ISBN: 978-628-96613-0-9. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LEIRD2024.1.1.237

Effectiveness of Machine Learning Models in Intrusion Detection in Information Systems and Their Applicability in the Context of Entrepreneurship and Innovation: A Systematic Literature Review

Daniel Ivan Llontop Alama, Estudiante de Ingeniería de Sistemas e Informática[1], Christian Abraham Dios-Castillo, Dr. en Administración de la Educación[1]
[1]Universidad Tecnológica del Perú, Perú, 1525293@utp.edu.pe, c16763@utp.edu.pe

Abstract*– As technology has been advancing and the world is becoming more and more digitalized, the use of information systems in companies has become a key point for their development, and with it, security has become more relevant than ever. The purpose of this research is to explore Machine Learning models and their effectiveness in intrusion detection, evaluating their impact and applicability in the context of entrepreneurship and innovation. Therefore, 50 papers obtained from the Scopus database in the period between 2015 and 2024 focused on the use of Machine Learning models for intrusion detection were thoroughly analyzed. This review covers aspects such as: the most common types of attacks, most used dataste, most studied Machine Learning models and their classification, and finally their effectiveness in intrusion detection. The results showed that the 58 Machine Learning models identified present a minimum of 79.00% effectiveness and a maximum of 99.99% effectiveness. The conclusion is that Machine Learning models are highly effective for intrusion detection and that the implementation of these models in Intrusion Detection Systems (IDS) ensures a high percentage of the continuity of the business that is implemented online, as well as the security of the company's information and its customers.

Keywords*-- Intrusion detection, machine learning, network security, intrusion detection systems, computer crime.

Efectividad de Modelos de Machine Learning en la detección de intrusos en sistemas de información y su Aplicabilidad en el Contexto del Emprendimiento e Innovación: Una Revisión Sistemática de Literatura

Daniel Ivan Llontop Alama, Estudiante de Ingeniería de Sistemas e Informática[1], Christian Abraham Dios-Castillo, Dr. en Administración de la Educación[1]
[1]Universidad Tecnológica del Perú, Perú, 1525293@utp.edu.pe, c16763@utp.edu.pe

Abstract*– A medida que la tecnología ha ido avanzando y el mundo se esta digitalizando cada vez más, el uso de sistemas de información en las empresas se ha vuelto un punto clave para su desarrollo, y con ello la seguridad ha tomado más relevancia que nunca. El propósito de esta investigación es explorar los modelos de Machine Learning y su efectividad en la detección de intrusos, evaluando su impacto y aplicabilidad en el contexto del emprendimiento e innovación. Por ello, se analizaron exhaustivamente 50 documentos obtenidos de la base de datos Scopus en el periodo comprendido entre 2015 y 2024 centrados en el uso de modelos de Machine Learning para la detección de intrusos. Esta revisión abarca aspectos como: los tipos de ataques más comunes, dataste más utilizados, modelos de Machine Learning más estudiados y su clasificación, y finalmente su efectividad en la detección de intrusos. Los resultados mostraron que los 58 modelos de Machine Learning identificados presentan un mínimo de 79.00% de efectividad y un máximo de 99.99% de efectividad. Llegando a la conclusión que los modelos de Machine Learning, son altamente efectivos para la detección de intrusos y que la implementación de estos modelos en los Sistemas de Detección de Intrusos (IDS), asegura en un gran porcentaje la continuidad de los negocios que se implementen en línea, así como la seguridad de la información de la empresa y sus clientes.

Keywords*-- Intrusion detection, machine learning, network security, intrusion detection systems, computer crime.

I. INTRODUCCIÓN

En la era digital actual, la ciberseguridad se ha convertido en un factor prioritario para empresas de todos los sectores y tamaños, ya que las amenazas cibernéticas evolucionan constantemente siendo estas cada vez más sofisticadas y complejas de detectar. El siglo XXI está experimentando más ataques de día cero cada año, cuyo volumen e intensidad fueron mayores que en años anteriores, según el Informe de amenazas a la seguridad en Internet de Symantec de 2017, asimismo se sabe que el número de datos perdidos o robados por piratas informáticos han aumentado a más de catorce mil millones desde 2013, según Data Bruch Statistics de 2023[1].

Es por ello que el uso de modelos de Machine Learning (ML) en Sistemas de detección de Intrusos (IDS) ha emergido como una solución prometedora para mejorar la detección de intrusiones, ofreciendo la capacidad de identificar patrones anómalos y predecir potenciales ataques con una precisión cada vez mayor.

En el ámbito del emprendimiento e innovación, la adopción de modelos ML para la detección de intrusos no solo representa una mejora en la seguridad, sino también una oportunidad para que empresas emergentes puedan diferenciarse en el mercado y desarrollar nuevas soluciones de ciberseguridad. Las startups y empresas innovadoras están en una posición única para aprovechar estos avances, pudiendo crear productos y servicios que no solo protegen a sus clientes (empresas y consumidores), sino que también impulsan la evolución del sector de la ciberseguridad.

Esta Revisión Sistemática de Literatura (RSL) tiene como objetivo explorar los modelos de Machine Learning y su efectividad en la detección de intrusos, evaluando su impacto y aplicabilidad en el contexto del emprendimiento e innovación. A través de este análisis de literatura, se busca proporcionar una visión comprensiva de las tendencias actuales, identificar las mejores prácticas y destacar las oportunidades de innovación que pueden ser capitalizadas por emprendedores y startups.

II. METODOLOGIA

La presente investigación consta de una revisión sistemática de la literatura (RSL), el cual es un método de investigación que sirve para recopilar, evaluar y sintetizar de manera sistemática y exhaustiva toda la evidencia disponible sobre un tema específico [2]. Para ello se utilizó la estrategia PICO, el cual es un método utilizado para construir preguntas de investigación y realizar búsquedas bibliográficas de manera estructurada y efectiva, PICO es un acrónimo que representa cuatro componentes, P: Problema, I: Intervención, C: Comparación y O: Resultado [3]. Con base en ello se planteó como pregunta general: ¿Cuál es la efectividad de los modelos de Machine Learning en la detección de intrusos en sistemas de información? y como preguntas especificas las siguientes, P: ¿Qué modalidades de ataques a sistemas de

Digital Object Identifier: (only for full papers, inserted by LEIRD).
ISSN, ISBN: (to be inserted by LEIRD).
DO NOT REMOVE

información existen?, I: ¿Qué modelos de Machine Learning existen para la detección de intrusos en sistemas de información?, O: ¿Qué nivel de efectividad muestran los casos de aplicación? Además, las palabras clave de los componentes PICO son las detalladas en la Tabla I.

