Revisión Teórica De Las Redes Neuronales Profundas Para La Detección De Malware.

Sebastián Andrés Montero Franco

Asesor: José Nayid Cardona Castañeda

Universidad Nacional Abierta y a Distancia

Vicerrectoría Académica y de Investigación

Programa: Ingeniería de sistemas

2025

Resumen analítico especializado (RAE)	
Título	Revisión teórica de las Redes Neuronales profundas para la Detección de Malware.
Modalidad de trabajo de grado	Monografía
Línea de investigación	Gestión de Sistemas
Núcleo problemático	La creciente sofisticación de los ataques de malware ha superado la capacidad de las técnicas tradicionales de detección, como el análisis de firmas y heurísticas. Esto ha generado la necesidad de explorar enfoques avanzados como las redes neuronales profundas, que permitan detectar patrones ocultos y amenazas emergentes de forma eficaz y automatizada.
Autores	Sebastián Andrés Montero Franco
Institución	Universidad Nacional Abierta y a Distancia
Fecha	Mayo, 2025
Palabras claves	Ciberseguridad, redes neuronales convolucionales, redes neuronales recurrentes, Inteligencia Artificial, Malware.
Descripción	Esta monografía presenta una revisión teórica sobre la aplicación de redes neuronales profundas —específicamente CNN y RNN— en la detección de malware. Se analizan sus fundamentos, ventajas y limitaciones, así como su efectividad frente a amenazas avanzadas. También se examinan modelos híbridos y su rendimiento comparativo frente a técnicas tradicionales de machine learning.

Fuentes	Se consultaron más de 25 fuentes académicas, incluyendo artículos científicos, informes técnicos y trabajos de grado nacionales e internacionales, extraídos de bases como IEEE, Elsevier, Springer, y repositorios institucionales como el de la UOC.
Contenidos	El trabajo se organiza en capítulos temáticos que cubren desde los fundamentos de la ciberseguridad y el malware, hasta las características de las redes neuronales recurrentes y convolucionales. Se incluyen análisis comparativos de modelos, técnicas de evasión de malware y tablas de desempeño con métricas.
Metodología	Se utilizó un enfoque de revisión teórica de literatura, seleccionando y analizando estudios de caso relevantes sobre la aplicación de CNN, RNN y modelos híbridos en la clasificación de malware. La información fue sistematizada y organizada.
Conclusiones	Las redes neuronales profundas representan una alternativa poderosa para superar las limitaciones de los métodos tradicionales de detección de malware. Su capacidad de aprendizaje automático y adaptación a patrones complejos las hace especialmente útiles frente a amenazas evasivas y emergentes. Los modelos híbridos que combinan CNN y RNN muestran resultados prometedores, con altas tasas de precisión y bajo nivel de falsos positivos, lo que los convierte en candidatos viables para su implementación en sistemas de ciberseguridad de nueva generación.

Resumen

El presente trabajo realiza una revisión teórica sobre la detección de malware mediante el uso de redes neuronales profundas, enfocándose en el análisis de la eficacia de modelos como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN). A partir de una revisión de la literatura especializada, se identifican las principales técnicas, retos y avances recientes relacionados con la detección de software malicioso, en especial aquellos que presentan variantes sofisticadas utilizando técnicas de evasión como el cifrado, la ofuscación y el polimorfismo. El propósito de este estudio es explorar cómo las CNN son capaces de extraer patrones visuales de archivos binarios representados como imágenes, y cómo las RNN permiten modelar comportamientos temporales de ejecuciones maliciosas, lo que facilita su clasificación precisa. Asimismo, se examina la efectividad de enfoques híbridos que integran ambas arquitecturas para mejorar el rendimiento en la detección de amenazas complejas. Esta monografía también considera aspectos como la escalabilidad de los modelos, su capacidad de generalización y las implicaciones prácticas en entornos reales. Finalmente, el estudio ofrece una visión integral sobre la evolución de las técnicas de ciberseguridad basadas en Inteligencia Artificial, destacando tanto sus limitaciones como sus oportunidades de mejora. Así, se brinda una base teórica sólida para futuras investigaciones y desarrollos en este campo en constante crecimiento.

Palabras Clave: Ciberseguridad, redes neuronales convolucionales, redes neuronales recurrentes, Inteligencia Artificial, Malware.

Abstract

This work presents a theoretical review on malware detection using deep neural networks, focusing on the analysis of the effectiveness of models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Based on a review of specialized literature, the study identifies the main techniques, challenges, and recent advances related to malicious software detection, particularly those involving sophisticated variants that employ evasion strategies such as encryption, obfuscation, and polymorphism. The purpose of this study is to explore how CNNs can extract visual patterns from binary files represented as images, and how RNNs can model temporal behaviors of malicious executions, thus enabling more accurate classification. Furthermore, the effectiveness of hybrid approaches that integrate both architectures to enhance performance in detecting complex threats is examined. This monograph also considers aspects such as model scalability, generalization capacity, and practical implications in real-world environments. Finally, the study provides a comprehensive overview of the evolution of cybersecurity techniques based on Artificial Intelligence, highlighting both their limitations and opportunities for improvement. In doing so, it offers a solid theoretical foundation for future research and developments in this constantly evolving field.

Keywords: Cybersecurity, Convolutional Neural Networks, Recurrent Neural Networks, Artificial Intelligence, Malware.

Tabla de Contenido

Introducción 9
Planteamiento del problema 9
Justificación 11
Objetivos 13
Objetivo General 13
Objetivos Específicos 13
Marco Teórico 14
Arquitecturas de Aprendizaje Profundo: Evolución Conceptual y Taxonomía 14
Taxonomía Arquitectónica: Complejidad Estructural y Conectividad 15
Malware: Caracterización Técnica y Evolución Adaptativa 16
Paradigmas de Detección: Limitaciones Tradicionales y Oportunidades Emergentes 17
Metodología 19
Bases de datos consultadas 19
Palabras clave empleadas 19
Filtros aplicados 20
Criterios de inclusión y exclusión 20
Proceso de selección y análisis 21
Ciberseguridad 22
Malware 25
Principales Categorías del Malware (Mundaca, 2020) 28
Técnicas de Evasión del Malware 30
Redes Neuronales Recurrentes 39
Redes Neuronales Convolucionales 42
Combinación de CNN y RNN 45
Discusión 49
Conclusiones 52
Referencias 55

Lista de tablas

Tabla 1 Tabla comparativa entre Machine Learning y Deep Learning aplicada a la detección de malware..... 35
Tabla 2 Tabla comparativa de modelos híbridos (CNN+RNN) 48

Lista de figuras

Figura 1 Arquitectura interna de una celda LSTM (Long Short-Term Memory)..... 41
Figura 2 Arquitectura de una Red Neuronal Convolucional (CNN)..... 44

Introducción

En los últimos años, el malware se ha convertido en una de las amenazas más graves para la seguridad de los sistemas informáticos. A medida que estas amenazas evolucionan, los métodos tradicionales de detección, como el análisis de firmas o reglas fijas, se vuelven menos efectivos. Por eso, es necesario buscar nuevas herramientas que permitan identificar estas amenazas de forma más precisa y rápida.

En este contexto, las redes neuronales profundas, como las redes convolucionales (CNN) y las recurrentes (RNN), han demostrado ser útiles para detectar malware. Estas redes pueden analizar grandes cantidades de datos, encontrar patrones ocultos y adaptarse a nuevas amenazas sin necesidad de programar reglas específicas.

Este trabajo tiene como objetivo revisar cómo estas redes neuronales se están utilizando para la detección de malware, analizando sus ventajas, desventajas y casos de uso. De esta manera, se busca aportar al conocimiento de estas nuevas técnicas de ciberseguridad que ayuden a proteger mejor los sistemas ante ataques informáticos cada vez más complejos.

Planteamiento del problema

El crecimiento acelerado del malware y la constante evolución de sus técnicas de ataque presentan un desafío cada vez mayor para la ciberseguridad global. De acuerdo con el Trellix Advanced Research Center Threat Report de junio de 2024, se ha observado un aumento significativo en la aparición de variantes de malware diseñadas para evadir las soluciones de detección tradicionales, como el análisis basado en firmas y las tecnologías EDR (Endpoint Detection and Response). Estos ataques, que afectan a sectores críticos

como las telecomunicaciones, el transporte y las finanzas, subrayan la necesidad urgente de desarrollar métodos más avanzados y automatizados para la identificación y clasificación de estas amenazas emergentes (Trellix, 2024).

A medida que las soluciones basadas en análisis de firmas y heurísticas se vuelven ineficaces frente a nuevos ataques que alteran su estructura y comportamiento, surge la necesidad de explorar enfoques más robustos y adaptativos para mitigar estos riesgos (Pinhero et al., 2021). El uso de técnicas de aprendizaje profundo se ha convertido en un enfoque prometedor para enfrentar los desafíos que presenta la detección de malware avanzado. Las redes neuronales, en particular las convolucionales (CNN) y recurrentes (RNN), permiten analizar grandes volúmenes de datos, identificar patrones complejos y modelar secuencias temporales sin necesidad de un proceso extenso de ingeniería de características (Ahmad, 2020).

No obstante, la implementación de estos modelos en contextos reales no está exenta de dificultades, como el tiempo de procesamiento, la necesidad de grandes volúmenes de datos para entrenamiento y la susceptibilidad a ataques adversarios que pueden manipular la predicción de los modelos (Han et al., 2019; Pinhero et al., 2021). Esto genera un reto importante de seleccionar adecuadamente qué tipo de red neuronal se ajusta mejor a cada situación y tipo de amenaza, considerando tanto sus fortalezas como sus limitaciones inherentes.

Las CNN son particularmente eficaces para identificar patrones visuales en datos de malware representados como imágenes (Chen et al., 2019), mientras que las RNN se han utilizado para analizar secuencias temporales de datos, como flujos de red y registros de

actividad. Sin embargo, la implementación de estos modelos en entornos de ciberseguridad sigue siendo un desafío debido a problemas de sobrecarga computacional y a su susceptibilidad a técnicas de evasión, lo que plantea la necesidad de desarrollar arquitecturas híbridas o adaptativas que combinen lo mejor de ambos enfoques (Anandhi et al., 2022).

En este contexto, surge la siguiente pregunta de investigación: ¿Cómo se han utilizado las redes neuronales profundas, tanto convolucionales como recurrentes, para superar las limitaciones de las técnicas tradicionales y ofrecer una solución robusta en la detección de malware avanzado? Esta monografía ofrece una revisión teórica de la literatura sobre el uso de CNN y RNN en la detección de malware, recopilando y sintetizando hallazgos clave que permitan entender mejor su rendimiento y aplicabilidad en diversos entornos de ciberseguridad.

Justificación

La detección de malware es uno de los principales desafíos de la ciberseguridad a nivel global. Según el AV-Test Report, la cantidad de malware se ha multiplicado exponencialmente en los últimos años, alcanzando cifras de más de 800 millones de muestras únicas (Chen et al., 2019). Este aumento no solo ha incrementado la frecuencia de los ciberataques, sino que también ha diversificado las formas en las que el malware puede manifestarse, utilizando técnicas avanzadas de evasión como ofuscación de código y cifrado para evitar ser detectado. Estas características han hecho que las herramientas tradicionales de detección, basadas en firmas y métodos heurísticos, se vuelvan ineficaces frente a nuevas variantes que alteran su estructura y comportamiento para engañar a los sistemas de seguridad (Mudzfirah et al., 2019).

