Revista Innovación y Software
Vol. 7, No. 1, Mes Marzo - Agosto, 2026
ISSN: 2708-0935
Pág. 183-206
https://revistas.ulasalle.edu.pe/innosoft

Tipo de artículo: Artículos de revisión
Temática: Redes y seguridad informática
Recibido: 7/12/2025 | Aceptado: 3/2/2026 | Publicado: 30/3/2026

Identificadores persistentes:
DOI: 10.48168/innosoft.s29.a342
ARK: ark:/42411/s29.a342

Enfoques de detección de ransomware basados en aprendizaje automático

Machine learning-based ransomware detection approaches

Luis Fernando Avila Reyes[1][0009-0000-0827-7491]*, **Kevin Eduardo Galvez Carrillo**[2][0009-0007-4013-9081], **Alberto Carlos Mendoza De Los Santos**[3][0000-0002-0469-915X]

[1]Universidad Nacional de Trujillo. Trujillo, Perú.. t1013300721@unitru.edu.pe
[2]Universidad Nacional de Trujillo. Trujillo, Perú.. t1023300621@unitru.edu.pe
[3]Universidad Nacional de Trujillo. Trujillo, Perú.. amendozad@unitru.edu.pe

*Autor para correspondencia: t1013300721@unitru.edu.pe

Resumen
El estudio tuvo como objetivo analizar los enfoques y técnicas de detección de ransomware basados en aprendizaje automático, a fin de identificar las propuestas más eficaces reportadas en la literatura reciente. Se aplicó la metodología PRISMA para seleccionar artículos originales publicados entre 2020 y 2025 en bases de datos especializadas. Los hallazgos muestran que los métodos tradicionales basados en firmas resultan insuficientes ante variantes de día cero, mientras que algoritmos como Random Forest, Gradient Boosting y redes neuronales profundas ofrecen mayor precisión y capacidad de adaptación. Asimismo, se destacan enfoques híbridos y emergentes que incorporan análisis forense con modelos de lenguaje o inteligencia artificial explicable. Se concluye que las técnicas de aprendizaje automático representan una alternativa robusta y en evolución para la detección temprana de ransomware, contribuyendo a mejorar la resiliencia de los sistemas de ciberseguridad.

Palabras claves: Aprendizaje automático, Ciberseguridad, Detección, Ransomware, Redes neuronales

Abstract
The study aimed to analyze machine learning-based ransomware detection approaches in order to identify the most effective proposals reported in recent literature. The PRISMA methodology was applied to select original articles published between 2020 and 2025 in specialized databases. Findings show that traditional signature-based methods are insufficient against zero-day variants, while algorithms such as Random Forest, Gradient Boosting, and deep neural networks provide higher accuracy and adaptability. Likewise, hybrid and emerging approaches that incorporate forensic analysis with language models or explainable artificial intelligence stand out. It is concluded that machine learning techniques represent a robust and evolving alternative for early ransomware detection, contributing to strengthening the resilience of cybersecurity systems.

Keywords: *Machine learning, Cybersecurity, Detection, Ransomware, Neural networks.*

Introducción

En los países desarrollados, el uso de sistemas informáticos e Internet se ha convertido en un parte fundamental de la economía y de la vida cotidiana. Esta creciente interconexión tambien ha abierto nuevas oportunidades para que los ciberdelincuentes actúen, desarrollando ataques cada vez más sofisticados [1]. A medida que crece el acceso a Internet y se amplía el uso de la tecnología, no solo se incrementa el número de ciberataques, tambien sus métodos evolucionan. Un ciberataque es el intento de vulnerar o explotar sistemas y redes informáticas de personas u organizaciones [2].

Dentro de las multiples formas de malware, que es el software diseñado para causar daño o infiltrarse en sistemas ajenos, el ransomware ha surgido como una amenaza grave en los últimos años. Este tipo de malware afecta a individuos, empresas, hospitales e incluso a infraestructuras críticas como los sistemas de suministro de energía [3].

El ransomware representa actualmente una de las amenazas más relevantes en el ámbito de la ciberseguridad, ya que provoca considerables perjuicios económicos y compromete la confidencialidad tanto de individuos como de instituciones [4]. Se trata de un riesgo de ciberseguridad en aumento, ya que cifra la información y exige un pago para su recuperación. La evolución constante de esta amenaza suele superar los métodos tradicionales de detección de ransomware [5]. En este escenario, el ransomware se ha distinguido por su capacidad de restringir el acceso de los usuarios a sus sistemas o archivos, bien sea mediante el bloqueo de la pantalla o el cifrado de documentos esenciales, hasta que se pague un rescate [6]. Lo que lo ha consolidado como una de las amenazas digitales más perjudiciales de la última década. Del mismo modo, su impacto económico ha quedado en evidencia con ataques de gran magnitud como WannaCry y NotPetya, responsables de pérdidas globales que superan los ocho mil millones de dólares [7].

En especial, los ataques de ransomware de día cero, que buscan aprovechar fallos aún no identificados, constituyen una seria amenaza para las defensas de ciberseguridad actuales. La ausencia de datos de entrenamiento hace que su detección continúe siendo un reto considerable [8]. Asimismo, este tipo de ataques suele valerse de métodos avanzados de cifrado para explotar nuevas vulnerabilidades [9].

Las soluciones antiransomware habituales con frecuencia no logran identificar ataques de ransomware de día cero, pues no es posible disponer anticipadamente de sus firmas para entrenar los modelos de detección [9]. Las soluciones centradas en la mitigación suelen enfocarse en alertar o detener la ejecución del ransomware, pero pocas veces abordan mecanismos que permitan prevenir de forma proactiva la ocurrencia de estos ataques [10].

Es importante señalar que la modalidad de ransomware como servicio (RaaS) ha intensificado este escenario de riesgo, ya que facilita que incluso ciberdelincuentes sin conocimientos técnicos avanzados puedan desplegar

campañas a gran escala mediante infraestructuras alojadas en la nube [11]. Está dinámica ha contribuido no solo a aumentar la frecuencia de los ataques, sino también a elevar su nivel de sofisticación en los ultimos años.