TABLA I
KEYWORDS POR CADA COMPONENTE PICO

Componente	Keyword
P = Problema	Intrusion Detection, intrusion prevention, Intrusion mitigation
I = Intervención	Machine Learning Models, Deep Learning, ML
O = Resultados	Information Systems, Information Technology, IT Systems

La base de datos científica utilizada para la búsqueda de información es Scopus, cuya cadena de búsqueda es la siguiente:
(TITLE-ABS-KEY ("intrusion detection" OR "IDS" OR "intrusion prevention" OR "intrusion mitigation" OR "network intrusion") AND TITLE-ABS-KEY ("Machine learning" OR "deep learning" OR "ML") AND TITLE-ABS-KEY ("information systems" OR "information technology" OR "IT systems"))

La búsqueda fue realizada el 26 de abril a las 4:00pm. Entre los criterios de elegibilidad se han contemplado los siguientes, mostrados en la Tabla II y Tabla III.

TABLA II
CRITERIOS DE INCLUSION

Ítem	Criterio
CI1	Publicaciones relacionadas con los modelos de Machine Learning para la detección de intrusos
CI2	Publicaciones enmarcadas en tipos de ataque
CI3	Publicaciones que reflejen el nivel de efectividad de los modelos de Machine Learning

TABLA III
CRITERIOS DE EXCLUSION

Ítem	Criterio
CE1	Publicaciones anteriores al 2015
CE2	Publicaciones que no son de acceso abierto
CE3	Publicaciones que no son artículos originales y revisiones
CE4	Publicaciones distintas al idioma inglés

Estos criterios fueron insertados en los filtros de la ecuación de búsqueda, quedando como resultante, la siguiente:
(TITLE-ABS-KEY ("intrusion detection" OR "IDS" OR "intrusion prevention" OR "intrusion mitigation" OR "network intrusion") AND TITLE-ABS-KEY ("Machine learning" OR "deep learning" OR "ML") AND TITLE-ABS-KEY ("information systems" OR "information technology" OR "IT systems")) AND PUBYEAR > 2014 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (OA , "all"))

Para un análisis más detallado se utilizó la guía PRISMA, la cual está diseñada para garantizar la transparencia, integridad y precisión en la publicación de revisiones sistemáticas, proporcionando orientación sobre la documentación de los objetivos, métodos y hallazgos de la revisión [4]. En la Fig. 1 se muestra que en la primera búsqueda se obtuvo 496 publicaciones, se recuperaron 145 publicaciones de acceso abierto, 138 publicaciones de los últimos 10 años, 96 publicaciones entre Artículos originales y Revisiones. Finalmente, luego del cribado, se obtuvieron 93 publicaciones en idioma inglés. Cabe indicar que luego de realizar la lectura del resumen de los artículos se descartaron 11 publicaciones por no tener relación con el tema de investigación, además no se logró descargar 27 artículos a texto completo. Finalmente, después de haber realizado la lectura del contenido de los artículos a texto completo se descartaron 5 publicaciones por no cumplir con el esquema PICO, quedando 50 artículos para ser incluidos en el análisis.

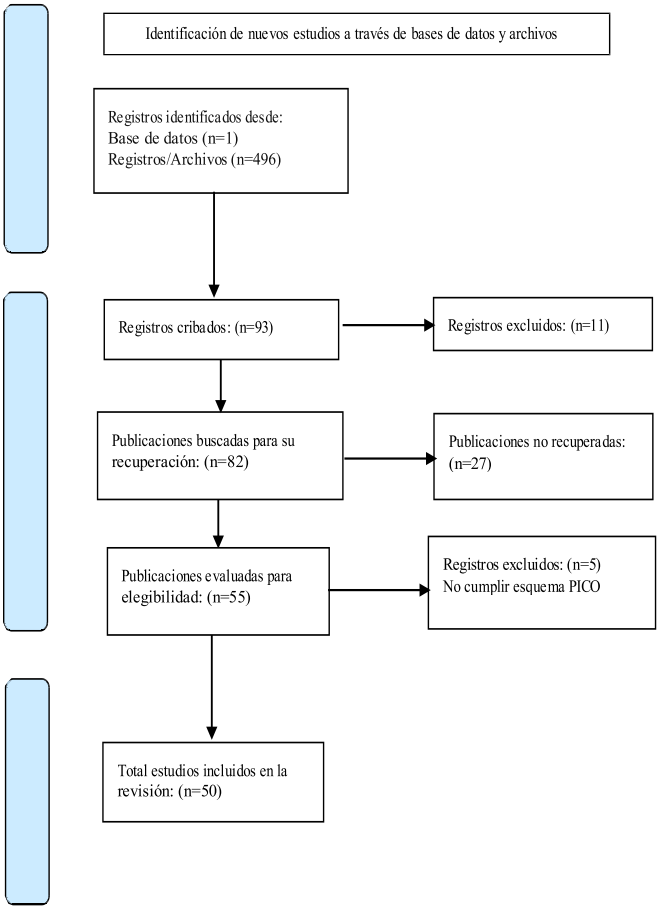

Fig. 1 Diagrama de flujo PRISMA.

III. RESULTADOS

A. *Bibliometría*

Para realizar la bibliometría de los documentos seleccionados para esta RSL, se utilizó VOSviewer, el cual es una herramienta de software para construir y visualizar redes bibliométricas a partir de literatura científica.

En la Fig. 2 Se muestra la cantidad de documentos por año, desde 2015 hasta 2024.

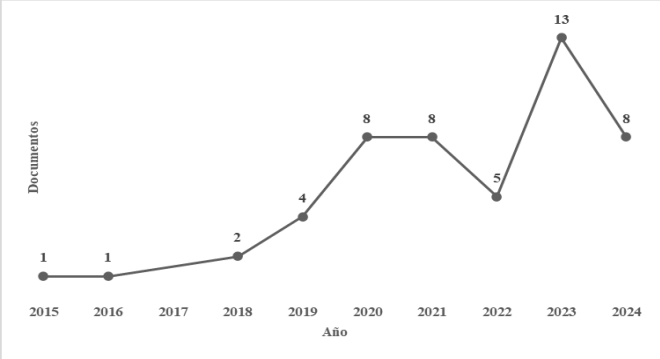

Fig. 2 Documentos indizados por Año

En 2015 y 2016, la producción fue mínima, con solo un documento publicado cada año. El número de documentos comenzó a aumentar lentamente en 2018, con 2 publicaciones, seguido de un incremento del 100% en 2019 con 4 documentos, el año 2020 marca un punto de inflexión con la publicación de 8 documentos duplicando la cantidad del año anterior (100% de crecimiento). La producción se mantuvo alta en 2021 con otros 8 documentos, indicando una estabilización en el interés y la investigación en este campo. En 2022, hubo una ligera disminución con 5 documentos publicados (-37.50%). Sin embargo, el año 2023 mostró un fuerte repunte con 13 documentos, el número más alto en el periodo analizado (160.00% de crecimiento con respecto al año anterior). Finalmente, en 2024, el número de publicaciones se reduce a 8 documentos, dado que, a la fecha de redacción de este manuscrito, el período de indización aún se encuentra en curso; sin embargo, el 2024 tiende a tener un mayor número de publicaciones relacionadas al estudio de modelos de Machine Learning para la detección de intrusos.