En 2023, Colombia enfrentó aproximadamente 12,000 millones de intentos de ciberataques, una cifra notablemente inferior a los 20,000 millones reportados en 2022. Sin embargo, la reducción en cantidad no representa una mejora, ya que los ataques actuales son más dirigidos y sofisticados, con mayor potencial de impacto si las defensas de ciberseguridad no están actualizadas y automatizadas (Cámara Colombiana de Informática y Telecomunicaciones, 2023). Además, en los primeros cinco meses de 2023, se registraron 23,640 delitos cibernéticos, destacando el hurto informático y violaciones de datos personales como los más frecuentes (Policía Nacional, 2023).

Dada esta situación, es fundamental estudiar enfoques más avanzados para mejorar la capacidad de detección de malware y proporcionar un marco teórico que aborde las deficiencias de las técnicas tradicionales. En este sentido, las redes neuronales profundas, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), han mostrado un gran potencial para identificar patrones complejos en el comportamiento del malware (Chen et al., 2019).

Las CNN permiten detectar patrones visuales en datos de malware representados como imágenes, mientras que las RNN son útiles para analizar secuencias temporales y flujos de comportamiento (Mudzfirah et al., 2019). A pesar de su potencial, la aplicación de estos modelos en la detección de malware presenta desafíos como la sobrecarga computacional, la necesidad de grandes volúmenes de datos para el entrenamiento y la susceptibilidad a ataques adversarios que pueden alterar la efectividad de los modelos (Gibert, 2016).

Esta monografía se justifica por la necesidad de realizar una revisión teórica que permita examinar el uso de redes neuronales profundas en la detección de malware,

aportando un marco conceptual que facilite la comprensión de cómo estas técnicas pueden superar las limitaciones de los métodos tradicionales. El trabajo no solo será relevante a nivel global, sino que también proporcionará un contexto específico para Colombia, resaltando la importancia de aplicar enfoques más robustos en un entorno donde las ciberamenazas siguen en aumento.

Objetivos

Objetivo General

Realizar una revisión teórica sobre el uso de redes neuronales profundas, específicamente CNN y RNN, en la detección de malware, enfocándose en su efectividad y limitaciones en comparación con técnicas tradicionales.

Objetivos Específicos

Explorar los conceptos y fundamentos teóricos de las redes neuronales profundas, con énfasis en su aplicación en la detección de malware.

Examinar el funcionamiento y aplicabilidad de las redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN) en la detección de malware, considerando su capacidad para detectar patrones complejos y secuencias temporales de comportamiento.

Comparar la efectividad de las CNN y las RNN en la detección de malware utilizando estudios de caso y revisiones bibliográficas para determinar sus fortalezas y debilidades.

Evaluar el desempeño de arquitecturas híbridas que combinan CNN y RNN en la detección de malware avanzado, con el fin de identificar si su integración mejora la precisión y capacidad de detección frente a amenazas complejas.

Marco Teórico

El presente capítulo esboza los fundamentos teóricos de la investigación centrada en la implementación de arquitecturas para el aprendizaje profundo en la identificación de malware. La confluencia entre la inteligencia artificial y la ciberseguridad ha dado lugar a un paradigma que elude los métodos tradicionales para la detección del malware, esto ha dado pie a interrogantes básicas sobre la efectividad, escalabilidad y capacidad de respuesta de estos nuevos métodos.

Arquitecturas de Aprendizaje Profundo: Evolución Conceptual y Taxonomía

La conceptualización de las redes neuronales artificiales como modelos computacionales bioinspirados ha experimentado una transformación radical desde sus fundamentos teóricos establecidos por McCulloch y Pitts en la década de 1940. El desarrollo del perceptrón por Rosenblatt en 1958 fue un avance importante, pero la revolución conceptual real comienza con el resurgimiento del deep learning a partir de la década de 2010, siendo posible de entrenar arquitecturas más complejas como consecuencia del poder computacional disponible y el acceso a cantidades masivas de datos (M., G., & Sethuraman, 2023). La taxonomía contemporánea de estas arquitecturas en cuanto a sus características funcionales en función del dominio donde son implementadas es evidente. Las redes neuronales convolucionales han sido una innovación arquitectónica que, aunque inicialmente desarrolladas para el procesamiento de información visual, han sido aplicadas con notable eficacia para la representación de archivos ejecutables con utilización de visualización a partir de técnicas de binarización (Yadav et al., 2022). Esta transposición ilustra con claridad la flexibilidad de los modelos asociados a redes neuronales artificiales

para migrar a dominios de aplicación que parecerían desvinculados entre sí. Las redes neuronales recurrentes introducen la temporalidad, una consideración muy a tener en cuenta en la modelización de secuencias de comportamiento como las trazas de llamadas a interfaces de programación de aplicaciones y patrones de actividad temporal (Jha et al., 2020; Rhode et al., 2017), constituyendo la capacidad de mantener estados en el interior de la arquitectura una consideración muy importante en la modelización de comportamientos dinámicos. La hibridación o fusión de estas arquitecturas son consideradas una aplicación efectiva dada la rico/abundante interrelación que destaca el procesamiento espacial y temporal frente a amenazas complejas (Musikawan et al., 2023).

Taxonomía Arquitectónica: Complejidad Estructural y Conectividad

La categorización de las arquitecturas neuronales de acuerdo a su grado de complejidad arquitectónica da cuenta de un recorrido que pasa de modelos simplicistas a modelos de creciente sofisticación. Las arquitecturas neuronales monocapa, que se caracterizan por la posibilidad de conectar la entrada con la salida de manera directa, dan cuenta de la forma más simple de la representación de estos modelos, mientras que las arquitecturas multicapa hacen uso de la jerarquización de las transformaciones por medio de las cuales se puede llevar a cabo la extracción de características de forma incremental. La dicotomía entre arquitecturas de alimentación directa o feedforward y arquitecturas recurrentes manifiesta dos visiones de computación diferentes. Las arquitecturas de alimentación directa tienen una concepción de computación de tipo unidireccional, en la cual se privilegian eficiencias computacionales, mientras que las arquitecturas recurrentes agrupan ciclos de computación que, aunque sean computacionalmente costosos, brindan

capacidades de modelización temporal más eficaces. La conectividad entre neuronas es otro parámetro arquitectónico fundamental. Las arquitecturas totalmente conectadas maximizan el tránsito de la información entre capas, pero a costa de una complejidad computacional exponencial. Las arquitecturas parcialmente conectadas son el acuerdo más práctico que triangula la capacidad expresiva del modelo con la computación (Catak et al., 2020).

Malware: Caracterización Técnica y Evolución Adaptativa

El malware pertenece a la clase de las amenazas digitales cuya sofisticación ha ido en aumento a medida que las TIC han ido desarrollándose. Lo que se considera técnicamente malware es cualquier software cuyo diseño contemple dar lugar a situaciones de disponibilidad, confidencialidad o integridad sin la autorización descriptiva del usuario legítimo (Shu et al., 2023).

La taxonomía del malware, responde de alguna manera a una clasificación de tipo estratégico en virtud de los vectores de ataque y los objetivos a alcanzar. Los virus, que necesitan un programa huésped para su replicación, contraponen el diseño de los gusanos, que se replican y se difunden de forma autónoma gracias a la propia red. Los troyanos logran apelar a la ingeniería social disfrazándose como aplicaciones de confianza y el ransomware utiliza el cifrado como técnica para tratar de obtener información para la extorsión.

La evolución del malware hacia las variantes polimórficas y metamórficas ha generado problemas para los sistemas de detección tradicionales. Las técnicas de ofuscación de código, cifrado dinámico y mutación automática son también ejemplos de adaptaciones evolutivas con la intención de sustraerse a la detección convencional (Gao et al., 2024).

Paradigmas de Detección: Limitaciones Tradicionales y Oportunidades Emergentes

Los enfoques clásicos de detección de malware se apoyan en técnicas que, si bien han sido eficaces de modo tradicional, son insuficientes frente a la sofisticación actual de las amenazas. La detección basada en firmas, la cual está dotada de la capacidad de detectar firmas de malware conocidas, se ve en desventaja de modo inherente a polimórficas y ataques de día cero; el análisis estático, el cual examina propiedades estructurales de forma que no ejecuta el código, proporciona información relevante pero es vulnerable a métodos de ofuscación y empaquetado; el análisis dinámico, aunque a priori pueda resultar más resistente ya que hace ejecutar las muestras en un entorno controlado, resulta en el mismo momento vulnerable a problemas de detección de sandboxes y a comportamientos evasivos (Tobiyama et al., 2016).

Por su parte, la adopción de arquitecturas de aprendizaje profundo en este ámbito supone un cambio de paradigma que permite superar las limitaciones de los enfoques tradicionales. Esas metodologías permiten el aprendizaje automático de representaciones para el aprendizaje sin que se precise de la ingeniería manual de características, resultando en una aplicación particularmente eficaz para el tratamiento de datos no estructurados como las secuencias binarias, flujos de ejecución y representaciones visuales de archivos ejecutables (Qiang et al., 2022).

La evidencia empírica indica que esos métodos suponen ventajas importantes para la detección y la reducción de falsos positivos (Falana et al., 2022). Sin embargo, su implementación práctica está condicionada por unos requerimientos computacionales elevados y unos largos procesos de entrenamientos, lo cual plantea interrogantes sobre la viabilidad de su uso en producción con limitaciones de recursos.

La promesa de esos enfoques es su que son más adaptables a las amenazas emergentes y ataques de día cero, aunque su efectividad final dependerá de la capacidad de actualizar los modelos conforme se va modificando el panorama de amenazas (Nayak et al., 2024).

La fundamentación teórica expuesta establece la base conceptual que permite la evaluación crítica de la aplicación de arquitecturas de aprendizaje profundo para detecciones de malware. Esta base conceptual es fundamental para poder interpretar los hallazgos empíricos y para poder evaluar tanto lo que ofrecen como las limitaciones que suponen estos enfoques emergentes en el contexto de la ciberseguridad moderna.

Metodología

La presente monografía se elaboró bajo la modalidad de revisión teórica con la finalidad de analizar el estado de la técnica en torno a la utilización de redes neuronales profundas para la detección de malware. Este tipo de revisión permite recopilar, indagar y sintetizar información de interés de fuentes científicas para así localizar avances, obstáculos y huecos en la línea de investigación (Hernández-Sampieri et al., 2014).

Con el objetivo de asegurar la exigencia y la pertinencia de la información reunida, se realizó la pertinente búsqueda robotizada en bases científicas de alta consideración; para ello se han aplicado los unos criterios de inclusión y exclusión previamente establecidos. A continuación, se narra el camino seguido que ha llevado a encontrar las referencias obtenidas y los elementos metodológicos más importantes:

Bases de datos consultadas

Para obtener la información se recurrió a las plataformas científicas de impacto elevado y de reconocimiento internacional: IEEE Xplore, ScienceDirect (Elsevier), SpringerLink, Scopus, ACM Digital Library, Google Académico. Esas bases permiten el acceso a publicaciones sometidas a revisión por pares, lo que resulta esencial para asegurar la validez académica de los trabajos a seleccionar (Gómez, 2019).