Los sistemas de detección tradicionales se apoyan fundamentalmente en enfoques basados en firmas, los cuales identifican malware mediante patrones previamente definidos. Pero, estos métodos enfrentan importantes limitaciones ante la rápida evolución del ransomware y la aparición de ataques de día cero [11]. Tambien, requieren actualizaciones constantes y personal con conocimientos especializados [12].

Como respuesta, la comunidad investigadora ha enfocado sus esfuerzos en la detección de ransomware mediante el uso de tecnologías como el aprendizaje automático [13]. Los enfoques convencionales para la detección de malware, como aquellos sustentados en técnicas estadísticas, resultan insuficientes frente a la evolución del ransomware, pues tienden a producir un elevado número de falsos positivos [14].

Los métodos de aprendizaje automático (ML) han mostrado una mayor efectividad en la detección de ransomware en comparación con las técnicas tradicionales basadas en firmas [15]. Aunque las técnicas de aprendizaje automático y profundo representan alternativas prometedoras, la falta de transparencia de los complejos modelos de caja negra puede limitar su implementación en contextos de seguridad sensibles [16].

En consecuencia, se vuelve fundamental diseñar estrategias innovadoras e inteligentes que permitan una protección más efectiva contra este tipo de amenaza [14]. Entre ellos, los algoritmos fundamentados en estructuras de árboles, como los árboles de decisión (DT), los bosques aleatorios (RF) y el eXtreme Gradient Boosting (XGBoost), han captado una atención significativa dentro de la comunidad investigadora en ciberseguridad [15].

Frente a este panorama, resulta necesario ordenar y examinar lo que la investigación ha producido en torno al uso del aprendizaje automático para la detección de ransomware. Aunque en la literatura se describen multiples enfoques, todavía no existe una vision integrada que permita reconocer con claridad cuáles son las técnicas más empleadas, cúales han mostrado mejores resultados y qué limitaciones persisten en su aplicación.

Bajo esta premisa, se plantea la siguiente pregunta de investigación: ¿Cuáles son los enfoques, técnicas y tendencias más relevantes en la detección de ransomware basados en aprendizaje automático reportados en la literatura científica reciente?

Materiales y métodos o Metodología computacional

Marco metodológico y términos de busqueda

Para el desarrollo de este estudio se aplicó la metodología PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), la cual ofrece un marco estructurado y transparente que facilita la identificación, selección y análisis crítico de los trabajos más relevantes dentro de la literatura científica [17].

A partir de la pregunta de investigación planteada, se definieron los siguientes terminos de búsqueda: "ransomware detection", "machine learning", "approaches", "techniques". Posteriormente, mediante el uso de operadores booleanos para ampliar y refinar los resultados, la cadena final quedó conformada de la siguiente manera: "ransomware detection" AND "machine learning" AND (approaches OR techniques).

Bases de datos consultadas

La búsqueda se llevó a cabo en bases de datos académicas reconocidas por su cobertura y fiabilidad, a fin de garantizar la calidad de evidencia recopilada. Se consultaron ScienceDirect, IEEE Xplore y ACM Digital Library. Para garantizar la actualidad de los hallazgos, se estableció como criterio temporal el intervalo comprendido entre los años 2020 y 2025. Además, se restringió los resultados a publicaciones en español e inglés. Finalmente, se filtró la inclusión solo de artículos originales y de libre acceso.

Criterios de inclusión y exclusión

Para garantizar la pertinencia y consistencia de los estudios incluidos, se definieron criterios de elegibilidad claros. Los criterios de inclusión y exclusión se muestran en la Tabla 1.

Tabla 1. Criterios de inclusión y exclusión

Tipo	**Criterio**
Inclusión	Estudios publicados entre 2020 y 2025 Artículos originales y de acceso abierto Propuestas basadas en ML aplicadas a la detección de ransomware
Exclusión	Revisiones, encuestas, artículos de opinión o editoriales Investigaciones sobre malware genérico Métodos no basados en ML

Dentro de la metodología PRISMA se emplea un diagrama de flujo de cuatro fases (Figura 1), el cual per-

mite representar el proceso de depuración y selección de los artículos que serán considerados en la revisión sistemática.

Revista Innovación y Software
Vol. 7, No. 1, Mes Marzo - Agosto, 2026
ISSN: 2708-0935
Pág. 183-206
https://revistas.ulasalle.edu.pe/innosoft