En resumen, la tendencia general muestra un crecimiento constante en la producción de documentos sobre ciberseguridad y sistemas de detección de intrusiones desde 2015, con un notable aumento en 2020 y un pico significativo en 2023. Este patrón refleja el creciente reconocimiento de la importancia de la ciberseguridad y la respuesta de la comunidad científica a los desafíos emergentes en este ámbito. En segundo lugar, se realizó el análisis de co-autoria por país, este análisis nos permite comprender como es la distribución de los documentos en termino de países, la cantidad de citas recibidas y cómo se distribuyen las colaboraciones científicas en el área de investigación, el cual es representado por el "total link strength", que en el contexto de este análisis es una medida que significa que los autores de un documento están afiliados a países distintos. La Tabla IV, muestra los 8 países con mayor cantidad de documentos publicados, otros países no fueron considerados en la tabla por tener una cantidad de documentos inferior a 5.

TABLA IV
DOCUMENTOS INDIZADOS POR PAIS

country	documento	citations	total link strength
india	19	131	14
china	16	292	10
saudi arabia	13	169	10
united states	8	358	11
united kingdom	7	128	5
canada	5	206	5
pakistan	5	103	12
turkey	5	64	5

Nota: Elaborada a partir de los documentos analizados.

En primer lugar, India se destaca como el país con el mayor número de documentos (19), aunque ha recibido 131 citas y tiene una fuerza total del enlace de 14. Esto indica una alta productividad y una notable capacidad de colaboración, aunque el impacto en términos de citas es moderado en comparación con otros países. Seguido de China, con 16 documentos, ha recibido 292 citas y tiene una fuerza total del enlace de 10, Lo que refleja una fuerte colaboración con otros países. Arabia Saudita, con 13 documentos y 169 citas, también muestra una alta capacidad de colaboración (fuerza del enlace de 10). Estados Unidos, aunque ha producido solo 8 documentos, lidera en términos de citas recibidas con 358, lo que indica un altísimo impacto y relevancia de sus investigaciones. La fuerza total del enlace de 11 también indica una sólida red de colaboración internacional. El Reino Unido, con 7 documentos y 128 citas, y una fuerza del enlace de 5, muestra un impacto moderado y una conectividad de colaboración notable, aunque menor en comparación con otros países líderes. Canadá y Pakistán, ambos con 5 documentos, presentan diferentes perfiles de impacto y colaboración. Canadá ha recibido 206 citas y tiene una fuerza del enlace de 5. Pakistán, con 103 citas y una fuerza del enlace de 12, muestra una fuerte colaboración internacional a pesar de un menor número de citas. Finalmente, Turquía, también con 5 documentos, ha recibido 64 citas y tiene una fuerza del enlace de 5, indicando un impacto y colaboración más limitados.

En resumen, India y China son los países más productivos en términos de número de documentos, con China mostrando un mayor impacto en citas. Estados Unidos lidera en impacto de citas, reflejando la relevancia de sus investigaciones. Arabia Saudita y Pakistán destacan por su alta conectividad y colaboración internacional

Por último, se realizó el análisis de co-ocurrencia por keyword, este análisis nos permite corroborar que los documentos seleccionados guardan relación con el tema de investigación de esta RSL. En la Tabla V, se muestran las 8 keywords con mayor ocurrencia en los documentos. Cabe mencionar que, el "Total link strength" en el contexto de este análisis es una medida que indica la intensidad de las conexiones entre las palabras clave, basada en la frecuencia con la que aparecen juntas en los documentos.

TABLA V
CO-OCURRENCIA DE KEYWORDS

keyword	occurrences	total link strength
intrusion detection	60	345
machine learning	36	163
network security	35	236
intrusion detection systems	30	210
computer crime	26	184
deep learning	24	124
cybersecurity	22	147
learning systems	21	142

Nota: Elaborada a partir de los documentos analizados.

Las keywords “intrusion detection”, “network security”, “intrusion detection systems”, “computer crime” y “cybersecurity” destacan por su alto nivel de ocurrencia en la literatura. Por otro lado, las keywords “machine learning”, “deep learning” y “learning systems” muestran también altos niveles de ocurrencia en los documentos analizados, sin embargo, cabe destacar que estas keywords tienen un total link strength menor si las comparamos con las mencionadas líneas anteriores, esto debido a que los modelos de machine learning son diversos en su campo de aplicación, siendo la detección de intrusos, uno de ellos.

Finalmente, en la Tabla VI se muestra la distribución de los documentos seleccionados por tipo.

TABLA VI
DOCUMENTOS POR TIPO

Tipo	Cantidad	Porcentaje
Artículos	46	92.00%
Revisiones	4	8.00%
Total	50	100%

B. Respuesta a las preguntas especificas

Para responder a la primer sub-pregunta: "¿Qué modalidades de ataques a sistemas de información existen?", se realizó un análisis del contenido de los documentos descargados, revelando tendencias claras en términos de frecuencia y enfoque. Para abordar este tema se brinda en primer lugar una breve definición acerca de los ataques a sistemas de información (Ciberataques).

Un ciberataque es un ataque al sistema de una computadora que se utiliza para comprometer la confidencialidad, integridad y disponibilidad de dichos datos [5]. En la Tabla VII y VIII se muestra la Distribución de referencias por tipos de ataques y Tipos de ataques respectivamente

Tabla VII
DISTRIBUCIÓN DE REFERENCIAS POR TIPOS DE ATAQUES

Referencia	Ataque	No. de Papers
[5],[6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18],[19],[20],[21],[22],[23],[24],[25],[26],[27],[28],[29],[30],[31],[32],[33],[34],[35],[36],[37],[38],[39],[40],[41],[42],[43],[44],[45]	DoS	41
[6],[7],[12],[13],[14],[15],[16],[17],[18],[26],[27],[28],[29],[31],[32],[34],[35],[38],[41],[42],[43],[44],[45],[46]	U2R	24
[5],[6],[10],[11],[13],[14],[15],[17],[23],[24],[25],[26],[30],[34],[35],[37],[47],[48],[49],[50],[51],[52],[53]	DDoS	23
[6],[7],[12],[13],[14],[15],[16],[17],[18],[26],[27],[28],[29],[31],[32],[34],[35],[38],[41],[42],[44],[45],[46]	R2L	23
[14],[16],[17],[18],[26],[27],[28],[29],[31],[32],[34],[35],[38],[42],[44],[45],[46]	Probe	17
[5],[9],[10],[17],[19],[24],[34],[35],[39],[45]	XSS	10
[9],[11],[17],[25],[30],[34],[35],[40],[54]	Botnet	9
[5],[9],[10],[17],[19],[24],[34],[35],[39]	SQL injection	9
[5],[14],[16],[20],[22],[23],[26],[33],[54]	Worms	9
[14],[19],[20],[22],[23],[26],[33],[43]	Shellcode	8
[14],[19],[20],[22],[23],[26],[33],[43]	Fuzzers	8
[17],[24],[34],[35],[36],[39],[45]	SSH	7
[19],[20],[22],[26],[33],[43]	Exploits	6
[5],[17],[34],[35],[54]	Rasomware	5
[14],[22],[23],[26],[33]	Reconnaissance	5
[17],[34],[35],[40]	Bruteforce	4
[5],[54]	Trojan	2

Nota: Elaborada a partir de los documentos analizados.