Palabras clave empleadas

La planificación de la estrategia de búsquedas se concreta a partir de términos clave, vinculados estrechamente con el objeto de estudio. Entre las que constituyen las principales palabras clave utilizadas figuran las siguientes: "malware detection", "deep learning", "convolutional neural network", "recurrent neural networks", "CNN" (red neuronal

convolucional), "RNN" (red neuronal recurrente), "adversarial malware", "hybrid model" (modelo híbrido) y "Android malware". La combinación de estas expresiones fue llevada a cabo con la ayuda de operadores booleanos para obtener unos determinados resultados y una determinada pertinencia temática.

Filtros aplicados

Estos filtros nos permitirán acotar el corpus de análisis:

- Idioma: Se consideraron prioritariamente los documentos publicados en lengua inglesa y bien en español siendo estos los idiomas que más se usan en el área de la ingeniería para la divulgación científica.

- Año de publicación: Se consideraron trabajos publicados entre 2016 y 2024 ya que en este periodo se ha producido un claro auge de la aplicación de modelos de deep learning en ciberseguridad (Mudzfirah et al., 2019).

- Tipo de documento científico: Se consideraron artículos científicos, tesis de máster, revisiones sistemáticas y actas de congresos académicos siempre que se encontraran correctamente indexadas o estuviesen soportadas por comités científicos.

Criterios de inclusión y exclusión

Con el objetivo de garantizar la calidad del material revisado, se establecieron los siguientes criterios:

Inclusión: Estudios centrados en la detección de malware mediante técnicas de aprendizaje profundo (CNN, RNN o híbridas). Además, investigaciones que emplearan datasets reconocidos, metodologías replicables y enfoques cuantitativos con métricas claras de

evaluación (como precisión, recall, F1-score, AUC). Y publicaciones con resultados relevantes y discusión crítica de hallazgos.

Exclusión: Documentos sin revisión por pares o sin respaldo académico. Además, estudios centrados únicamente en productos comerciales (antivirus) sin aporte técnico o metodológico. Y trabajos con información incompleta, sin metodología clara o sin reporte de resultados cuantificables.

Proceso de selección y análisis

De entrada, se habían identificado a más de 80 fuentes. Después y tras aplicar los filtros descriptos para ello y revisar los resúmenes y textos completos, se obtenían 34 estudios que se consideraron relevantes para su análisis detallado.

Se realizó la clasificación de los documentos según la tipología de red neuronal empleada (CNN, RNN o hibridos), el planteamiento metodológico adoptado y las métricas de desempeño reportadas mediante la creación de las fichas de análisis comparativo, lo que contribuyó a identificar patrones compartidos, corrientes que surgen en la actualidad y vacíos persistentes en la literatura.

Este tipo de enfoque metodológico realiza una revisión de la literatura y se adecua a las recomendaciones para las revisiones teóricas propuestas por Kitchenham (2007), quien argumenta la necesidad de que sean sistematizadas, trazables y replicables en estudios basados en literatura científica.

Ciberseguridad

La ciberseguridad, o seguridad informática, abarca un conjunto de métodos, estrategias y recursos enfocados en salvaguardar la confidencialidad, integridad y disponibilidad de la información y los sistemas informáticos ante posibles amenazas o ataques cibernéticos. Su objetivo principal es prevenir, detectar y responder a incidentes que comprometan los activos de información, incluyendo datos personales, recursos tecnológicos y aplicaciones críticas para las organizaciones (Cando-Segovia & Chicaiza, 2021). En el contexto actual, la ciberseguridad ha adquirido una importancia estratégica debido al aumento exponencial de las ciberamenazas y la creciente dependencia de las tecnologías de la información en la sociedad moderna. La ciberseguridad abarca una amplia variedad de prácticas, desde la protección contra virus y malware hasta la implementación de políticas de control de acceso, cifrado de datos y auditorías de seguridad. Estas prácticas buscan garantizar la continuidad operativa de las empresas y la protección de la información frente a actores maliciosos que intentan explotar vulnerabilidades en los sistemas (Cando-Segovia & Chicaiza, 2021).

Además, con la evolución de las tecnologías y la digitalización de los procesos, la ciberseguridad ha tenido que adaptarse rápidamente para abordar nuevas amenazas que surgen en entornos como el Internet de las Cosas (IoT), la computación en la nube y el uso de inteligencia artificial (IA). En este sentido, las estrategias de ciberseguridad modernas, basadas en tecnologías avanzadas como la inteligencia artificial y el aprendizaje automático, han demostrado ser más efectivas para identificar y mitigar ciberamenazas emergentes, especialmente cuando se combinan con políticas adecuadas y un entendimiento profundo de los factores humanos (Abrahams et al.2024).

La protección de la información no solo es un tema técnico, sino también ético, legal y organizacional, por eso las organizaciones, sin importar su tamaño, deben fomentar una cultura de seguridad que promueva la concienciación de todos los usuarios frente a los peligros que representan el mal uso de la tecnología. Muchas legislaciones nacionales e internacionales han establecido marcos regulatorios para asegurar el tratamiento adecuado de los datos y exigir a las empresas que adopten políticas estrictas de seguridad.

Por ejemplo, en la Unión Europea, el Reglamento General de Protección de Datos (RGPD) establece requisitos claros sobre cómo deben protegerse los datos personales y las consecuencias legales en caso de vulneraciones. En América Latina, países como Colombia, México y Argentina también han avanzado en la creación de leyes de protección de datos y ciberseguridad, aunque aún existen desafíos importantes en materia de implementación, coordinación interinstitucional y recursos técnicos (Cambo, 2019).

En el ámbito gubernamental y geopolítico, la ciberseguridad se ha posicionado como una prioridad. Las infraestructuras críticas como la energía, las telecomunicaciones, el transporte o los servicios financieros son constantemente blancos de ataques que pueden tener efectos devastadores. Por esta razón, muchos Estados han desarrollado estrategias nacionales de ciberseguridad, que buscan no solo proteger estos sectores, sino también fomentar la cooperación internacional, la innovación tecnológica y la resiliencia digital de la sociedad (Méndez, 2021).

En el entorno académico y de investigación, la ciberseguridad ha generado un campo interdisciplinario en constante expansión. Ingenieros, matemáticos, abogados, psicólogos y expertos en ciencias sociales trabajan juntos para entender las complejidades del

ciberespacio, identificar patrones delictivos y desarrollar soluciones innovadoras. Una de las líneas más prometedoras es el uso del aprendizaje profundo para detectar intrusiones, malware, fraudes o amenazas persistentes avanzadas (APT). Según Quirumbay Yagual et al. (2022), los algoritmos de redes neuronales ofrecen una alternativa eficaz para abordar amenazas sofisticadas, especialmente en escenarios donde los ataques se camuflan entre datos normales.

Es por eso que la ciberseguridad no puede ser vista únicamente como un problema técnico que afecta a las áreas de sistemas de las organizaciones. Es un componente fundamental de la seguridad moderna, que abarca aspectos económicos, sociales, políticos y éticos. La transformación digital, acelerada por fenómenos como la pandemia de COVID-19, ha incrementado la exposición de millones de personas a riesgos cibernéticos que pueden afectar su identidad, patrimonio e incluso su bienestar emocional. Frente a este panorama, se requiere un compromiso colectivo y continuo para fortalecer las defensas digitales, promover la educación en seguridad y fomentar una cultura de prevención que permita a las sociedades aprovechar al máximo los beneficios de la tecnología sin poner en riesgo su integridad.

Malware

El malware es un software malicioso diseñado para infiltrarse y dañar sistemas informáticos sin el consentimiento del usuario. Se utiliza para diversos fines, como robo de información, secuestro de sistemas, espionaje y destrucción de datos. Este tipo de software puede adoptar múltiples formas, y su evolución ha llevado a la creación de variantes complejas que utilizan técnicas avanzadas para evitar la detección (Anandhi et al., 2022).

El malware (abreviatura de *malware software* , o software malicioso en español) es un término general que hace referencia a cualquier tipo de software diseñado específicamente para causar daño a un sistema informático, robar información o interrumpir su funcionamiento normal. Este tipo de software malicioso puede presentarse de diversas formas, como virus, gusanos, troyanos, ransomware, adware y spyware, cada uno con objetivos y métodos de propagación diferentes. A pesar de que su nombre puede variar, todos los tipos de malware tienen en común la intención de explotar vulnerabilidades en los sistemas informáticos, redes y dispositivos de los usuarios para llevar a cabo actividades ilícitas o perjudiciales (Cando-Segovia & Chicaiza, 2021).

El malware se introduce en los sistemas a través de diversos vectores, como correos electrónicos maliciosos, enlaces infectados, sitios web comprometidos o aplicaciones de terceros no verificadas. Uno de los métodos más comunes es el uso de *phishing* , una táctica en la que los atacantes engañan a los usuarios para que ingresen sus credenciales o información personal a través de correos electrónicos que parecen legítimos pero que en realidad están diseñados para robar esos datos. Otra vía popular de propagación es el uso de vulnerabilidades conocidas en los sistemas operativos o aplicaciones, que los atacantes explotan para instalar malware de forma silenciosa, sin el conocimiento del usuario.

Una de las formas más peligrosas de malware es el **ransomware** , que cifra los archivos del sistema afectado y exige un rescate en criptomonedas para liberarlos. En los últimos años, el ransomware ha aumentado significativamente en número y sofisticación, afectando a individuos, empresas e incluso instituciones gubernamentales. En muchos casos, los ataques de ransomware no solo afectan a una computadora individual, sino que se propagan rápidamente a través de redes, infectando millas de dispositivos en una organización. Un ataque de ransomware puede tener consecuencias devastadoras para las empresas, provocando pérdidas financieras, daños a la reputación y la posible exposición de información confidencial.

Además del ransomware, otro tipo de malware ampliamente conocido son los **troyanos** , que se disfrazan de programas legítimos o inofensivos para engañar a los usuarios. Una vez que un troyano ha sido instalado en un sistema, puede realizar una variedad de acciones maliciosas, como robar información sensible, permitir el acceso remoto no autorizado a un atacante, o incluso crear una puerta trasera que permite a los ciberdelincuentes instalar otros tipos de malware. Los **gusanos** , por otro lado, son programas que se propagan por sí mismos a través de redes, aprovechando vulnerabilidades en los sistemas para duplicarse y distribuirse sin la intervención del usuario.

El **adware** y el **spyware** son otras formas de malware que, aunque menos destructivas en términos inmediatos, pueden tener efectos perjudiciales a largo plazo. El adware suele mostrar anuncios intrusivos sin el consentimiento del usuario, lo que puede generar molestias y ralentizar el rendimiento del sistema. Por su parte, el spyware recopila información del usuario, como hábitos de navegación, contraseñas o información financiera, y la transmite a terceros sin el conocimiento o consentimiento de la víctima. Ambos tipos de

malware se utilizan generalmente con fines comerciales, ya sea para promocionar productos o servicios específicos o para vender los datos recopilados a otras entidades.