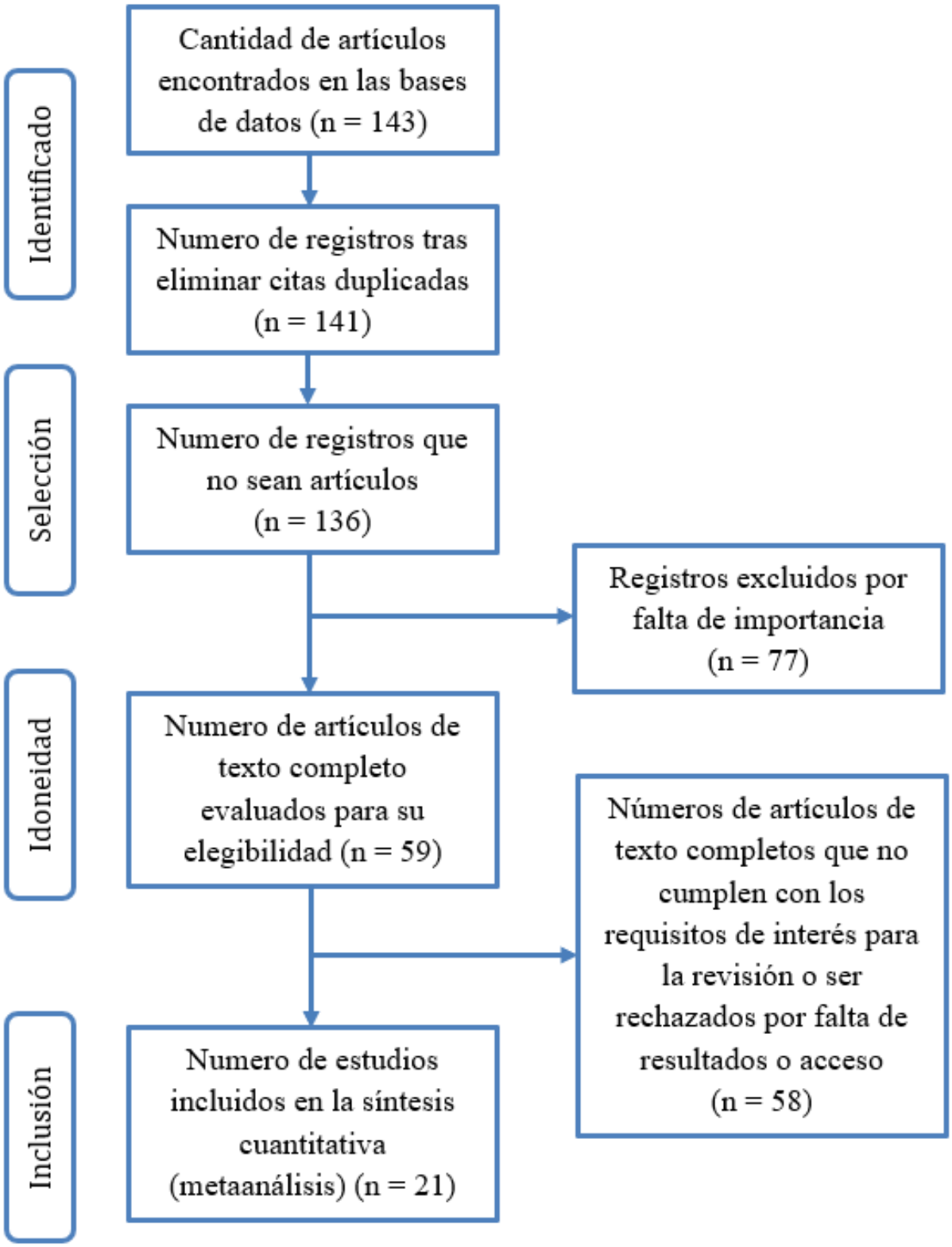

Figura 1. Diagrama de flujo de 4 estados

Revista Innovación y Software
Vol. 7, No. 1, Mes Marzo - Agosto, 2026
ISSN: 2708-0935
Pág. 183-206
https://revistas.ulasalle.edu.pe/innosoft

Resultados y discusión

Resultados

El aprendizaje automático (ML) se ha consolidado como una estrategia de gran relevancia en la detección de ransomware, al posibilitar la identificación de patrones en extensos volúmenes de datos. En especial, los métodos de ML fundamentados en enfoques basados en conjuntos destacan por ofrecer elevados niveles de precisión, una notable capacidad de adaptación frente a amenazas emergentes y una gestión eficiente de grandes cantidades de información [18].

Teniendo en cuenta los notables progresos en redes informales, computación en la nube, entornos web, banca electrónica, sistemas adaptativos e inteligentes, la seguridad informática se presenta como un ámbito en constante expansión que abre múltiples líneas de análisis. En este contexto, diversas soluciones basadas en aprendizaje automático han demostrado ser efectivas para abordar la complejidad de los problemas asociados a la ciberseguridad [19]. El aprendizaje automático puede aplicarse de distintas formas dentro de este campo. La Figura 2 ilustra algunas de sus aplicaciones en el ámbito de la seguridad informática.

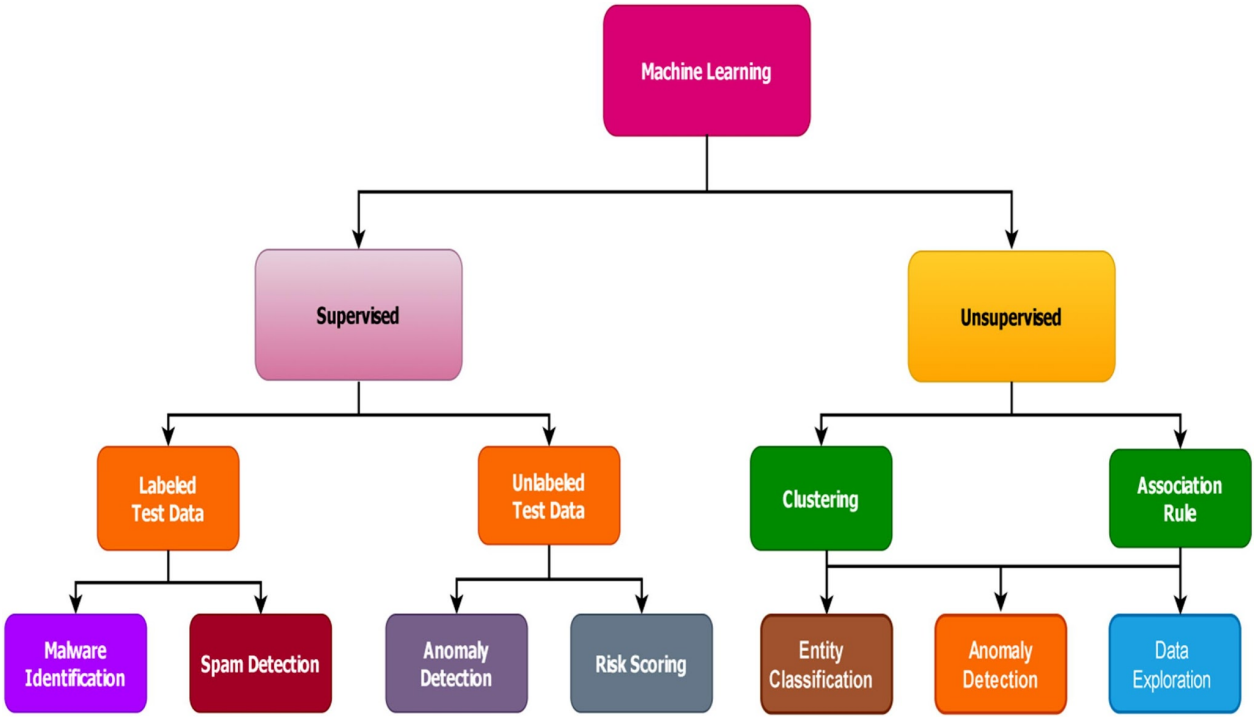

Figura 2. Aplicaciones de aprendizaje automático en ciberseguridad

En la Tabla 2 se resumen los trabajos que han propuesto métodos de detección de ransomware en equipos y redes que no son dispositivos móviles. La información se organiza de acuerdo con el enfoque de detección planteado, el algoritmo de aprendizaje automático utilizado, las métricas de desempeño reportadas y las limitaciones.