TABLA VIII

TIPOS DE ATAQUES

Ataque	No. de Papers	% de Papers
DoS	41	82.00%
U2R	24	48.00%
DDoS	23	46.00%
R2L	23	46.00%
Probe	17	34.00%
XSS	10	20.00%
Botnet	9	18.00%
SQL injection	9	18.00%
Worms	9	18.00%
Shellcode	8	16.00%
Fuzzers	8	16.00%
SSH	7	14.00%
Exploits	6	12.00%
Rasomware	5	10.00%
Reconnaissance	5	10.00%
Bruteforce	4	8.00%
Trojan	2	4.00%

Nota: Elaborada a partir de los documentos analizados.

Haciendo un análisis de los cinco tipos de ataques más estudiados, tenemos en primer lugar al DoS (Denial of Service), estudiado en el 82.00% de los documentos (41 papers). Este alto porcentaje indica que la denegación de servicio es una preocupación central en los documentos analizados, debido a su potencial para interrumpir la disponibilidad de servicios y causar daños significativos a las infraestructuras de red. Por otro lado, U2R (User to Root) es el segundo ataque más estudiado, apareciendo en el 48.00% de los documentos (24 papers). Los ataques U2R, que implican la escalada de privilegios desde un usuario normal a un usuario con privilegios de root, son críticos debido a su capacidad para comprometer completamente la seguridad de un sistema. En tercer lugar, DDoS (Distributed Denial of Service) también recibe una considerable atención, siendo estudiado en el 46.00% de los documentos (23 papers). Los ataques DDoS son una variante de los ataques DoS que utilizan múltiples sistemas comprometidos para abrumar el objetivo, haciéndolos aún más difíciles de mitigar y defender. R2L (Remote to Local), estudiado en el 46.00% de los documentos (23 papers), es otro tipo de ataque significativo. Estos ataques permiten a un atacante remoto obtener acceso local en la máquina objetivo, representando una seria amenaza a la integridad y confidencialidad de los datos. Probe es el quinto tipo de ataque más estudiado, mencionado en el 34.00% de los documentos (17 papers). Los ataques de sondeo o escaneo son utilizados por atacantes para recopilar información sobre las redes y sistemas, lo cual puede ser un paso preliminar antes de lanzar otros tipos de ataques más dañinos.

En resumen, la investigación en documentos analizados se centra significativamente en los ataques DoS y sus variantes (DDoS), así como en los ataques que implican la escalada de privilegios (U2R) y el acceso remoto no autorizado (R2L). La atención considerable a los ataques de sondeo (Probe) también destaca la importancia de la prevención y detección temprana de posibles intrusiones.

Por otro lado, es importante analizar los conjuntos de datos (Datasets) estudiados en los documentos, ya que estos conjuntos de datos contienen información acerca de uno o más tipos de ataques y son utilizados en los métodos experimentales para la validación de cualquier modelo de Machine Learning, porque permiten evaluar qué tan bien el modelo sugerido puede identificar actividades intrusivas [1]. En la Tabla IX se muestra las referencias por dataset y en la Tabla X, los ocho datasets más estudiados en la literatura, los otros datasets no fueron considerados para dicha tabla por tener una única ocurrencia en los documentos.

TABLA IX

DISTRIBUCIÓN DE REFERENCIAS POR DATASET

Referencia	Dataset	No. de Papers
[5],[6],[8],[12],[13],[14],[15], [16],[17],[18],[22],[24],[26],[27],[31], [34],[37],[39],[41],[42],[46],[48]	NSL-KDD	22
[5],[6],[12],[14],[16],[17],[19],[20],[22], [23],[26],[33],[34],[35],[43],[45]	UNSW-NB15	16
[5],[7],[13],[15],[17],[24],[28],[29], [32],[34],[35],[38],[44],[45]	KDDcup99	14
[10],[15],[17],[19],[22],[31], [32],[34],[36],[37],[39],[52]	CICIDS2017	12
[9],[11],[17],[22],[25],[30],[34],[40],[53]	CSE-CIC-IDS2018	9
[48],[49],[50],[51],[52]	CICDDoS2019	5
[15],[17],[34],[35],[45]	DARPA	5
[15],[34],[35],[45]	ISCX	4
[54]	25-DGA	1
[48]	CAIDA	1
[17]	CIDDS-001	1
[54]	CTU13	1
[13]	DARPA1999	1
[54]	EMBER	1
[52]	InSDN	1
[47]	IoT-23	1
[17]	ISCXIDS2012	1
[54]	Kasperski	1
[5]	Kyoto dataset	1

[54]	Malicia	1
[15]	NSA	1
[21]	Suricata logs	1
[23]	UGR' 16	1

Nota: Elaborada a partir de los documentos analizados.

TABLA X

DATASETS

Dataset	No. de Papers	% de Papers
NSL-KDD	22	44.00%
UNSW-NB15	16	32.00%
KDDcup99	14	28.00%
CICIDS2017	12	24.00%
CSE-CIC-IDS2018	9	18.00%
CICDDoS2019	5	10.00%
DARPA	5	10.00%
ISCX	4	8.00%

Nota: Elaborada a partir de los documentos analizados.

Como se puede apreciar en la Tabla X. NSL-KDD es el dataset más utilizado, apareciendo en el 44.00% de los documentos (22 papers). Este dataset es una versión mejorada del famoso KDDcup99 y se utiliza ampliamente debido a sus mejoras en la reducción de registros redundantes y la eliminación de las limitaciones del dataset original, haciéndolo más adecuado para la evaluación de sistemas de detección de intrusiones. UNSW-NB15 es el segundo dataset más estudiado, presente en el 32.00% de los documentos (16 papers). Este dataset es valorado por su capacidad para proporcionar un conjunto de datos más reciente y completo que abarca múltiples tipos de ataques y características modernas, lo que lo hace relevante para la investigación actual en ciberseguridad. KDDcup99, a pesar de ser uno de los datasets más antiguos, sigue siendo relevante y es utilizado en el 28.00% de los documentos (14 papers). Su persistente popularidad se debe a su extensa base de datos y su uso histórico. CICIDS2017 es mencionado en el 24.00% de los documentos (12 papers). Este dataset es apreciado por la inclusión de una variedad de ataques modernos y tráfico benigno, proporcionando un contexto más realista para la evaluación de sistemas de detección de intrusiones. CSE-CIC-IDS2018, utilizado en el 18.00% de los documentos (9 papers), es otro dataset moderno que combina características de tráfico de red con ataques actualizados. CICDDoS2019 y DARPA, cada uno mencionado en el 10.00% de los documentos (5 papers), son valorados por su enfoque en ataques DDoS y su uso en la evaluación de técnicas de mitigación de estos ataques distribuidos. Finalmente, ISCX, utilizado en el 8.00% de los documentos (4 papers), es conocido por sus escenarios de ataque realistas y detallados, proporcionando un valioso recurso para la investigación en detección de anomalías y evaluación de sistemas de seguridad.