El impacto del malware en la seguridad informática y en la vida cotidiana de los usuarios es considerable. Las organizaciones y los individuos pueden enfrentar pérdidas significativas debido al robo de datos confidenciales, la interrupción de servicios o el costo de reparar los sistemas comprometidos. En algunos casos, el malware se utiliza como parte de campañas más amplias de ciberespionaje o ciberguerra , donde los atacantes buscan obtener información sensible de gobiernos, empresas o competidores, con el fin de obtener ventajas económicas o políticas. La infraestructura crítica , como las redes eléctricas, los sistemas de transporte y los servicios financieros, es especialmente vulnerable a estos ataques, lo que pone de manifiesto la necesidad urgente de proteger estos sistemas.

Las estrategias de defensa contra el malware se centran en varios frentes, desde la prevención y detección temprana hasta la respuesta rápida ante incidentes. Una de las primeras líneas de defensa es la instalación de software antivirus y antimalware actualizado, que puede identificar y eliminar muchos tipos de malware antes de que causen daño. Sin embargo, los atacantes están desarrollando continuamente nuevas técnicas para evadir estas defensas, lo que hace que los sistemas de protección sean menos efectivos si no se mantienen al día. Además, la capacitación de los usuarios es fundamental, ya que muchas infecciones de malware ocurren debido a comportamientos imprudentes, como hacer clic en enlaces sospechosos o descargar archivos de fuentes no confiables.

Otra práctica recomendada es el cifrado de datos, especialmente en entornos empresariales donde la información confidencial debe estar protegida. El cifrado asegura

que, incluso si un atacante logra acceder a los datos, estos sean ilegibles sin las claves adecuadas. Además, las políticas de control de acceso y autenticación multifactor pueden ayudar a prevenir que los atacantes accedan a sistemas comprometidos, incluso si logran obtener credenciales válidas.

A medida que el malware sigue evolucionando, la respuesta ante incidentes debe incluir la capacidad de aislar y contener rápidamente los sistemas afectados para evitar la propagación a otros dispositivos y redes. Esto puede incluir el uso de tecnologías como sandboxing , que permite ejecutar archivos o programas en un entorno controlado y aislado para observar su comportamiento antes de permitir su ejecución en el sistema principal.

En cuanto a la prevención de malware, es crucial que las empresas y organizaciones implementen políticas de actualización regular de sistemas y software, ya que muchas infecciones ocurren debido a la explotación de vulnerabilidades conocidas que no han sido corregidas. El uso de firewalls y sistemas de detección y prevención de intrusiones (IDS/IPS) también puede bloquear intentos de acceso no autorizados, previniendo la instalación de malware desde el exterior.

Principales Categorías del Malware (Mundaca, 2020)

Virus: Es un software malicioso que altera el funcionamiento de un dispositivo al infectar archivos mediante un código maligno. Necesita que el usuario lo ejecute para activarse y propagarse a otros archivos y dispositivos.

Gusano: Similar al virus, pero con la capacidad de propagarse automáticamente sin intervención del usuario. No requiere modificar archivos existentes y se expande rápidamente por redes conectadas.

Spyware: Programa que se instala de forma oculta para recopilar información confidencial del usuario, como hábitos de navegación o datos personales, sin su conocimiento. Generalmente se oculta en segundo plano y evita ser detectado.

Troyano: Aparenta ser una aplicación legítima para engañar al usuario. Su objetivo principal es crear una puerta trasera en el sistema para permitir el acceso de otros programas maliciosos, generalmente sin ser detectado.

Adware: Programa que muestra publicidad de manera intrusiva. Aunque no siempre causa daño directo, puede afectar el rendimiento del dispositivo e instalar otros programas no deseados.

Ransomware: Secuestra los datos del dispositivo cifrándolos y solicita un rescate económico para devolver el acceso. Generalmente se propaga a través de correos electrónicos maliciosos o mensajes sospechosos y muestra una advertencia con el monto a pagar y el método de pago, como SMS, PayPal o bitcoins.

Criptominería: Se instala junto con software pirateado o programas maliciosos y utiliza los recursos del dispositivo infectado para minar criptomonedas sin que el usuario lo sepa, afectando el rendimiento y la vida útil del sistema.

Técnicas de Evasión del Malware

El malware moderno ha evolucionado para eludir las herramientas de seguridad tradicionales y sofisticadas mediante diversas técnicas de evasión. Estas técnicas permiten al malware adaptarse a diferentes entornos y ocultar sus intenciones maliciosas, dificultando su detección por parte de sistemas de seguridad automatizados. Entre las técnicas más utilizadas evaluadas por Planells 2023: Ofuscación de Código: Es una técnica que consiste en alterar la estructura del código fuente o del archivo binario del malware sin modificar su funcionalidad. Al modificar las cadenas de caracteres, eliminar comentarios y utilizar nombres de variables sin significado, la ofuscación hace que el análisis estático del código sea más complejo. Por ejemplo, el uso de técnicas como la codificación XOR y la reordenación de instrucciones en el ensamblado del malware son formas de ofuscación que complican la identificación de patrones maliciosos mediante análisis automatizados. Esta técnica es particularmente efectiva contra motores de detección basados en firmas. Cifrado: El cifrado se utiliza para ocultar el contenido del malware, protegiendo sus componentes de la detección y análisis. Oculta sus instrucciones en un formato ilegible que sólo se descifra en el momento de la ejecución. Una de las formas más comunes de cifrado en malware es el uso de criptografía simétrica, donde una clave se utiliza para cifrar y descifrar el contenido malicioso. Al combinar cifrado con otras técnicas como el empaquetado, el malware puede evitar ser detectado tanto por análisis estático como dinámico. Metamorfismo y Polimorfismo: Estas técnicas consisten en la modificación automática del código del malware en cada nueva infección o ejecución, alterando su apariencia sin cambiar su funcionalidad. - Malware Metamórfico: Reescribe su propio código cada vez que se ejecuta, generando una versión diferente de sí mismo con cada infección. Este método permite que el malware evada la detección basada en firmas, ya que las diferentes versiones no

comparten un patrón constante. - Malware Polimórfico: Cambia su código mediante un módulo de cifrado que altera la secuencia de bytes en cada ejecución, generando nuevas variantes con cada propagación. Al ser descifrado en la memoria del sistema, el malware polimórfico utiliza un decodificador dinámico que altera la estructura del código en tiempo real, haciendo imposible la detección mediante firmas convencionales. Empaquetado y Compresión: El empaquetado es una técnica que comprime el archivo ejecutable del malware, ocultando sus secciones maliciosas dentro de un contenedor comprimido. Al utilizar empaquetadores personalizados o empaquetadores múltiples, el malware se descomprime solo en el momento de la ejecución, evitando ser detectado durante el escaneo estático. Esta técnica también se combina con la compresión y el cifrado para añadir capas adicionales de protección contra la detección. Inyección de Código y Desvío de Flujos: El malware puede inyectar su código en procesos legítimos para ejecutar sus instrucciones dentro de programas de confianza, lo que le permite evadir los análisis dinámicos y basados en comportamientos. Técnicas como el process hollowing (relleno de procesos) o DLL injection (inyección de librerías dinámicas) permiten que el malware se ejecute de manera encubierta, alterando los flujos de ejecución sin ser detectado. Esto le otorga la capacidad de permanecer activo en la memoria del sistema sin ser identificado por análisis convencionales.

Ataques Basados en Entornos: Algunas variantes de malware están diseñadas para detectar entornos virtuales y sandboxes utilizados por los analistas de malware. Cuando el malware identifica que se está ejecutando en un entorno controlado, cambia su comportamiento para no activar sus rutinas maliciosas o se autodestruye, evitando ser analizado. Esta técnica se conoce como evasión basada en entornos y es cada vez más

utilizada por cibercriminales para evitar la detección durante el análisis dinámico. Estas técnicas de evasión permiten que el malware moderno se mantenga un paso adelante de las herramientas de detección y análisis, desafiando constantemente la capacidad de los sistemas de ciberseguridad para identificar y neutralizar amenazas. Para contrarrestar estas tácticas, las soluciones de seguridad deben combinar enfoques avanzados de análisis, como el uso de machine learning y redes neuronales profundas, para detectar patrones de comportamiento y características que los métodos tradicionales no pueden identificar.

Enfoques de Detección de Malware Tradicionales: Los métodos tradicionales de detección de malware se basan en técnicas como la detección basada en firmas y el análisis estático y dinámico, basados en Souri & Hosseini, 2018: - Detección Basada en Firmas: Este enfoque se centra en la identificación de patrones de código malicioso previamente conocidos. Consiste en comparar la firma digital de un archivo con una base de datos de firmas de malware. Aunque es eficiente para detectar malware conocido, resulta ineficaz frente a nuevas variantes o malware polimórfico, ya que no puede identificar amenazas que no coincidan con los patrones almacenados - Análisis Estático: Examina el código fuente o binario de un archivo sin ejecutarlo, permitiendo detectar propiedades sospechosas como llamadas a APIs o secuencias de instrucciones inusuales. Aunque es más rápido que el análisis dinámico, es vulnerable a técnicas de ofuscación y cifrado, lo que dificulta su efectividad para identificar amenazas avanzadas. - Análisis Dinámico: Este método ejecuta el archivo sospechoso en un entorno controlado (sandbox) para observar su comportamiento en tiempo real. A diferencia del análisis estático, el análisis dinámico puede detectar comportamientos maliciosos incluso cuando el código ha sido ofuscado. Sin embargo, este

enfoque consume muchos recursos y es costoso de implementar, lo que limita su uso en tiempo real y lo hace vulnerable a técnicas de evasión basadas en entornos.

Inteligencia artificial: Deep learning: La inteligencia artificial (IA) es un campo de la informática cuyo objetivo es desarrollar máquinas capaces de realizar tareas propias de la inteligencia humana, como resolver problemas y alcanzar objetivos de forma autónoma. En 1950, Alan Turing propuso las condiciones para evaluar si una máquina podía considerarse inteligente, y en 1956, John McCarthy acuñó el término "inteligencia artificial". Estos eventos sentaron las bases para el desarrollo de la IA, abriendo un nuevo campo de investigación que buscaba replicar y emular procesos cognitivos humanos en sistemas computacionales (Soria et al., 2022). Con el paso del tiempo, la IA ha avanzado gracias al aprendizaje automático (Machine Learning, ML), un enfoque que permite a los sistemas mejorar su rendimiento a partir del análisis de datos. ML se ha convertido en una técnica clave que permite a las máquinas aprender patrones, hacer predicciones y adaptarse con base en experiencias previas. Esto ha ampliado considerablemente el alcance de la IA, ya que ML permite abordar problemas complejos sin la necesidad de programar reglas explícitas para cada situación (Portela, 2022). El aprendizaje profundo (Deep Learning, DL), considerado una subdisciplina avanzada de ML, emplea redes neuronales profundas para procesar grandes cantidades de datos y mejorar su precisión a medida que se adquiere experiencia. DL ha demostrado un rendimiento superior en comparación con los métodos tradicionales de ML, especialmente en contextos que requieren el procesamiento de volúmenes masivos de datos, como en la ciberseguridad. Esta capacidad de análisis profundo convierte a DL en una herramienta poderosa para tareas que requieren la detección de patrones complejos y la toma de decisiones informadas.