Tabla 2. Síntesis de estudios sobre detección de ransomware en entornos no moviles

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Arabo et al., 2020	Detección de Ransomware basado en el comportamiento de procesos de ejecución	Neural Net, Nearest Neighbors, Linear SVM, RBF SVM, Gaussian Process, Decision Tree, Random Forest, AdaBoost, QDA, Naive Bayes	Se empleó como métrica Accuracy para comparar los algoritmos de ML, donde Random Forest fue el mejor con una exactitud de 75,01 %	Dependencia de umbrales y patrones concretos (extensiones, frecuencia de API calls), susceptibles a evasión
Singh et al., 2024	El trabajo propone RansoDetect Fusion, un modelo de detección de Ransomware-as-a-Service (RaaS) enfocado en datos cifrados en la nube	Arquitectura principal: un ensemble de Deep Learning. Modelos individuales: tres Multilayer Perceptrons (MLPs) con distinta función de activación: MLP1: ReLU MLP2: SeLU MLP3: ELU	RansoDetect Fusion (ensemble): Accuracy: 98,79 % Recall: 98,79 % Precision: 98,85 % F1-score: 98,80 % Modelos individuales: MLP3: Accuracy, Recall y F1 ≈ 96,8 % MLP2: Accuracy ≈ 96,97 % MLP1: Accuracy ≈ 96,39 %	El desempeño puede variar frente a ataques zero-day o nuevas variantes de RaaS no presentes en los datasets

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Almashhadani et al., 2022	El trabajo propone MFMCNS (Multi-Feature Multi-Classifier Network-based System), un sistema de detección network-based diseñado para identificar actividades de ransomworm en su fase de propagación	- Decision Tree (Fine Tree) para vista session-based - Ensemble (Bagged Tree) para vista time-based - Fusión mediante majority voting	- Accuracy: 99,6–99,9 % - Recall: ≈ 100 % - FPR: ˜0,002–0,004 - F1: ≈ 0,997	Dependencia de la naturaleza del dataset y familias estudiadas. Alcance limitado: se centra en ransomworms propagados en red, no en otras variantes de ransomware
Kok et al., 2022	Detección de crypto-ransomware en la etapa anterior al cifrado (pre-cifrado) mediante el Pre-Encryption Detection Algorithm (PEDA). Combina la coincidencia de firmas (hash SHA-256) y análisis dinámico de APIs precifrado extraídas desde sandbox	Random Forest (RF) entrenado sobre secuencias de API pre-cifrado (con y sin discretización)	Principalmente Recuperación (Recall/Tasa de verdaderos positivos): 100 % (80:20 split) y 99.9 % (10-fold CV); también reportan Precisión, F1-Score, MCC, ROC y PRC	Dependencia de APIs de Windows para cifrado: si el ransomware usa código de cifrado propio (sin CryptoAPI), PEDA no lo detecta; debe verse como complemento, no solución única. Ámbito estrecho: diseñado para un tipo de malware (crypto-ransomware)

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Oh et al., 2024	El trabajo introduce volGPT, una integración entre la herramienta de análisis de memoria Volatility y un Large Language Model (LLM) para asistir en la triage forense de procesos de ransomware en memoria RAM	El núcleo ML es un Large Language Model (LLM) (basado en GPT 3.5), utilizado para interpretar la salida de Volatility y generar evaluaciones sobre procesos sospechosos	Accuracy general: 94,1 % en clasificación de familias: STOP/Djvu: 99 % Cerber: 93,8 % Locky: 91,9 % TeslaCrypt: 87,1 % Triage efficiency: 10,4 % (es decir, logra reducir ˜90 % de muestras sin descartar verdaderos positivos)	Dependencia de un LLM externo: implica riesgos de privacidad y seguridad si se emplea un modelo alojado en la nube

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Hernandez-Jaimes et al., 2024	Detección temprana basada en análisis de archivos y patrones de comportamiento para entornos de Internet of Medical Things (IoMT)	Los vectores de Nilsimsa fingerprints son usados como características principales. Clasificadores probados: Naïve Bayes (NB) Support Vector Machine (SVM) Random Forest (RF) k-Nearest Neighbors (k-NN) Decision Tree (DT)	Métricas evaluadas: Accuracy, Precision, Recall, F1-score. Resultados clave: El uso de Nilsimsa fingerprinting mejorado mejoró el desempeño en todos los clasificadores Random Forest alcanzó la mayor precisión global (≈ 96–97 % accuracy, con F1 superior al 0,95) SVM y k-NN también mostraron resultados competitivos, pero NB quedó rezagado	Evaluación realizada con un dataset limitado, sin acceso a repositorios amplios de ransomware dirigido a IoMT

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Khan et al., 2020	Propone DNAact-Ran, un motor de "secuenciación de ADN digital" para la detección de ransomware. Se centra en caracterizar el software como ransomware o goodware a partir de su "genoma digital", evitando depender de firmas tradicionales que requieren un ataque previo para crear la huella.	Active Learning con regresión lineal Random Forest (RF). Naïve Bayes (NB). Sequential Minimal Optimization (SMO)	Métricas evaluadas: Precisión, Recall, F1, Accuracy, TP y FP rates Active Learning: 87.9 % accuracy, superando a AdaBoost (83,2 %), Naïve Bayes (78,5 %) y Decision Stump (75,8 %)	Dataset relativamente pequeño (solo ˜1,5K muestras, de las cuales 300 usadas para entrenamiento)

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Kabuye et al., 2025	Detección de ransomware basado en inteligencia artificial explicable y consciente de la incertidumbre	Random Forest (RF) Support Vector Machine (SVM) K-Nearest Neighbors (KNN) Deep Neural Networks (DNN) El sistema incorpora técnicas de XAI (Explainable AI), como SHAP (SHapley Additive exPlanations), para interpretar la importancia de las características	Métricas utilizadas: Accuracy, Precision, Recall, F1-score y AUC Resultados principales: Deep Neural Networks (DNN) alcanzaron el mejor rendimiento: Accuracy: 98,6 % Precision: 98,4 % Recall: 98,7 % F1-score: 98,5 %	Aunque se logró explicar las decisiones, el costo computacional de las técnicas de interpretabilidad es alto para despliegues en tiempo real
Al-Hawawreh et al., 2021	Modelo federado asincrónico de detección de ransomware orientado a entornos de Internet Industrial de las Cosas (IIoT)	Random Forest (RF) Support Vector Machine (SVM) Decision Tree (DT) K-Nearest Neighbors (KNN) Naive Bayes (NB)	Métricas utilizadas: Accuracy, Precision, Recall y F1-score Resultados destacados: Random Forest obtuvo el mejor desempeño: Accuracy: 98,2 % Precision: 97,9 % Recall: 98,1 % F1-score: 98,0 %	El modelo se validó en entornos simulados, no en implementaciones reales de IIoT