En resumen, los datasets NSL-KDD y UNSW-NB15 son los más utilizados en métodos experimentales para la validación de la efectividad de los modelos de Machine Learning, reflejando su relevancia y adaptabilidad a las necesidades actuales de la investigación en ciberseguridad. La continua utilización de KDDcup99 y la creciente popularidad de datasets más modernos como CICIDS2017 y CSE-CIC-IDS2018 subrayan la importancia de disponer de datos diversos y actualizados para la evaluación eficaz de sistemas de detección de intrusiones (IDS).

Continuando con la segunda sub-pregunta: "¿Qué modelos de Machine Learning existen para la detección de intrusos en sistemas de información? En primer lugar, para abordar esta pregunta debemos entender cuáles son los tipos de modelos de Machine Learning y cómo se clasifican. Los modelos de Machine Learning se dividen en 2 tipos: Shallow Learning y Deep Learning.

El Shallow Learning hace referencia los modelos tradiciones de Machine Learning como ANN, SVM, LR, KNN, DT y modelos híbridos combinados. La Tabla XI muestra las abreviaturas y sus significados. Algunos de estos métodos se han estudiado durante varias décadas y su metodología está madura. Se concentran no solo en los efectos de la detección, sino también en temas útiles como la eficiencia de la detección y la gestión de datos [55].

Por otro lado, el Deep Learning se refiere a un tipo específico de aprendizaje automático en el que se utiliza una red neuronal (NN) multicapa para resolver problemas. Estas redes neuronales "aprenden" a partir de amplios conjuntos de datos para imitar la actividad del cerebro humano, aquí podemos encontrar modelos como RNN, CNN, DNN, etc. [25].

La principal diferencia entre ambos tipos es el rendimiento, Los modelos de Deep Learning requieren una gran cantidad de datos para que funcionen correctamente, por el contrario, el Shallow Learning puede dar mejores resultados con una menor cantidad de datos [44].

Asimismo, estos dos tipos de modelos tienen 3 clasificaciones: Supervisado, No Supervisado y Semi-Supervisado (Híbridos). Por un lado, Modelo Supervisado utiliza datos etiquetados para aprender. Después de analizar los datos, el algoritmo utiliza los patrones para determinar qué etiqueta aplicar a los datos nuevos, luego conecta los patrones con los datos nuevos sin etiquetar [17]. El modelo No Supervisado se utiliza para analizar la estructura de los datos, extraer información útil, encontrar patrones y aumentar la eficiencia, luego incorpora esta información a su funcionamiento [17]. Por último, el modelo Semi-Supervisado combina las ventajas del Aprendizaje Supervisado y No

supervisado. Cuando tenemos pocos datos etiquetados, pero muchos sin etiquetar, una estrategia de aprendizaje Semi-Supervisado funciona bien, Las características de aprendizaje Supervisado pueden usarse para aprovechar la pequeña cantidad de datos etiquetados mientras que las características de aprendizaje No Supervisado permiten aprovechar una gran cantidad de datos sin etiquetar [54]. La Fig. 3 muestra la taxonomía del Machine Learning. Además, en la Tabla XI, se provee una lista de abreviaturas acerca de los modelos de Machine Learning y su significado.

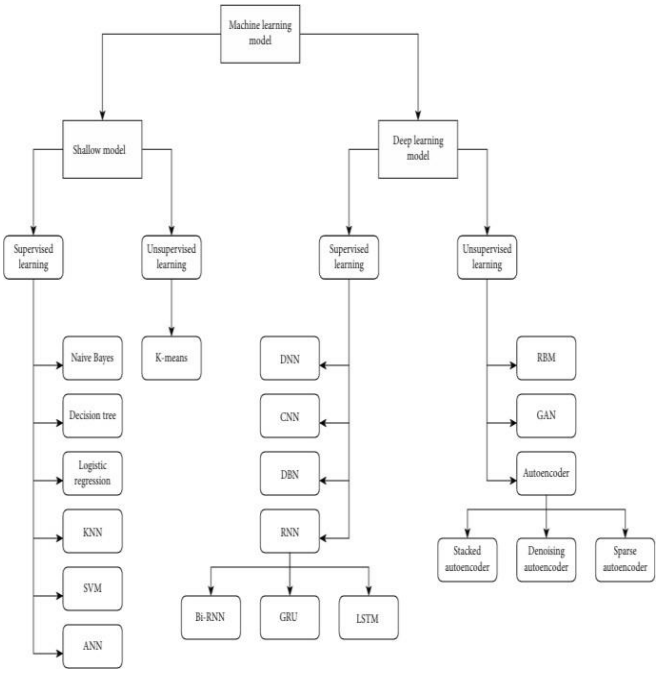

Fig. 3 Taxonomía del Machine Learning [17].

TABLA XI

LISTA DE ABREVIATURAS

Abreviatura	Significado
ANN	Artificial Neural Network
BaysCNN	Bayesian-based Convolutional Neural Network
BaysFusCNN	Bayesian-based Data Fusion Convolutional Neural Network
BiLSTM	Bidirectional Long-Short Term Memory network
BiRNN	Bidirectional Recurrent Neural Network
CFS_FPA	Correlation Feature Selection coupled with Forest Panelized Attributes
CNN	Convolutional Neural Network
CNN-LSTM	Convolutional Long-Short Term memory network
DCNN	Deep Convolutional Neural Network
DNN	Deep Neural Network
DT	Decisión Tree
GRU-RNN	Deep Neural Network with Gated Recurrent Units
GSAFS-OQNN	Gravitational Search Algorithm-Based Feature Selection with Optimal Quantum Neural
KNN	K-Nearest Neighbor
LR	Logistic Regresion
LSTM	Long Short-Term Memory
MRA S-Ddca	Multiresolution Analysis Segmented DeterministicDCA
NB	Naive Bayes
RC	Random Committee
RF	Random Forest
RNN	Recurrent Neural Networks
SVM	Support Vector Machine

La Tabla XII y XIII muestran la distribución de referencias por tipo de modelo y la distribución de los tipos de modelos estudiados en los documentos, respectivamente.