En la práctica, la principal diferencia entre Machine Learning (ML) y Deep Learning (DL) radica en la forma en que procesan los datos y en su capacidad para identificar patrones complejos. Los métodos tradicionales de ML, como los árboles de decisión, los algoritmos de vecinos más cercanos (K-NN) o las máquinas de vectores de soporte (SVM), requieren una fase previa de ingeniería de características, en la cual expertos definen manualmente qué atributos del malware se analizarán, tales como el tamaño del archivo, la frecuencia de aparición de ciertos opcodes, el número de conexiones a puertos o los permisos solicitados por una aplicación. El modelo aprende a partir de estos atributos y construye reglas o hiperplanos para clasificar si un archivo es malicioso o benigno.

Por ejemplo, un algoritmo como Random Forest podría entrenarse para detectar malware analizando características tabuladas como: "¿el archivo solicita más de 5 permisos?", "¿se conecta a una IP externa al iniciar?", "¿modifica el sistema de arranque?". Con estas reglas, el modelo podría lograr una precisión aceptable en malware conocido, pero sería vulnerable frente a variantes que utilicen técnicas de evasión o modifiquen superficialmente estos atributos. (Fernández Khatiboun, 2019).

En contraste, los modelos de DL como las redes neuronales convolucionales (CNN) o las redes neuronales recurrentes (RNN) no necesitan que un experto diseñe las características, ya que pueden aprender directamente de los datos crudos, como secuencias de instrucciones máquina (opcodes), archivos binarios o flujos de red. Esto les permite encontrar patrones más complejos, incluso aquellos invisibles al ojo humano o a modelos simples.

Un ejemplo concreto de esta diferencia se observa en el estudio de Mudzfirah et al. (2019), donde se compararon modelos tradicionales de ML con una arquitectura híbrida de DL para la detección de malware Android utilizando el dataset Drebin. El modelo basado en MLP (un modelo clásico de ML) alcanzó una precisión de 94.73%, mientras que el modelo LSTM-CNN, que combina redes profundas, logró una precisión del 98.53%. Esta mejora significativa se debe a que el modelo DL fue capaz de identificar patrones espaciales y temporales en las secuencias de código malicioso, incluso sin haber visto antes ciertas variantes de malware. Esto evidencia que el proyecto es financieramente viable demuestra que, en contextos donde los atacantes modifican constantemente el comportamiento del código para evadir la detección, los modelos de DL ofrecen una capacidad de generalización y adaptación mucho mayor.

Con el fin de destacar las diferencias más relevantes entre los enfoques de Machine Learning y Deep Learning en el ámbito de la detección de malware, se incluye a continuación una tabla comparativa. En ella se sintetizan elementos clave como el tipo de procesamiento, la dependencia de ingeniería de características, el rendimiento obtenido y la eficacia frente a amenazas nuevas o sofisticadas. (Tabla 1)

Tabla 1

Tabla comparativa entre Machine Learning y Deep Learning aplicada a la detección de malware.

Características	Machine Learning (ML)	Deep Learning (DL)
Requiere ingeniería de características	Sí, manual y dependiente del experto	No, aprende automáticamente a partir de los datos
Capacidad de procesamiento	Moderada (datos estructurados)	Alta (datos complejos y no estructurados)
Precisión en detección de malware	Buena en escenarios simples	Superior, especialmente para malware ofuscado o avanzado

Tiempo de entrenamiento	Más rápido y menos costoso computacionalmente	Requiere más tiempo y recursos computacionales
Robustez frente a nuevas amenazas	Limitada (depende de las características definidas)	Alta (detecta patrones complejos y nuevas variantes)
Ejemplo de uso	Detección basada en permisos o llamadas al sistema	Detección basada en secuencias de opcodes o imágenes binarias

Las redes neuronales artificiales (RNA) tienen sus orígenes en los años 40, cuando Warren McCulloch y Walter Pitts desarrollaron los primeros modelos algorítmicos inspirados en la actividad del cerebro humano. Posteriormente, en 1957, Frank Rosenblatt introdujo el perceptrón, un modelo pionero para el reconocimiento de patrones. Aunque las RNA enfrentaron limitaciones debido a la capacidad computacional de la época, estos modelos experimentaron un resurgimiento en los años 2000, impulsados por mejoras tecnológicas y el acceso a grandes cantidades de datos, lo que facilitó su evolución y aplicación en diversos campos. (Thakur et al., 2021).

Hoy en día, las redes neuronales y el aprendizaje profundo se han convertido en componentes esenciales en la ciberseguridad. Estas tecnologías se utilizan para abordar problemas complejos, como la detección de intrusiones, el análisis de malware y la predicción de comportamientos anómalos. La combinación de redes neuronales en modelos de seguridad híbridos permite a los sistemas adaptarse y responder de manera inteligente a diversas amenazas cibernéticas, ofreciendo una solución eficiente y escalable para proteger entornos digitales en constante evolución (Quirumbay et al., 2022).

***Redes Neuronales Artificiales*:** Las redes neuronales artificiales forman parte de los sistemas de aprendizaje automático, una rama de la inteligencia artificial que busca emular, en términos simplificados, el funcionamiento del cerebro humano para resolver problemas

complejos. Estas redes se componen de nodos (neuronas) interconectados que procesan información de manera similar a las neuronas biológicas, permitiendo el aprendizaje y la adaptación a partir de la experiencia y los datos. Según Raschka (2015, Cap. 11), las redes neuronales convierten los datos en conocimiento mediante el aprendizaje supervisado, no supervisado o reforzado, facilitando la identificación de patrones y la clasificación de información. En el ámbito de la detección de malware, las redes neuronales han demostrado ser efectivas para abordar problemas que los métodos tradicionales no pueden resolver debido a la creciente complejidad y diversidad de las amenazas cibernéticas. Dentro de los fundamentos de las redes neuronales artificiales (ANN) son sistemas de procesamiento distribuidos inspirados en la estructura y funcionamiento del cerebro humano. Una ANN está compuesta por múltiples nodos o neuronas interconectadas, que se organizan en capas. Cada neurona procesa información proveniente de una o varias entradas y produce una salida que se transmite a otras neuronas. La capacidad de las ANN para aprender y adaptarse se logra mediante un proceso denominado entrenamiento, donde se ajustan los pesos sinápticos (la fuerza de las conexiones entre neuronas) para minimizar el error entre la salida esperada y la obtenida.

El aprendizaje en las ANN puede ser supervisado, no supervisado o de refuerzo. En el aprendizaje supervisado, la red se entrena con un conjunto de datos de entrada y sus respectivas salidas, ajustando los pesos hasta que la salida generada se ajuste al objetivo deseado. Por otro lado, en el aprendizaje no supervisado, la red recibe datos de entrada sin una referencia clara de salida, buscando patrones y estructuras inherentes en los datos (Palacios, 2023).

Las redes neuronales se pueden clasificar de acuerdo con varios criterios: Número de capas: - Monocapa: Contiene una sola capa de neuronas que proyectan las entradas directamente a una capa de neuronas de salida. - Multicapa: Posee una o más capas intermedias entre la entrada y la salida, llamadas capas ocultas, las cuales permiten a la red aprender representaciones más complejas (Benites Grado & Pernalete Lugo, 2022).

Tipo de conexiones:

No recurrentes: La propagación de las señales se realiza en un solo sentido, desde la capa de entrada hacia la capa de salida, sin retroalimentación.

Recurrentes: Poseen lazos de retroalimentación que conectan neuronas de diferentes capas, la misma capa o incluso de la misma neurona, permitiendo el almacenamiento de información a lo largo del tiempo y el modelado de secuencias temporales. Grado de conexión:

Totalmente conectadas: Cada neurona de una capa está conectada con todas las neuronas de la siguiente capa o de la capa anterior, dependiendo del tipo de red.

Parcialmente conectadas: No existe una conexión completa entre neuronas de las diferentes capas, lo que puede reducir la complejidad y mejorar la eficiencia computacional de la red.

Redes Neuronales Recurrentes

Las Redes Neuronales Recurrentes (RNN) son un tipo especializado de ANN que se caracterizan por su capacidad para procesar secuencias temporales de datos. A diferencia de las redes no recurrentes, las RNN permiten que la salida de una neurona en un momento dado se retroalimenta como entrada a la misma red, posibilitando el modelado de secuencias temporales y dependencias a lo largo del tiempo. Esta arquitectura las hace particularmente útiles en el análisis de secuencias, como las llamadas a API en la detección de malware (Mudzfirah et al., 2019). Dentro de las RNN, existen variaciones como las Long Short-Term Memory (LSTM), que incorporan celdas de memoria y gates (puertas) para controlar el flujo de información y retener datos relevantes durante periodos largos de tiempo (Figura 1). Estas puertas —de olvido, entrada y salida— regulan de forma dinámica qué información se descarta, cuál se actualiza y cuál se transmite como salida. Gracias a esta estructura, las LSTM son capaces de preservar el contexto en secuencias extensas, superando limitaciones clásicas como el desvanecimiento del gradiente, por eso son ampliamente utilizadas en la detección de intrusos y en la predicción de comportamientos anómalos en sistemas de ciberseguridad (Quirumbay et al., 2022).