Continúa en la página siguiente

Tabla 2 – Continuación de la página anterior

Referencia	**Enfoque de detección**	**Algoritmo de ML utilizado**	**Métrica de desempeño**	**Limitaciones**
Hill et al., 2024	Clasificación de ransomware mediante el monitoreo de contadores de rendimiento de hardware (HPCs) en un sistema físico no virtualizado	Árboles de decisión Random Forest K-Nearest Neighbors (KNN) Support Vector Machine (SVM) Naive Bayes	Métricas reportadas: Accuracy, Recall, Precision, y F1-score Random Forest fue el mejor modelo: Accuracy: 97,3 % Precision: 97,6 % Recall: 97,3 % F1-score: 97,4 %	Evaluación realizada en un único sistema no virtualizado, lo que limita la generalización

La Tabla 3 expone un análisis comparativo de los enfoques más recientes que emplean técnicas de aprendizaje automático para la detección de ransomware en dispositivos Android. En dicho análisis se subrayan elementos centrales como las principales contribuciones, los métodos de clasificación aplicados, los tipos de ransomware considerados, los conjuntos de datos utilizados, las metodologías implementadas, los resultados alcanzados, así como las fortalezas y limitaciones inherentes a cada propuesta. [18]

Revista Innovación y Software
Vol. 7, No. 1, Mes Marzo - Agosto, 2026
ISSN: 2708-0935
Pág. 183-206
https://revistas.ulasalle.edu.pe/innosoft

Tabla 3. Evaluación de técnicas de aprendizaje automático de vanguardia para la detección de ransomware de Android

Referencia	**Contribución importante**	**Tipos de ransomware**	**Conjunto de datos**	**Método**	**Resultados**	**Fortalezas**
Farhan y Salman, 2024	Aprendizaje profundo para la detección de ransomware en Android con un enfoque en redes neuronales de avance (FNN)	Simplocker, LockerPin, DoubleLocker, ransomware de temática policial, Fusob	AndroZoo (benigno), RansomProber (malicioso)	FNN usando Keras Sequential con 3 capas densamente conectadas	Precisión 98,9 %, F1 0,662, Precisión 0,5, Recall 1,0	Alta precisión y recall; aprendizaje a partir de datos sin procesar
Ahmed et al., 2023	Uso de ML y DL para crear modelos eficientes, precisos y robustos para clasificación binaria de ransomware	10 tipos diferentes de ransomware para Android	Ransomware de Android	DT, SVM, KNN, Conjunto, FNN, TabNet	DT: Precisión 97,24 %, Exactitud 98,50 %, F1 98,45 %	Análisis exhaustivo; uso de datasets recientes
Sharma et al., 2021	ML supervisado con conjuntos, reducción de dimensionalidad y análisis en GPU/CPU	Locker y Crypto Ransomware	RansomProber (2721 muestras) y AndroZoo (2000 benignas)	PCA + Random Forest	Precisión 99,67 %	Extracción integral, PCA, análisis GPU/CPU

Continúa en la página siguiente

Tabla 3 – Continuación de la página anterior

Referencia	**Contribución importante**	**Tipos de ransomware**	**Conjunto de datos**	**Método**	**Resultados**	**Fortalezas**
Oneil Victoriano, 2018	ML con varios clasificadores	HelDroid (varios tipos)	HelDroid	DT, RF, Gradient Boost, AdaBoost	Precisión 98,05 %; DT 99,08 % en dataset transformado	Múltiples clasificadores; alta precisión
Faris et al., 2020	Máquina de aprendizaje extremo optimizada con Salp Swarm	CryptoLocker, WannaCry, Petya, etc.	1000 apps Android	Salp Swarm + Kernel Extreme Learning Machine	Precisión, exactitud y recall 98 %	Alta precisión, baja tasa FP
Hiba Zuhair	Rasgos de clúster híbrido para la seguridad de smartphones	Genric.17.1762, LockDroid, Koler, Pletor, Simplocker	Hel-Droid, Virus Total, Google Play Store, Herramienta APK	Análisis híbrido estático + dinámico con ML y DL	Precisión 96,50 % (DNN)	Evaluación holística; clustering híbrido
Bagui y Woods, 2021	Detección usando datos de tráfico de red	Cargador, Jisut, Koler, Lockerpin, Pletor, PornDroid, RansomBo, SVPeng, Simplocker, WannaLocker	CICAnMal2017	DT, NB, OneR + selección de características	DT: Precisión 99,67 %, Exactitud 99,68 %, Recall 99,67 %, F1 99,67 %	Selección de características eficaz; alto rendimiento

Continúa en la página siguiente

Tabla 3 – Continuación de la página anterior

Referencia	**Contribución importante**	**Tipos de ransomware**	**Conjunto de datos**	**Método**	**Resultados**	**Fortalezas**
Samah e Iman, 2019	Sistema estático basado en API para detectar ransomware en Android	No especificado	Dataset propio: 2959 ransomware, 500 benignos	Análisis estático con API	Precisión 97 %, reducción complejidad 26 %	Reducción de funciones mejora rendimiento
Masum et al., 2021	Marco basado en selección de características con múltiples clasificadores ML	Casillero, Cripto	Dataset propio con 138.047 muestras	DT, RF, NB, LR, NN	RF: Precisión 99 ± 0,01, Recall 0,97 ± 0,03, Precisión 0,99 ± 0,00	Selección de características integral; alta precisión
Arabo et al., 2020	Mecanismo de detección basado en análisis de comportamiento de procesos	ViraLock, WannaCry, Cerber, WinLocker	Dataset propio con 7 ransomwares	Análisis de procesos + ML	Precisión 98,9 %, FPR 1,5 %, FNR 2,6 %	Alta precisión; distingue entre benigno y ransomware

Los modelos Extra Trees y Random Forest se caracterizan por ofrecer un desempeño consistente y equilibrado en la mayoría de las métricas evaluadas. De manera similar, Gradient Boosting y CatBoost muestran resultados altamente competitivos. A su vez, K-Neighbors y Decision Tree mantienen un nivel de eficacia aceptable. Sin embargo, al analizar el rendimiento global en clasificación y la capacidad discriminativa, como se detalla en la Tabla 4, los primeros modelos evidencian una superioridad más marcada [20].