Tabla XII

Distribución de referencias por tipo de modelo

Referencia	Tipo Modelo	No. de Papers
[6],[7],[8],[9],[10],[11],[13],[22],[23],[29],[32],[36],[38],[40],[42],[43],[47],[49]	Shallow Learning (Supervised)	18
[12],[16],[19],[20],[21],[25],[26],[27],[30],[37],[41],[44], [48],[50],[51],[52]	Deep Learnig (Supervised)	16
[14],[18],[31]	Deep Learnig (Unsupervised)	3
[15],[17],[34]	Shallow Learning (Supervised - Unsupervised) , Deep Learnig (Supervised - Unsupervised)	3
[35],[45]	Deep Learning (Supervised - Usupervised-Semisupervised)	2
[24],[46]	Shallow Learning (Semi-Supervised)	2
[33],[39]	Shallow Learning (Supervised) , Deep Learnig (Supervised)	2
[28],[54]	Shallow Learning (Unsupervised)	2
[5]	Shallow Learning (Supervised - Unsupervised)	1
[53]	Deep Learning (Semisupervised)	1

Nota: Elaborada a partir de los documentos analizados.

TABLA XIII

TIPOS DE MODELOS

Tipo Modelo	No. de Papers	% de Papers
Shallow Learning (Supervised)	18	36.00%
Deep Learnig (Supervised)	16	32.00%
Deep Learnig (Unsupervised)	3	6.00%
Shallow Learning (Supervised - Unsupervised) , Deep Learnig (Supervised - Unsupervised)	3	6.00%
Deep Learning (Supervised - Usupervised-Semisupervised)	2	4.00%
Shallow Learning (Semi-Supervised)	2	4.00%
Shallow Learning (Supervised) , Deep Learnig (Supervised)	2	4.00%
Shallow Learning (Unsupervised)	2	4.00%
Shallow Learning (Supervised - Unsupervised)	1	2.00%
Deep Learning (Semisupervised)	1	2.00%
Total Papers		50

Nota: Elaborada a partir de los documentos analizados.

Como se puede apreciar en la Tabla XIII. Los modelos Shallow Learning (Supervised) son los más estudiados, apareciendo en el 36.00% de los documentos (18 papers). Estos modelos supervisados, que incluyen algoritmos como árboles de decisión, SVM y KNN, son populares debido a su simplicidad y eficacia en la clasificación de datos cuando se dispone de un conjunto de entrenamiento etiquetado. Por otro lado, el Deep Learning (Supervised) es el segundo tipo más frecuente, presente en el 32.00% de los documentos (16 papers). La popularidad de los modelos de aprendizaje profundo supervisados, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), refleja su capacidad para manejar grandes volúmenes de datos y aprender representaciones complejas, lo que es crucial para la detección de intrusiones en entornos de red complejos. Deep Learning (Unsupervised) y las combinaciones de Shallow Learning (Supervised-Unsupervised) y Deep Learning (Supervised-Unsupervised) son mencionados en el 6.00% de los documentos cada uno (3 papers). Los modelos de aprendizaje profundo no supervisados, como autoencoders y redes generativas adversariales (GANs), son valiosos para la detección de anomalías sin necesidad de datos etiquetados. Las combinaciones de modelos supervisados y no supervisados sugieren enfoques híbridos que buscan aprovechar las fortalezas de ambos métodos. Las categorías menos frecuentes incluyen Deep Learning (Supervised-Unsupervised-Semisupervised), Shallow Learning (Semi-Supervised), y combinaciones de Shallow Learning (Supervised) y Deep Learning (Supervised), cada una apareciendo en el 4.00% de los documentos (2 papers cada una). Estos modelos representan enfoques más avanzados y específicos que intentan abordar diversas limitaciones de los modelos puramente supervisados o no supervisados, como la escasez de datos etiquetados. Finalmente, los modelos de Shallow Learning (Unsupervised) y Shallow Learning (Supervised-Unsupervised), junto con Deep Learning (Semisupervised), son los menos comunes, apareciendo en solo el 2.00% de los documentos (1 paper cada uno).

En resumen, los modelos de Shallow Learning (Supervised) y Deep Learning (Supervised) son los más estudiados de manera individual en la literatura, reflejando el interés por parte de los investigadores en estos tipos de modelo en específico. La presencia de modelos híbridos o semisupervisados indica una tendencia hacia la búsqueda de métodos más adaptativos que puedan mejorar la precisión y eficiencia en la detección de intrusiones, tomando los beneficios de los modelos supervisados y no supervisados. Asimismo, el estudio de diversos modelos en conjunto, representan un índice bajo de investigación.

Continuando con el análisis de modelos de Machine Learning para la detección de intrusos, ahora de manera individual, Se identificaron un total de 58 modelos de Machine Learning a lo largo de los 50 documentos analizados. En la tabla XIV se muestra la distribución de referencias por modelo de ML y en la Tabla XV se muestran los 12 modelos de ML más estudiados en los documentos, los demás modelos no fueron considerados en la tabla por tener un índice de estudio en los documentos menor a 5.

Tabla XIV

DISTRIBUCIÓN DE REFERENCIAS POR MODELO DE ML

Referencia	Modelo	No. De Papers
[5],[8],[10],[13],[15],[17],[29],[32],[33], [35],[36],[38],[42],[43], [45],[49],[54]	SVM	17
[5],[8],[9],[49],[10],[13],[15], [17],[22],[23],[39],[42],[43],[54]	DT	14
[5],[6],[8],[9],[10],[17],[23], [33],[34],[36],[42],[43],[54]	KNN	13
[5],[8],[10],[13],[15],[17], [33],[34],[36],[49],[54]	NB	11
[5],[9],[10],[13],[15],[23], [36],[39],[47],[49],[54]	RF	11
[5],[17],[30],[33],[34],[35],[39],[45]	CNN	8
[34],[35],[39],[44],[45],[50],[53]	LSTM	7
[5],[17],[34],[35],[39],[45]	RNN	6
[5],[10],[13],[15],[17]	ANN	5

[5],[23],[32],[49],[54]	LR	5
[12],[17],[52],[26],[35]	DNN	5
[5],[15],[17],[28],[54]	K-means	5

Nota: Elaborada a partir de los documentos analizados.

TABLA XV

MODELOS DE MACHINE LEARNING

Modelo	No. de Papers	% de Papers
SVM	17	34.00%
DT	14	28.00%
KNN	13	26.00%
NB	11	22.00%
RF	11	22.00%
CNN	8	16.00%
LSTM	7	14.00%
RNN	6	12.00%
ANN	5	10.00%
LR	5	10.00%
DNN	5	10.00%
K-means	5	10.00%

Nota: Elaborada a partir de los documentos analizados.