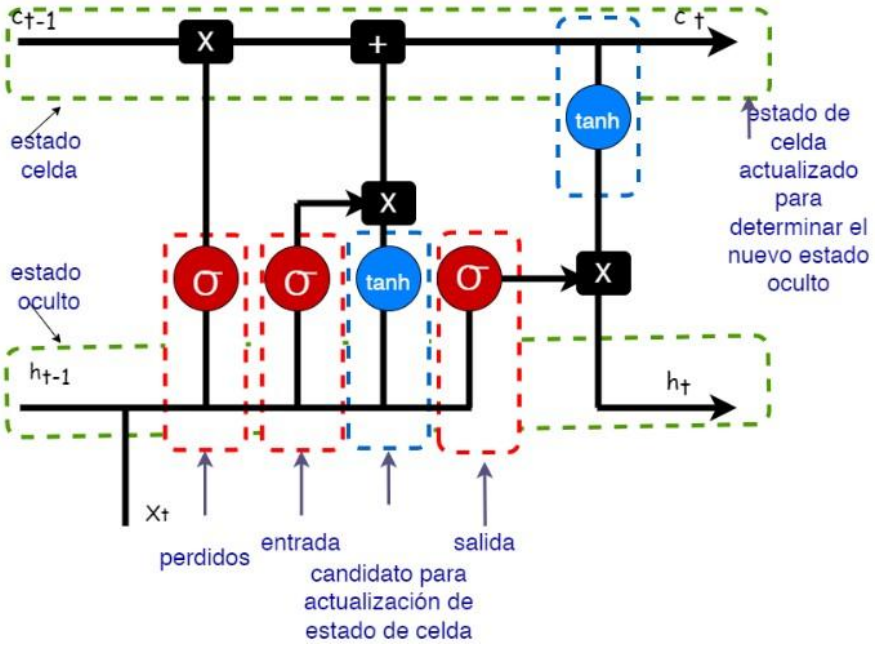
Jha et al. (2020) exploraron la aplicación de RNN en la detección de malware, enfocándose en cómo la elección de vectores de características afecta el rendimiento del modelo. En su estudio, utilizaron tres tipos de representaciones de características: codificación one-hot, vectores aleatorios y vectores Word2Vec, evaluando la capacidad de cada una para capturar información contextual de las secuencias de datos maliciosos. El uso de vectores Word2Vec resultó ser el más efectivo, permitiendo que la RNN alcance un área bajo la curva (AUC) significativamente superior a la de otras configuraciones. Este enfoque

se justifica por la capacidad de Word2Vec para representar relaciones semánticas entre instrucciones o llamadas a API en secuencias de código, lo cual es crucial para identificar comportamientos anómalos en el malware (Jha et al., 2020). Los autores concluyeron que las RNN, combinadas con representaciones semánticas avanzadas como Word2Vec, pueden mejorar considerablemente la detección de variantes desconocidas de malware, contribuyendo a una detección más robusta en entornos dinámicos. Por otro lado, un segundo estudio analizó el uso de LSTM, una variante de las RNN, para mejorar la precisión y eficiencia en la detección de malware en entornos IoT (Internet de las Cosas). Este enfoque híbrido combinó capas LSTM con un clasificador basado en memoria de corto y largo plazo, lo que permitió al modelo capturar tanto dependencias temporales como relaciones contextuales a lo largo de las secuencias de comportamiento del malware. Los autores lograron una precisión del 99.99% en la clasificación de conexiones maliciosas en dispositivos Android al emplear el optimizador NAdam, lo que demuestra la viabilidad de este enfoque en aplicaciones móviles. Además, destacaron la importancia de utilizar LSTM para mitigar el problema del desvanecimiento de gradiente, común en las RNN tradicionales, y para garantizar una detección continua y en tiempo real en sistemas con recursos limitados.

La implementación en dispositivos móviles refuerza la seguridad en tiempo real sin comprometer el rendimiento, lo que es particularmente relevante en redes IoT, donde la proliferación de dispositivos y conexiones expone a las redes a una amplia variedad de amenazas (Woźniak et al., 2021).

Figura 1

Arquitectura interna de una celda LSTM (Long Short-Term Memory)



Fuente: Tomada de Quirumbay et al., 2022.

Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (CNN) se han desarrollado principalmente para el procesamiento de imágenes, pero también han demostrado ser eficaces en la detección de malware al analizar archivos representados como imágenes. Las redes neuronales convolucionales (CNN) están diseñadas para trabajar con imágenes en 2D y 3D, y su arquitectura se compone principalmente de tres tipos de capas: convolucionales, de reducción (pooling) y densas (fully connected). Las capas convolucionales se encargan de extraer características locales de los datos, mientras que las capas de reducción (como el max pooling) reducen la dimensionalidad y la complejidad del modelo. Estas dos primeras conforman la fase de extracción de características, mientras que las capas densas (fully connected) se encargan de la fase de clasificación (Figura 2). En la detección de malware, las CNN permiten identificar patrones visuales que corresponden a firmas específicas de malware, lo que las hace útiles para detectar variantes polimórficas y metamórficas (Palacios, 2023). Al combinar las CNN con las RNN, se pueden obtener arquitecturas híbridas que permiten el análisis simultáneo de características espaciales y temporales, mejorando la detección de malware avanzado y la reducción de falsos positivos (Mudzfirah et al., 2019). Pinhero et al. (2021) propusieron un método basado en la visualización de malware y su clasificación mediante redes neuronales profundas. Este enfoque utiliza técnicas de visualización de malware para superar las limitaciones de la extracción de características en métodos tradicionales. A través de la conversión de muestras de malware en imágenes en escala de grises, imágenes RGB y representaciones de modelos de Markov, el estudio demuestra que es posible detectar y clasificar malware con una precisión notable, alcanzando un F-measure de hasta 99.97%. El uso de imágenes, junto con el filtro Gabor para la extracción de características, permitió clasificar 20,199 muestras de malware en 12

arquitecturas de redes neuronales diferentes. Además, el artículo resalta cómo la visualización de malware en imágenes permite comparar familias de malware de manera más efectiva, independientemente del tipo de archivo y con menor conocimiento específico del dominio en comparación con otros enfoques basados en ingeniería de características.

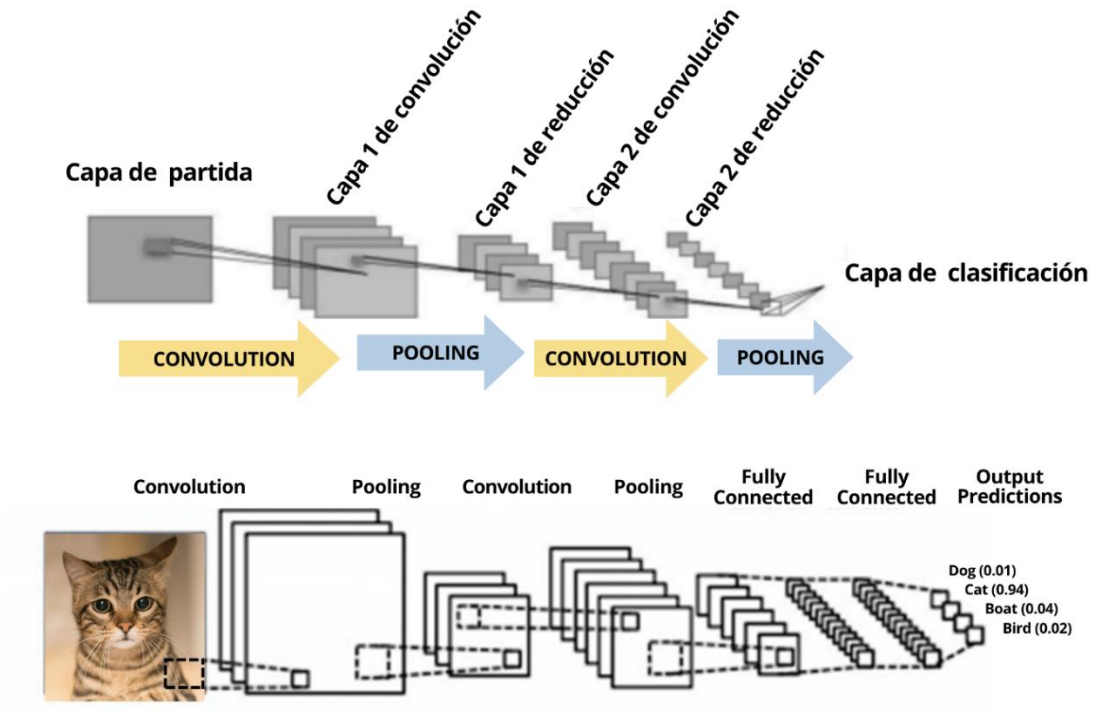
Por su parte, Vasan et al. (2020) desarrollaron un modelo denominado IMCFN (Image-based Malware Classification using Fine-tuned Convolutional Neural Network Architecture) para la clasificación de variantes de familias de malware. Utilizando imágenes en color generadas a partir de binarios de malware, y aplicando técnicas de fine-tuning en la arquitectura de red neuronal convolucional, lograron mejorar la precisión en la detección de malware en comparación con arquitecturas convencionales como VGG16, ResNet50 y Google InceptionV3. Los experimentos se realizaron con dos conjuntos de datos: Malimg (9,435 muestras) y un dataset de malware IoT-Android (14,733 malware y 2,486 muestras benignas). Los resultados mostraron que el modelo propuesto alcanzó una precisión del 98.82% en el dataset Malimg y más del 97.35% en el dataset IoT-Android, demostrando su eficacia en la detección de código oculto, malware ofuscado y variantes de familias de malware con tiempos de ejecución reducidos.

Finalmente, Anandhi et al. (2022) presentaron un enfoque de detección de malware utilizando redes neuronales convolucionales (CNN) para superar las limitaciones de la detección basada en firmas. Este estudio se centró en la clasificación de malware a través de imágenes generadas a partir de binarios de malware, implementando una arquitectura CNN optimizada. Los resultados experimentales, utilizando un conjunto de datos de 10,860 muestras de malware del Microsoft Malware Classification Challenge (Big 2015), demostraron una precisión del 98.2%. El enfoque se destaca por su capacidad para detectar

malware de día cero y variantes desconocidas con alta precisión. Además, el modelo propuesto se evaluó en diferentes arquitecturas de CNN, como LeNet y AlexNet, comparándolos con otros métodos convencionales, concluyendo que la visualización de malware como imágenes y su posterior clasificación con CNN no solo es efectiva, sino también eficiente en cuanto al tiempo de procesamiento, proponiendo su uso en entornos de ciberseguridad para la clasificación de malware avanzado y el desarrollo de sistemas de defensa proactivos.

Figura 2

Arquitectura de una Red Neuronal Convolucional (CNN)



Nota. La imagen ilustra el funcionamiento de una Red Neuronal Convolucional (CNN), donde una imagen pasa por capas de convolución y reducción que extraen y simplifican características. Luego, capas completamente conectadas procesan esa información para clasificar la imagen según su contenido, como en este caso, donde se identifica un gato. *Fuente:* Adaptada de Lubinus Badillo, F., et al., 2021.

Combinación de CNN y RNN

La combinación de CNN y RNN ha demostrado ser altamente efectiva en la detección de malware, al aprovechar la capacidad de las CNN para reducir la dimensionalidad y extraer características espaciales, y de las RNN para modelar la información secuencial (Mudzfirah et al., 2019). Mudzfirah Abdul Halim et al. (2019) desarrollaron y evaluaron dos modelos híbridos que combinan redes neuronales convolucionales (CNN) y redes Long Short-Term Memory (LSTM), con el propósito de mejorar la detección de malware al abordar simultáneamente la pérdida de información espacial y secuencial que generan enfoques tradicionales como el modelo Bag of Words (BOW). Este último, aunque comúnmente utilizado para representar patrones de malware, presenta limitaciones importantes al eliminar la estructura de las secuencias, lo que compromete su capacidad para capturar relaciones contextuales entre características.

Para superar esta deficiencia, los autores propusieron dos arquitecturas distintas: una en la que la red LSTM se coloca antes de la CNN (modelo LSTM-CNN), y otra en la que se invierte este orden (modelo CNN-LSTM). En la primera configuración, el modelo aplica inicialmente una capa LSTM que analiza las secuencias de características generadas a partir del BOW, permitiendo así modelar dependencias temporales y aprender representaciones contextuales más profundas. A continuación, se aplica una CNN que actúa sobre la salida de la LSTM para extraer patrones espaciales relevantes y reducir la variación en las representaciones, antes de pasar a la capa de clasificación final mediante un perceptrón multicapa (MLP). En la segunda configuración, se invierte el flujo: primero se aplica una CNN sobre los vectores generados por BOW para realizar una reducción dimensional

temprana, y luego se utiliza la LSTM para capturar la secuencialidad restante en los datos procesados.