Tabla 4. Comparación de rendimiento de la estrategia propuesta en diferentes algoritmos de aprendizaje automático

	Exactitud	**Precisión**	**Recordar**	**Puntuación de F1**	**ROC-AUC**
Bosque aleatorio	0,955	0,951	0,967	0,959	0,993
Impulso de gradiente	0,956	0,954	0,967	0,960	0,990
Árboles adicionales	0,961	0,961	0,967	0,964	0,993
KVecinos	0,946	0,958	0,942	0,950	0,970
Árbol de decisiones	0,925	0,926	0,936	0,931	0,924
CatBoost	0,956	0,954	0,967	0,960	0,993

Discusión

Enfoques

La investigación evidencia una notable diversidad de enfoques, que van desde la monitorización del comportamiento en equipos individuales hasta el despliegue de arquitecturas distribuidas en entornos industriales. En términos generales, estos aportes pueden agruparse en tres líneas: (1) soluciones basadas en host y red, (2) modelos híbridos y de ensembles, y (3) propuestas emergentes que explotan técnicas no convencionales.

En la prímera línea, los sistemas basados en host se centran en identificar aplicaciones maliciosas a partir de métricas de uso de recursos y llamadas a API en tiempo real [21]. Su principal fortaleza es la capacidad de detector variantes zero-day, aunque su aplicación práctica sigue limitada por la falta de métricas estandarizadas que respalden los resultados. En paralelo, los enfocados en red han cobrado relevancia frente a amenazas como los ransomworms, ya que permiten analizar distintas vistas del tráfico y combinar clasificadores de manera cooperativa. Un ejemplo es el modelo MFMCNS [22], capaz de identificar fases tempranas de propogación antes de la ejecución del payload.

La segunda categoría corresponde a estrategias híbridas y de ensembles, diseñadas para supercar las limitaciones de los modelos individuales. Singh et al. [23], por ejemplo, demuestran que la combinación de perceptrones multicapa con distintas funciones de activación puede mejorar entre 2 y 3 % la precición y el F1-score, lo que refuerza la robustez de estos esquemas. En la misma dirección, Kok et al. [24] plantean un enfoque práctico que une firmas criptográficas (SHA-256) con clasificación de llamadas a API en la fase de pre-cifrado, alcanzando un balance entre rapidez y capacidad de adaptación frente a variantes desconocidas.

La tercera línea reúne propuestas emergentes con un marcado carácter interdisciplinario y orientado a dominios críticos. Destacan, por ejemplo, DNAact-Ran [25], que incorpora conceptos de bioinformática y aprendizaje activo; el uso de fingerprints tipo Nilsimsa en entornos médicos sensibles [26]; o el aprendizaje federado asincrónico aplicado a sistemas IIoT, con módulos de diagnóstico basados en deep learning que resguardan la privacidad de los datos [27]. Más recientemente, se han explorado combinaciones de modelos generativos y explicables. Entre ellas, volGPT [28], que integra análisis forense de memoria con LLMs para apoyar la clasificación de procesos sospechosos, y el marco propuesto por Kabuye et al. [29], que combina generación sintética de datos, estimación de incertidumbre e interpretabilidad en un mismo pipeline. En paralelo, la detección basada en contadores de rendimiento de hardware (HPCs) ha surgido como una alternativa prometedora, al aprovechar señales de bajo nivel del procesador para reconocer patrones anómalos incluso en variantes inéditas [30].

Además la propuesta de un enfoque innovador que emplea algoritmos de aprendizaje automático (como Decision Tree, Markov Test, K-nearest Neighbors (KNN), Kernel Trick y técnicas de Deep Learning) para el análisis de la entropía de archivos, entendida como la medida de aleatoriedad presente en ellos, con el fin de identificar ransomware en sistemas de respaldo de manera eficiente. Este planteamiento ofrece una perspectiva prometedora para la detección temprana de ransomware, lo que podría contribuir a reducir significativamente el impacto generado por este tipo de ataques [31].

Técnicas

Tras el análisis de los artículos revisados, se identificó que las técnicas de aprendizaje automático aplicadas a la detección de ransomware han evolucionado hacia una notable diversidad de estrategias. Entre las más utilizadas se encuentran los clasificadores tradicionales, tales como Random Forest, Support Vector Machines y k-Nearest Neighbors, que continúan empleándose debido a su simplicidad y eficiencia en escenarios con recursos limitados. Sin embargo, varios estudios coinciden en que estos métodos presentan dificultades para enfrentar variantes polimórficas u ofuscadas, lo que ha impulsado la transición hacia técnicas más sofisticadas.

Un hallazgo importante es la creciente presencia de redes neuronales profundas, en particular los perceptrones multicapa y las arquitecturas de deep learning, que han mostrado mejoras en métricas como la precisión y el F1-score. Singh et al. [23], por ejemplo, demostraron que la combinación de distintas funciones de activación dentro de un esquema ensemble logró un rendimiento superior respecto a modelos individuales. Este tipo de resultados refuerza la idea de que las técnicas basadas en aprendizaje profundo tienen un mayor potencial para capturar patrones complejos en los datos, sobre todo en fases tempranas de los ataques.

De manera complementaria, algunos autores han optado por combinar técnicas de diferente naturaleza. Un ejemplo claro es el trabajo de Kok et al. [24], quienes integraron firmas criptográficas con modelos de clasifica-

ción sobre llamadas a APIs, logrando un equilibrio entre detección rápida y capacidad de adaptación. Este tipo de aproximaciones híbridas evidencian que no existe una técnica única capaz de resolver todos los escenarios, sino que la combinación de métodos puede resultar más efectiva frente a la diversidad de amenazas actuales.