Como resumen de los seis modelos de ML más estudiados en la literatura. En primer lugar, Support Vector Machine (SVM) es el modelo más estudiado, mencionado en el 34.00% de los documentos (17 papers). La popularidad de SVM se debe a su capacidad para manejar problemas de clasificación en espacios de alta dimensión y su efectividad en la detección de patrones complejos, lo que es crucial para la identificación de intrusiones. Decision Trees (DT) es el segundo modelo más estudiado, presente en el 28.00% de los documentos (14 papers). Los árboles de decisión son apreciados por su simplicidad y facilidad de interpretación, lo que los hace una opción popular para la detección de anomalías y la clasificación en ciberseguridad. K-Nearest Neighbors (KNN) aparece en el 26.00% de los documentos (13 papers). Este modelo es valorado por su simplicidad y efectividad en problemas de clasificación, especialmente en casos donde la distribución de los datos es compleja y no lineal. Naive Bayes (NB) y Random Forest (RF), ambos mencionados en el 22.00% de los documentos (11 papers cada uno), son modelos estadísticos que ofrecen buenas capacidades de clasificación. NB es conocido por su simplicidad y rapidez, mientras que RF, un conjunto de árboles de decisión, es popular por su capacidad para mejorar la precisión y reducir el sobreajuste mediante el uso de múltiples árboles. Finalmente, Convolutional Neural Networks (CNN) es estudiado en el 16.00% de los documentos (8 papers). Las CNN son particularmente efectivas en el procesamiento y análisis de datos con estructura espacial, como los datos de tráfico de red, y son valoradas por su capacidad para aprender características complejas a partir de los datos.

En resumen, los modelos SVM, DT y KNN son los más estudiados en la literatura, reflejando su importancia y aplicabilidad en la detección de intrusiones. La presencia significativa de NB y RF muestra una preferencia por modelos que equilibran simplicidad y efectividad. La inclusión de CNN destaca la tendencia hacia el uso de técnicas avanzadas de aprendizaje profundo (Deep Learning) en la detección de intrusos. Estos modelos representan un conjunto diversificado de enfoques que los investigadores están explorando para mejorar la detección y mitigación de amenazas en el ámbito de la ciberseguridad.

Por último, para responder a la tercera sub-pregunta: ¿Qué nivel de efectividad muestran los casos de aplicación?, se realizó un minucioso análisis de la literatura, por un lado, en una parte de los artículos originales se proponían nuevos modelos de Machine Learning y a su vez estos eran comparados con otros modelos ya existentes, dando como resultado el desempeño de estos modelos en valores cuantitativos. Asimismo, los documentos de tipo Revisión, en su mayoría no eran más que recopilatorios de métodos experimentales, dando como resultado también el desempeño de los modelos de Machine Learning en valores cuantitativos. En la tabla XVI se muestra los 20 modelos de Machine Learning con mayor efectividad en la literatura.

TABLA XVI

EFECTIVIDAD DE LOS MODELOS DE MACHINE LEARNING

Ref.	Modelo	Efectividad	TOP
[52]	DCNN	99.99%	TOP10
[28]	K-means	99.99%	
[11]	XGBoost	99.99%	
[22]	MRA S-Ddca+DT	99.97%	
[44]	BiLSTM	99.93%	
[44]	BiRNN	99.92%	
[13]	DT	99.88%	
[19]	DCF-IDS	99.88%	
[44]	CNN-LSTM	99.87%	
[17]	GRU-RNN	99.84%	
[51]	BaysFusCNN	99.79%	TOP20
[20]	GSAFS-OQNN	99.79%	

[17]	KNN	99.76%	
[36]	CFS_FPA With Voted (RF, NB, KNN, SVM)	99.70%	
[30]	CNN	99.70%	
[6]	RC	99.70%	
[8]	Bagging-DT	99.68%	
[51]	BaysCNN	99.66%	
[12]	DNN	99.63%	
[8]	PART	99.62%	

Nota: Elaborada a partir de los documentos analizados.

A continuación, se presenta un breve análisis de los 10 modelos de Machine Learning más efectivos (TOP10), cabe mencionar que los valores porcentuales mostrados en la Tabla XVI representan el valor más alto obtenido en cuanto a desempeño, ya que estos valores varían en cada prueba realizada por diversos factores tales como: dataset utilizado, tipo de ataque, etc.

Los modelos Deep Convolutional Neural Network (DCNN), K-means y XGBoost comparten la posición de los modelos más efectivos, cada uno con una efectividad del 99.99%. Estos modelos representan la vanguardia en la precisión de la detección de intrusiones, con DCNN aprovechando las capacidades de las redes neuronales convolucionales para el reconocimiento de patrones complejos, K-means destacando en la agrupación y segmentación de datos, y XGBoost siendo una poderosa técnica de boosting que optimiza la precisión a través de un enfoque iterativo. El modelo MRA S-Ddca+DT sigue de cerca con una efectividad del 99.97%. Este enfoque híbrido combina métodos avanzados de reducción de características con árboles de decisión, mejorando la precisión y eficiencia en la detección de intrusiones. Bidirectional Long Short-Term Memory (BiLSTM) y Bidirectional Recurrent Neural Network (BiRNN), con efectividades del 99.93% y 99.92% respectivamente, son modelos de aprendizaje profundo especializados en procesar secuencias temporales y capturar dependencias a largo plazo, cruciales para la detección de patrones de ataque en flujos de datos de red. Decision Trees (DT) y DCF-IDS ambos muestran una efectividad del 99.88%. Los árboles de decisión son conocidos por su simplicidad y claridad en la toma de decisiones, mientras que DCF-IDS es un sistema de detección de intrusiones basado en características que integra múltiples técnicas para mejorar la precisión de la detección. CNN-LSTM, con una efectividad del 99.87%, combina la capacidad de las redes neuronales convolucionales (CNN) para la extracción de características espaciales con las ventajas de las redes de memoria a largo plazo (LSTM) para la captura de patrones temporales, ofreciendo un enfoque robusto para la detección de anomalías en datos de red. Finalmente, GRU-RNN alcanza una efectividad del 99.84%, utilizando unidades de memoria de puerta (GRU) en redes neuronales recurrentes (RNN)para manejar eficientemente secuencias temporales y mejorar la detección de intrusiones.

En resumen, los modelos DCNN, K-means y XGBoost se destacan por su máxima efectividad del 99.99%, mientras que los enfoques híbridos y las técnicas de aprendizaje profundo como BiLSTM, BiRNN y CNN-LSTM también muestran una alta precisión en la detección de intrusiones. Los modelos de árboles de decisión (DT) y sistemas basados en características siguen siendo opciones robustas y efectivas en este campo. Estos resultados subrayan la diversidad y el potencial de los enfoques avanzados de Machine Learning en la mejora de la seguridad cibernética y la detección de intrusos.

Además, cabe mencionar que los modelos de Machine Learning no considerados en la Tabla XVI (21-58), muestran una efectividad entre 99.57% y 79.00%.