Ambos modelos fueron entrenados y evaluados utilizando el conjunto de datos Drebin donde los resultados experimentales mostraron que el modelo LSTM-CNN obtuvo una precisión del 98.53%, superando significativamente al modelo CNN-LSTM, que alcanzó una precisión del 96.76%, así como a los modelos individuales de LSTM (95.90%), CNN (87.91%) y MLP (94.73%). Esta diferencia de rendimiento no solo evidencia la efectividad de los modelos híbridos, sino también la importancia crítica del orden de las capas en su arquitectura.

Por otro lado, Jeon y Moon (2020) presentaron un modelo basado en una red recurrente convolucional (CRNN) que combina autoencoders convolucionales y RNN dinámicas para la detección de malware utilizando secuencias de opcode. En su trabajo, los autores describen un proceso que comienza con la extracción de secuencias de opcode de archivos ejecutables, las cuales son transformadas en representaciones compartidas a través de un autocorrector convencional (OCAE). Posteriormente, estas secuencias comprimidas son procesadas por una RNN dinámica para la clasificación de malware. Los resultados experimentales mostraron que el modelo propuesto logró una tasa de detección verdadera del 95% y un área bajo la curva (AUC) de 0.99. Aunque la precisión del modelo (96%) fue ligeramente inferior a otros métodos como el K-Nearest Neighbor (KNN), su capacidad para reducir falsos positivos y capturar características secuenciales lo hace altamente efectivo para la detección de malware en entornos reales. Los autores concluyeron que el uso de combinaciones de técnicas como CNN y RNN es fundamental para abordar los desafíos de la variabilidad de los datos y la detección de patrones en secuencias largas.

Más recientemente, Barcenas-Medina y Alcaraz-Cancio (2025) propusieron un sistema de detección de amenazas zero-day en entornos empresariales, basado también en la combinación de CNN y RNN, aplicado específicamente al tráfico de red. Su enfoque integró capas convolucionales para extraer patrones espaciales y capas LSTM para modelar las dependencias temporales de los datos, obteniendo una precisión del 96.84%, un recall del 96.5% y un F1-score del 96.5%. Estos resultados confirman que los modelos híbridos no solo son eficaces en la clasificación de malware clásico, sino que también representan una solución escalable y robusta para amenazas avanzadas como los ataques zero-day. A diferencia de los sistemas tradicionales basados en firmas o reglas estáticas, este modelo demostró una capacidad destacada para adaptarse dinámicamente a patrones desconocidos, reduciendo los falsos positivos a solo el 0.18%. Además, los autores enfatizan que esta arquitectura es viable para entornos empresariales en producción, gracias a su diseño modular y a su capacidad de procesamiento en tiempo real. La comparación del modelo CNN-RNN con técnicas como Random Forest y LightGBM mostró un rendimiento competitivo, destacándose principalmente por su balance entre precisión, sensibilidad y velocidad de inferencia, factores clave para su implementación práctica en sistemas de ciberseguridad.

Estos resultados evidencian cómo la combinación de CNN y RNN permite abordar de manera más eficaz los desafíos que plantea la detección de malware, especialmente frente a variantes avanzadas que escapan a los métodos tradicionales. Para ofrecer una visión comparativa del desempeño de los modelos híbridos analizados, se presenta a continuación una tabla que resume las principales métricas de evaluación reportadas en los estudios mencionados (precisión, recall, F1-score y AUC)(Tabla 2).

Tabla 2

Tabla comparativa de modelos híbridos (CNN+RNN)

Estudio	Modelo	Precisión (%)	Recall (TPR) (%)	F1-score (%)	AUC
Mudzfirah et al. (2019)	LSTM-CNN	98.53	No reportado	No reportado	No reportado
Mudzfirah et al. (2019)	CNN-LSTM	**96**.76	96.7*	96.7*	No reportado
Jeon & Moon (2020)	CRNN (OCAE + DRNN)	96.2	95.7	No reportado	0.99
Barcenas-Medina & Alcaraz-Cancio (2025)	CNN + RNN (Zero-Day)	96.84	96.5	96.5	0.982

Nota: En esta tabla se compara los modelos híbridos (CNN+RNN) con el reporte de diferentes métricas de desempeño.

Discusión

La revisión teórica que se ha llevado a cabo hace evidenciar el claro avance de las redes neuronales para la detección de malware, principalmente en lo que respecta la utilización de arquitecturas profundas como CNN, RNN y sus combinaciones híbridas. Para ilustrar esto, han sido desarrollados en los capítulos anteriores un análisis al detalle de cada una de estas arquitecturas siendo destacadas sus fortalezas y debilidades.

Como se ha mostrado, las redes neuronales convolucionales (CNN) han sido utilizadas para la detección de patrones visuales y el modelo estructural del malware. Su capacidad para gestionarse con imágenes y representaciones binarias hace que el modelo correspondiente de clasificación de muestras de malware sea bastante preciso.

Sin embargo, también se evidencian ciertas limitaciones para modelizar tanto las relaciones secuenciales como las relaciones temporales y, en ocasiones, su ejecución requiere importantes cargas computacionales las cuales limitan su uso a dispositivos con escasos medios de procesamiento. Las redes neuronales recurrentes (RNN), en especial las variantes LSTM, se muestran efectivas modelizando las secuencias temporales del comportamiento malicioso, por ejemplo, en el uso de llamadas a APIs o flujos de ejecución, aunque su entrenamiento puede ser arduo y costoso y sus modelos son bastante vulnerables a la calidad y la estructura del conjunto de datos que se ha utilizado. En lo que respecta a los modelos híbridos que combinan CNN y RNN, se ha evidenciado una clara ventaja al integrar las ventajas de ambos modelos.

Dicha combinación conlleva a una detección más robusta ante amenazas avanzadas tales como el malware polimórfico, los ataques zero-day, etc. Estudios recientes muestran

métricas de performance y recall mayores y su mayor capacidad de adaptación a entornos reales. A pesar de ello, la mayoría de las propuestas han sido validadas en entornos simulados o usando datasets pequeños, lo que nos deja la necesidad abierta de probar estos modelos en entornos de producción, en especial en sistemas críticos o en sistemas corporativos. Ese sentido, se identifican algunas contradicciones notables. Por ejemplo, existen trabajos que reportan métricas de performance muy altas sin especificar el tipo de datos o de fuentes utilizadas para su verificación, lo que a veces hace que la replicación de los resultados sea poco viable.

Por otro lado, si bien se destaca las ventajas de los modelos híbridos en la automatización, pocos trabajos analizan la aplicabilidad de la computación de estos sistemas, que es uno de los aspectos clave para su aplicación práctica. Otra de las carencias observadas es que existe una exploración escasa sobre este tipo de técnicas en el entorno industrial (p.e. sistemas SCADA) o en plataformas embebidas. También se detecta que hay poco desarrollo de arquitecturas ligeras que puedan ser ejecutadas en dispositivos con recursos restringidos en energía o procesamiento, áreas muy presentes en el IoT. Al final de este análisis, se plantea una clasificación funcional para agrupar los modelos revisados.:

- **Modelos visuales (CNN puros):** orientados a la detección basada en estructuras fijas o visuales.
- **Modelos secuenciales (RNN/LSTM):** diseñados para analizar patrones temporales y de comportamiento.
- **Modelos híbridos (CNN + RNN):** integran ambas capacidades, mostrando un mejor rendimiento general en escenarios complejos.

En última instancia, se recomienda definir modelos explicables y eficientes que incluyan mecanismos de atención o de destilación del conocimiento. Tales estrategias permitirían mejorar la comprensión de decisiones, el uso de los recursos y ayudar a las aplicaciones de estas tecnologías en los sistemas reales de defensa cibernética.

Conclusiones

Por medio de la presente monografía, hemos demostrado que el empleo de las redes neuronales profundas es una excelente opción como medio de abordar las limitaciones propias de las técnicas tradicionales que se emplean para la detección de malware. En comparación con los métodos que están fundamentados en firmas, en reglas estáticas o en análisis heurísticos, las redes neuronales (especialmente las CNN y RNN) tienen la capacidad de reconocer patrones complejos en ficheros o secuencias de comportamiento malicioso, también en los que han sido creados con el fin de contrarrestar las herramientas propias de los métodos clásicos. El valor añadido que ofrecen estas tecnologías reside en el aprendizaje automático, la generalización ante nuevas amenazas, así como la reducción de los falsos positivos. Las CNN son especialmente útiles en el análisis de ficheros como imágenes con el objetivo de localizar firmas visuales del malware, al paso que las RNN son competentes en el análisis de secuencias temporales como las que resultan de los registros de actividad o de las llamadas a API. Los modelos híbridos de CNN-RNN han dado buenos resultados, proporcionando las ventajas espaciales y temporales en una única arquitectura, obteniendo tasas de precisión superiores al 96% en los estudios revisados. También se observa que las arquitecturas a partir de transformar han sido propuestas como técnicas prometedoras para la detección avanzada y también para poner de manifiesto técnicas alternativas aplicables al procesamiento de patrones complejos específicamente para la detección de ataques malware.

Aun así, el análisis crítico de la literatura ha puesto de relieve importantes contradicciones y vacíos metodológicos que limitan la validez de los hallazgos. Así, estudios diferentes reportan niveles de precisión muy altos sin ofrecer un número suficiente

de detalles sobre los datasets o las propias metodologías de validación, además de las condiciones experimentales que no son las mismas, lo que sin duda lleva a tener una dificultad para compartir las propuestas. La ausencia de transparencia contrasta con trabajos más serios, que, si bien son muy extensos, tampoco permiten la replicabilidad total de los resultados obtenidos, ya que carecen de código fuente y/o datasets públicos. Se observan huecos también en relación a las amenazas actuales. Hay una extensa falta de estudio del malware específico para sistemas embebidos, o para entornos críticos industriales (ICS/SCADA) que puedan verse afectados dramáticamente por un ataque, y la aplicabilidad del modelo (elemento capital para ser adoptado en compañías y organizaciones gubernamentales) ha sido puesta de manifiesto sólo en una escasa variedad de los trabajos estudiados. Estos modelos también tienen limitaciones prácticas severas, ya que se requieren grandes cantidades de datos etiquetados para su entrenamiento, cosa que no está de manera habitual en todas las organizaciones. Requieren, de recursos computacionales elevados, lo que provoca barreras para su implementación en entornos en tiempo real y/o con limitaciones como pueden ser los TI. Su falta de interpretabilidad, así como su sensibilidad a ataques adversarios, se convertirán en obstáculos importantes para la adopción segura de las CNNs en entornos críticos.

Estos retos invitan a continuar explorando vías de optimización de modelos que sean ligeros, explicables y resistentes a manipulaciones. También invita, y en gran medida se hace necesario, a ir generando datasets públicos, actualizados y balanceados en pro del entrenamiento de redes neuronales en situaciones más realistas.