Otro aspecto relevante es el interés por metodologías innovadoras que amplían los límites de las técnicas tradicionales. En este sentido, se han explorado propuestas que incluyen desde el uso de fingerprints como Nilsimsa hasta la aplicación de modelos generativos para la creación de datos sintéticos que permitan entrenar de forma más balanceada los clasificadores. Asimismo, el empleo de métodos para cuantificar la incertidumbre y mejorar la interpretabilidad, como Monte Carlo Dropout o SHAP, refleja una preocupación creciente por dotar de mayor transparencia y confiabilidad a los modelos.

Se evidencia que las técnicas de aprendizaje automático aplicadas a la detección de ransomware no se limitan ya a los algoritmos clásicos, sino que tienden hacia arquitecturas profundas, híbridas e interpretables, en un esfuerzo por responder a los retos que plantean las variantes emergentes y los entornos críticos en los que estas amenazas se manifiestan.

Tendencias

Los estudios revisados muestran que la investigación sobre detección de ransomware está avanzando hacia soluciones más rápidas, automáticas y confiables. Una de las principales tendencias es lograr que los sistemas detecten la amenaza en apenas unos segundos y que, además, puedan activar respuestas inmediatas dentro de esquemas de seguridad más estrictos, como el modelo zero-trust [21]. Esto refleja un interés por no solo identificar el ataque, sino también contenerlo antes de que cause daños graves.

Otra línea clara es la apuesta por enfoques híbridos, que combinan distintas fuentes de información (como datos del equipo y del tráfico de red) con métodos de aprendizaje automático y profundo. Estos enfoques buscan ser más resistentes frente a variantes nuevas del ransomware, incluso aquellas que nunca han sido vistas antes. También se exploran esquemas de aprendizaje distribuido o federado, que permiten entrenar modelos sin necesidad de centralizar todos los datos, reduciendo riesgos de privacidad y mejorando la adaptación en entornos como la nube o el internet industrial de las cosas [27].

De igual forma, existe un creciente interés en hacer que los modelos sean explicables y transparentes, de modo que los analistas de seguridad puedan entender por qué un sistema tomó cierta decisión. Esto es clave para generar confianza y facilitar la adopción de estas soluciones en escenarios reales [29]. Asimismo, comienzan a aparecer propuestas que aprovechan nuevas señales de detección, como el uso de contadores de hardware o técnicas que analizan patrones en secuencias de tiempo, con el fin de anticiparse mejor a intentos de evasion

[24], [30].

Finalmente, varios autores coinciden en la necesidad de estandarizar las evaluaciones con bases de datos públicas y escenarios más realistas, como el análisis en tiempo real o en ambientes de producción [22]. Esto permitirá comparar enfoques de manera justa y avanzar hacia soluciones que sean realmente útiles y aplicables en la práctica cotidiana. En conjunto, las tendencias apuntan a un futuro en el que la detección de ransomware será más rápida, confiable y fácil de integrar en los sistemas de ciberseguridad existentes.

Conclusiones

El análisis realizado permitió establecer que el aprendizaje automático constituye una de las herramientas más prometedoras para la detección de ransomware, al superar las limitaciones de los métodos tradicionales basados en firmas y ofrecer mayor capacidad de adaptación frente a variantes emergentes. Los resultados evidencian que algoritmos como Random Forest, Gradient Boosting y redes neuronales profundas destacan por su precisión y robustez, especialmente cuando se aplican en esquemas híbridos o combinados con técnicas de inteligencia artificial explicable.

Asimismo, se identificó que la tendencia de la investigación se orienta hacia soluciones que integren rapidez, confiabilidad y transparencia, con aplicaciones en entornos críticos como la nube, IoT industrial y sistemas médicos. Este panorama confirma que no existe un único método capaz de abordar la complejidad del ransomware, sino que la combinación de modelos y el desarrollo de enfoques innovadores representan el camino más viable para fortalecer la ciberseguridad.

Finalmente, el trabajo contribuye a consolidar una visión integrada del estado actual de la detección de ransomware basada en aprendizaje automático, lo que abre la posibilidad de diseñar estrategias más efectivas y adaptativas. Futuras investigaciones deberían centrarse en la validación de estos enfoques en escenarios reales, con bases de datos estandarizadas y sistemas en tiempo real, a fin de garantizar su aplicabilidad práctica y su impacto positivo en la protección de infraestructuras digitales.

Contribución de Autoría

Avila Reyes Luis Fernando: Conceptualización, Investigación, Metodología, Análisis formal, Validación, Visualización, Redacción - borrador original.

Galvez Carrillo Kevin Eduardo: Conceptualización, Investigación, Metodología, Análisis formal, Validación, Visualización, Redacción - borrador original.

Mendoza De Los Santos Alberto Carlos: Supervisión, Administración de proyectos, Escritura, revisión y edición.

Referencias

[1] K. Basu, P. Krishnamurthy, F. Khorrami, and R. Karri, "A theoretical study of hardware performance counters-based malware detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 512–525, 2019.

[2] B. A. S. Al-rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions," *Computers & Security*, vol. 74, pp. 144–166, 2018.

[3] D. M. Nicol, "The ransomware threat to energy-delivery systems," *IEEE Security & Privacy*, vol. 19, no. 3, pp. 24–32, 2021.

[4] C. Mingcan, F. Jiang, and R. Doss, "Ransoguard: un marco basado en rnn que aprovecha las api sensibles previas al ataque para la detección temprana de ransomware," *Computadoras y seguridad*, vol. 150, p. 104293, 2025.

[5] M. Rele, J. Samuel, D. Patil, and U. Krishnan, "Explorando la detección de ransomware basada en inteligencia artificial y aprendizaje automático," *Procedia Ciencias de la Computación*, vol. 252, pp. 548–556, 2025.

[6] P. O'Kane, S. Sezer, and D. Carlin, "Evolution of ransomware," *IET Networks*, vol. 7, no. 5, pp. 321–327, 2018.

[7] E. Berrueta, D. Morato, E. Magaña, and M. Izal, "A survey on detection techniques for cryptographic ransomware," *IEEE Access*, vol. 7, pp. 144 925–144 944, 2019.

