: A Novel Approach Using Optimistic Rollups and Fraud Detection,” *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3486029.