La principal recomendación es que futuras investigaciones continúen explorando arquitecturas híbridas más eficientes que conjuguen CNNs ligeros como medio de

extracción de características visuales, y mecanismos de atención en la etapa RNN como método de procesamiento de secuencias. Esta opción pudiera ser capaz de conjugar de mejor forma precisión, aplicabilidad y criterios de eficiencia computacional, las cuales son fundamentales para su implementación práctica. En segundo lugar, sería interesante ir desarrollando técnicas de defensa adversarial robustas y explorar su aplicación práctica en el ámbito real como por ejemplo el colombiano, donde el aumento exponencial de los ciberataques ha deparado la necesidad de soluciones tecnológicas que no sean solo válidas y eficaces, sino que también sean adecuadas a los recursos del país y a sus particularidades del panorama de amenazas.

Referencias

Abrahams, T. O., Ewuga, S. K., Dawodu, S. O., Adegbite, A. O., & Hassan, A. O. (2024). A review of cybersecurity strategies in modern organizations: Examining the evolution and effectiveness of cybersecurity measures for data protection. *Computer Science & IT Research Journal*, 5(1), 1–25. https://doi.org/10.51594/csitrj.v5i1.699

Ahmad, M. (2020). Malware in computer systems: Problems and solutions. International *Journal on Informatics for Development*, 9(1), 20–30. https://doi.org/10.14421/ijid.2020.09101

Anandhi, V., Vinod, P., Menon, V. G., & Aditya, K. M. (2022). Performance evaluation of deep neural network on malware detection: Visual feature approach. *Cluster Computing*, 25, 4601–4615. https://doi.org/10.1007/s10586-022-03702-3

Barcenas-Medina, A.-L., & Alcaraz-Cancio, A.-J. (2025). Identificación de amenazas zero-day en entornos empresariales con modelos de aprendizaje profundo. *Revista de Investigación Científica, Tecnológica e Innovación (RICT),* 3(5), 46–54. https://doi.org/10.5281/zenodo.15149234

Benites Grado, PC, & Pernalete Lugo, J. (2022). Redes Neuronales Artificiales: Historia y Actualidad. *Diario de Mount Scopus*, 2(3), 35-54. https://doi.org/10.17613/1vcs-vc87

Cámara Colombiana de Informática y Telecomunicaciones. (2024, 16 de Abril). Colombia sufrió 12.000 millones de intentos de ciberataques en 2023 según reporte de Fortinet.

Cambo, O. M. (2019). Análisis al Reglamento General de Protección de Datos en la Unión Europea: Un vistazo a la actualidad de la era digital. *La Revista de Derecho*, 40, 93–104.

Cando-Segovia, M. R., & Chicaiza, R. P. M. (2021). Prevención en ciberseguridad: Enfocada a los procesos de infraestructura tecnológica. *3C TIC: Cuadernos de Desarrollo Aplicados a las TIC*, 10(1), 17–41.

Catak, F., Ahmed, J., Sahinbas, K., & Khand, Z. (2020). Data augmentation based malware detection using convolutional neural networks. PeerJ Computer Science, 7. https://doi.org/10.7717/peerj-cs.346

Chen, C.-M., Wang, S.-H., Wen, D.-W., Lai, G.-H., & Sun, M.-K. (2019). Applying convolutional neural network for malware detection. *In 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)* (pp. 1–5). https://doi.org/10.1109/ICAwST.2019.8923568

Falana, O., Sodiya, A., Onashoga, S., & Badmus, B. (2022). Mal-Detect: An intelligent visualization approach for malware detection. J. King Saud Univ. Comput. Inf. Sci., 34, 1968-1983. https://doi.org/10.1016/j.jksuci.2022.02.026

Fernández Khatiboun, A. (2019). *Machine learning en ciberseguridad* [Trabajo de fin de máster, Universitat Oberta de Catalunya]. Repositorio Institucional O2. http://hdl.handle.net/10609/97546

Gao, C., Du, Y., Lan, Q., Chen, J., & Wu, J. (2024). A new adversarial malware detection method based on enhanced lightweight neural network. Comput. Secur., 147, 104078. https://doi.org/10.1016/j.cose.2024.104078

Gibert, D. (2016). *Convolutional neural networks for malware classification* [Master's thesis, Universitat de Barcelona, Universitat Rovira i Virgili, Universitat Politècnica de Catalunya].

Gómez, R. (2019). Cómo elaborar una revisión sistemática de la literatura científica: Fundamentos y aplicación práctica. Universidad de La Rioja.

Han, W., Xue, J., Wang, Y., Huang, L., Kong, Z., & Mao, L. (2019). MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. *Computers & Security*, 83, 208–233. https://doi.org/10.1016/j.cose.2019.02.007

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). Metodología de la investigación (6.ª ed.). McGraw-Hill Education.

Jeon, S., & Moon, J. (2020). Malware-detection method with a convolutional recurrent neural network using opcode sequences. *Information Sciences*, 512, 1212–1225. https://doi.org/10.1016/j.ins.2020.05.026

Jha, S., Prashar, D., Long, H. V., & Taniar, D. (2020). Recurrent neural network for detecting malware. *Computers & Security*, 99, 102037. https://doi.org/10.1016/j.cose.2020.102037

Kitchenham, B. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering (Version 2.3). EBSE Technical Report. Keele University and Durham University Joint Report. https://www.cs.waikato.ac.nz/~bernhard/Teaching/CS507/papers/Kitchenham-SystematicReviewGuidelines.pdf

Lubinus Badillo, F., Rueda Hernández, C. A., Narváez, B. M., & Arias Trillos, Y. E. (2021). Redes neuronales convolucionales: un modelo de Deep Learning en imágenes diagnósticas. Revisión de tema. *Revista Colombiana de Radiología*, *32(3)*, 5591–5599. https://doi.org/10.53903/01212095.161

M., G., & Sethuraman, S. (2023). A comprehensive survey on deep learning based malware detection techniques. Comput. Sci. Rev., 47, 100529. https://doi.org/10.1016/j.cosrev.2022.100529

Méndez, A. E. L. (2021). Propuesta de estrategias de seguridad cibernética: Aproximaciones teórico–prácticas hacia el aprestamiento en países latinoamericanos. *Dominio de las Ciencias*, 7(1), 1186–1207.

Mudzfirah, A. H., Abdul Rahman, M. R., Shahrin, S., & Syazril, S. A. (2019). Hybrid deep learning models for malware detection: LSTM-CNN and CNN-LSTM approaches. Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI), 1–6. https://doi.org/10.1109/ICCCI.2019.8822113

Mundaca, R. (2020, 7 de Julio*). Malware y otros "ware" que te hacen sufrir y cómo debes cuidarte.* Tecnologías.uchile.cl. https://tecnologias.uchile.cl/malware-y-otros-ware/

Nayak, S., Kurup, S., & J, A. (2024). Malware Detection Employing Deep Neural Networks. ICACCS 2024, 1, 44-49. https://doi.org/10.1109/ICACCS60874.2024.10717146

Palacios, A. G. V. (2020). *Detección de malware a través de redes neuronales* [Tesis de licenciatura, Universidad Tecnológica Centroamericana UNITEC]. Repositorio

UNITEC. https://repositorio.unitec.edu/items/1e9b1070-836a-4269-861b-a2b2711782e4

Pinhero, A., Anupama, M. L., Vinod, P., Visaggio, C. A., Aneesh, N., Abhijith, S., & AnanthaKrishnan, S. (2021). Malware detection employed by visualization and deep neural network. *Computers & Security, 105*, 102247. https://doi.org/10.1016/j.cose.2021.102247

Planells García, G. (2023). *Análisis de técnicas de evasión usadas por malware para su orquestación en entornos aislados de ejecución* [Tesis de maestría, Universitat Politècnica de València].

Policía Nacional de Colombia. (2023, 1 de Junio). Los delitos cibernéticos se han reducido en el 2023: Policía Nacional. Radio Nacional de Colombia. https://www.radionacional.co/actualidad/delitos-ciberneticos-en-colombia-estadisticas-actuales

Portela García-Miguel, S. (2022). *Panorama de la inteligencia artificial en el dominio de la ciberseguridad*. *RUIDERAe: Revista de Unidades de Información*, (19), 1–10.

Qiang, W., Yang, L., & Jin, H. (2022). Efficient and Robust Malware Detection Based on Control Flow Traces Using Deep Neural Networks. Comput. Secur., 122, 102871. https://doi.org/10.1016/j.cose.2022.102871

Quirumbay Yagual, D. I., Castillo Yagual, C. A., & Coronel Suárez, I. A. (2022). Una revisión del aprendizaje profundo aplicado a la ciberseguridad. *Revista Científica y Tecnológica UPSE (RCTU), 9(1)*, 57–65.

Raschka, S. (2015). Python machine learning. Packt Publishing Ltd.

Rhode, M., Burnap, P., & Jones, K. (2017). Early Stage Malware Prediction Using Recurrent Neural Networks. Comput. Secur., 77, 578-594. https://doi.org/10.1016/j.cose.2018.05.010

Shu, L., Dong, S., Su, H., & Huang, J. (2023). Android Malware Detection Methods Based on Convolutional Neural Network: A Survey. IEEE Trans. Emerging Topics Comput. Intell., 7, 1330-1350. https://doi.org/10.1109/TETCI.2023.3281833

Soria Olivas, E., Rodríguez Belenguer, P., García Vidal, E., Vaquer, F., Camisón, J. V., & Vila Tomás, J. (2022). *Inteligencia artificial: Casos prácticos con aprendizaje profundo*. Grupo Editorial RA-MA.

Souri, A., & Hosseini, R. (2018). A state-of-the-art survey of malware detection approaches using data mining techniques. *Human-Centric Computing and Information Sciences, 8(3*). https://doi.org/10.1186/s13673-018-0125-x

Thakur, A., & Konde, A. (2021). Fundamentals of neural networks. *International Journal for Research in Applied Science and Engineering Technology, 9*(8), 407–426. https://doi.org/10.22214/ijraset.2021.37362

Tobiyama, S., Yamaguchi, Y., Shimada, H., Ikuse, T., & Yagi, T. (2016). Malware Detection with Deep Neural Network Using Process Behavior. COMPSAC, 2, 577-582. https://doi.org/10.1109/COMPSAC.2016.151

Trellix. (2024). Trellix Advanced Research Center Threat Report: June 2024. https://www.trellix.com/advanced-research-center/threat-reports/june-2024

Vasan, D., Alazab, M., Wassan, S., Naeem, H., Safaei, B., & Qin, Z. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network

architecture. *Computer Networks, 171,* 107138. https://doi.org/10.1016/j.comnet.2020.107138

Woźniak, M., Siłka, J., Wieczorek, M., & Alrashoud, M. (2021). Recurrent neural network model for IoT and networking malware threat detection. *IEEE Transactions on Industrial Informatics, 17(8),* 5583–5594.https://doi.org/10.1109/TII.2020.3021689

Yadav, P., Menon, N., Ravi, V., Vishvanathan, S., & Pham, T. (2022). EfficientNet convolutional neural networks-based Android malware detection. Comput. Secur., 115, 102622. https://doi.org/10.1016/j.cose.2022.102622