[8] B. Cui, Y. Hu, T. Qub, Y. He, and L. Sun, "Un nuevo enfoque de detección de ransomware de día cero basado en cvae y 1d-cnn," *Computación de alta confianza*, vol. 5, no. 4, p. 100338, 2025.

[9] C. Mingcan, X. Deng, F. Jiang, and R. Doss, "Zero-ran sniff: un método de detección temprana de ransomware de día cero basado en el aprendizaje de disparo cero," *Computadoras y seguridad*, vol. 142, p. 103849, 2024.

[10] J. Von, C. Feng, A. Huertas, R. Oles, G. Bovet, and B. Stiller, "Guardfs : un sistema de archivos para la detección y mitigación integradas de ransomware basado en linux," *Revista de Seguridad de la Información y Aplicaciones*, vol. 93, p. 104078, 2025.

[11] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," *Computers & Security*, vol. 111, p. 102490, 2021.

[12] S. Razaulla, C. Fachkha, C. Markarian, A. Gawanmeh, W. Mansoor, B. Fung, and C. Assi, "The age of ransomware: A survey on the evolution, taxonomy, and research directions," *IEEE Access*, vol. 11, pp. 40 698–40 723, 2023.

[13] J. Ispahany, R. Islam, Z. Islam, and A. Khan, "Detección de ransomware mediante aprendizaje automático: una revisión, limitaciones de la investigación y futuras direcciones," *IEEE Xplore*, vol. 12, pp. 68 785–68 813, 2024.

[14] I. Almomani, R. Qaddoura, M. Habib, S. Alsoghyer, A. A. Khayer, I. Aljarah, and H. Faris, "Detección de ransomware en android basada en un enfoque evolutivo híbrido en el contexto de datos altamente desequilibrados," *IEEE Access*, vol. 9, pp. 57 674–57 691, 2021.

[15] A. Gajjar, P. Kashyap, A. Aysu, P. Franzon, Y. Choi, C. Cheng, G. Pedretti, and J. Ignowski, "Rd-faxid: Detección de ransomware con xgboost acelerado por fpga," *ACM Digital Library*, vol. 17, no. 4, pp. 1936–7406, 2024.

[16] A. Alvi and Z. Jalil, "Xrguard: Un enfoque independiente del modelo para la detección de ransomware mediante análisis dinámico e ia explicable," *IEEE Access*, vol. 13, pp. 53 159–53 170, 2025.

[17] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, and A. Hróbjartsson, "Declaración prisma 2020: una guía actualizada para la publicación de revisiones sistemáticasthe prisma 2020 statement: an updated guideline for reporting systematic reviews," *Revista Española de Cardiología*, vol. 74, no. 11, pp. 790–799, 2021.

[18] A. Hossain, T. Hasan, F. Ahmed, S. Hasib, M. Hasan, and A. Haque, "Hacia una detección superior de ransomware en android: una perspectiva de aprendizaje automático conjunto," *Ciberseguridad y aplicaciones*, vol. 1, p. 100076, 2025.

[19] M. Azeem, D. Khan, S. Iftikhar, S. Bawazeer, and M. Alzahrani, "Análisis y comparación de la eficacia de la detección de malware: un estudio de enfoques de aprendizaje automático," *Heliyon*, vol. 10, no. 1, p. 23574, 2024.

[20] M. Sibtain, M. Hussain, Q. Riaz, S. Qadir, N. Riaz, and K.-H. Jung, "Detección de ransomware para android robusta y ligera mediante análisis de comportamiento y reducción de características," *Computadoras, materiales y continua*, vol. 84, no. 3, pp. 5177–5199, 2025.

[21] A. Arabo, R. Dijoux, T. Poulain, and G. Chevalier, "Detecting ransomware using process behavior analysis," *Procedia Computer Science*, vol. 168, pp. 289–296, 2020.

[22] A. Almashhadani, D. Carlin, M. Kaiiali, and S. Sezer, "Mfmcns: a multi-feature and multi-classifier network-based system for ransomworm detection," *Computers & Security*, vol. 121, p. 102860, 2022.

[23] A. Singh, H. A. Abosaq, S. Arif, Z. Mushtaq, M. Irfan, G. Abbas, A. Ali, and A. A. Mazroa, "Securing cloud-encrypted data: Detecting ransomware-as-a-service (raas) attacks through deep learning ensemble," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 857–873, 2024.

[24] S. H. Kok, A. Abdullah, and N. Z. Jhanjhi, "Early detection of crypto-ransomware using pre-encryption detection algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1984–1999, 2022.

[25] F. Khan, C. Ncube, L. K. Ramasamy, S. Kadry, and Y. Nam, "A digital dna sequencing engine for ransomware detection using machine learning," *IEEE Access*, vol. 8, pp. 119 710–119 719, 2020.

[26] M. L. Hernandez-Jaimes, A. Martínez-Cruz, K. A. Ramírez-Gutiérrez, and E. Guevara-Martínez, "Enhancing machine learning approach based on nilsimsa fingerprinting for ransomware detection in iomt," *IEEE Access*, vol. 12, pp. 153 886–153 897, 2024.

[27] M. Al-Hawawreh, E. Sitnikova, and N. Aboutorab, "Asynchronous peer-to-peer federated capability-based targeted ransomware detection model for industrial iot," *IEEE Access*, vol. 9, pp. 148 738–148 755, 2021.

[28] D. B. Oh, D. Kim, D. Kim, and K. H. Kim, "volgpt: Evaluation on triaging ransomware process in memory forensics with large language model," *Forensic Science International Digital Investigation*, vol. 49, p. 301756, 2024.

[29] H. Kabuye, B. Issac, R. Yumlembam, and J. Neera, "Explainable and uncertainty aware ai-based ransomware detection," *IEEE Access*, vol. 13, pp. 106 573–106 589, 2025.

[30] J. E. Hill, T. O. Walker, J. A. Blanco, R. W. Ives, R. Rakvic, and B. Jacob, "Ransomware classification using hardware performance counters on a non-virtualized system," *IEEE Access*, vol. 12, pp. 63 865–63 884, 2024.

[31] U. Tariq, "Combatir el ransomware en entornos de iot industriales activados por zephyros," *Heliyon*, vol. 10, no. 9, p. e29917, 2024.