IV. Conclusion

El análisis exhaustivo de la literatura sobre la efectividad de modelos de Machine Learning para la detección de intrusiones ha revelado importantes hallazgos y tendencias. A través de los 58 modelos identificados en la literatura, se ha evidenciado que dichos modelos son altamente efectivos en la detección de intrusos, con un mínimo de 79.00% de efectividad y un máximo de 99.99% de efectividad.

Entre los modelos evaluados, los diez con mayor efectividad, incluyendo DCNN, K-means, XGBoost, y enfoques híbridos como MRA S-Ddca+DT, demostraron una precisión excepcional con valores de efectividad cercanos al 99.99%. Estos modelos destacan por su capacidad para manejar grandes volúmenes de datos y detectar patrones complejos, esenciales en la ciberseguridad moderna.

Los modelos Deep Learning supervisado y no supervisado, como BiLSTM, BiRNN, y CNN-LSTM, han mostrado su alta capacidad en el procesamiento de datos secuenciales y en la captura de dependencias a largo plazo. Por otro lado, modelos más tradicionales como los árboles de decisión (DT) y los modelos de bosque aleatorio (RF) continúan siendo robustos y efectivos, combinando simplicidad y claridad en la toma de decisiones.

Asimismo, la literatura refleja un patron significativo en cuanto al estudio de los modelos de ML, siendo de mayor estudio, los modelos de tipo Supervisado tanto en Shallow Learning como en Deep Learning. Por otro lado, los Datasets como NSL-KDD, UNSW-NB15, y CICIDS2017 son los más utilizados en los métodos experimentales para la validación de la efectividad de los modelos de Machine Learning.

En conclusión, el estudio sistemático de la literatura confirma que los modelos de Machine Learning, son altamente efectivos para la detección de intrusos. La implementación de estos modelos en los Sistemas de Detección de Intrusos (IDS), asegura en un gran porcentaje la continuidad de los negocios que se implementen en línea, así

como la seguridad de la información de la empresa y sus clientes.

REFERENCIAS

[1] Z. Azam, Md. M. Islam, y M. N. Huda, «Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree», IEEE Access, vol. 11, pp. 80348-80391, 2023, doi: 10.1109/ACCESS.2023.3296444.

[2] M. A. Babar y H. Zhang, «Systematic literature reviews in software engineering: Preliminary results from interviews with researchers», en 2009 3rd International Symposium on Empirical Software Engineering and Measurement, Lake Buena Vista, FL, USA: IEEE, oct. 2009, pp. 346-355. doi: 10.1109/ESEM.2009.5314235.

[3] C. M. D. C. Santos, C. A. D. M. Pimenta, y M. R. C. Nobre, «The PICO strategy for the research question construction and evidence search», Rev. Lat. Am. Enfermagem, vol. 15, n.o 3, pp. 508-511, jun. 2007, doi: 10.1590/S0104-11692007000300023.

[4] M. J. Page et al., «Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas», Rev. Esp. Cardiol., vol. 74, n.o 9, pp. 790-799, sep. 2021, doi: 10.1016/j.recesp.2021.06.016.

[5] J. Note y M. Ali, «Comparative Analysis of Intrusion Detection System Using Machine Learning and Deep Learning Algorithms», Ann. Emerg. Technol. Comput., vol. 6, n.o 3, pp. 19-36, jul. 2022, doi: 10.33166/AETiC.2022.03.003.

[6] Christian University of Rwanda et al., «Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches», Int. J. Intell. Eng. Syst., vol. 13, n.o 3, pp. 433-445, jun. 2020, doi: 10.22266/ijies2020.0630.39.

[7] P. Amudha, S. Karthik, y S. Sivakumari, «A Hybrid Swarm Intelligence Algorithm for Intrusion Detection Using Significant Features», Sci. World J., vol. 2015, pp. 1-15, 2015, doi: 10.1155/2015/574589.

[8] A. Iqbal, S. Aftab, I. Ullah, M. A. Saeed, y A. Husen, «A Classification Framework to Detect DoS Attacks», Int. J. Comput. Netw. Inf. Secur., vol. 11, n.o 9, pp. 40-47, sep. 2019, doi: 10.5815/ijcnis.2019.09.05.

[9] «Implementation of Six Single Classifiers and Feature Selection for Performance Enhancement in Anomaly-Based Intrusion Detection», Int. J. Electron. Commun. Eng., vol. 11, n.o 3, mar. 2024, doi: 10.14445/23488549/IJECE-V11I3P118.

[10] M. A. Paracha, M. Sadiq, J. Liang, M. H. Durad, y M. Sheeraz, «Multi-Layered Filtration Framework for Efficient Detection of Network Attacks Using Machine Learning», Sensors, vol. 23, n.o 13, p. 5829, jun. 2023, doi: 10.3390/s23135829.

[11] W. Chimphlee y S. Chimphlee, «Hyperparameters optimization XGBoost for network intrusion detection using CSE-CIC-IDS 2018 dataset», IAES Int. J. Artif. Intell. IJ-AI, vol. 13, n.o 1, p. 817, mar. 2024, doi: 10.11591/ijai.v13.i1.pp817-826.

[12] P. Illy y G. Kaddoum, «A Collaborative DNN-Based Low-Latency IDPS for Mission-Critical Smart Factory Networks», IEEE Access, vol. 11, pp. 96317-96329, 2023, doi: 10.1109/ACCESS.2023.3311822.

[13] B. Ahmad, W. Jian, y Z. Anwar Ali, «Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions», J. Comput. Netw. Commun., vol. 2018, pp. 1-10, jul. 2018, doi: 10.1155/2018/6383145.

[14] Q. Zhou y Z. Wang, «A Network Intrusion Detection Method for Information Systems Using Federated Learning and Improved Transformer»:, Int. J. Semantic Web Inf. Syst., vol. 20, n.o 1, pp. 1-20, dic. 2023, doi: 10.4018/IJSWIS.334845.

[15] F. A. Vadhil, M. F. Nanne, y M. L. Salihi, «Importance of Machine Learning Techniques to Improve the Open Source Intrusion Detection Systems», Indones. J. Electr. Eng. Inform. IJEEI, vol. 9, n.o 3, pp. 774-783, sep. 2021, doi: 10.52549/ijeei.v9i3.3219.

[16] Y. Yu y N. Bian, «An Intrusion Detection Method Using Few-Shot Learning», IEEE Access, vol. 8, pp. 49730-49740, 2020, doi: 10.1109/ACCESS.2020.2980136.

[17] A. Momand, S. U. Jan, y N. Ramzan, «A Systematic and Comprehensive Survey of Recent Advances in Intrusion Detection Systems Using Machine Learning: Deep Learning, Datasets, and Attack Taxonomy», J. Sens., vol. 2023, pp. 1-18, feb. 2023, doi: 10.1155/2023/6048087.

[18] A. Ugendhar et al., «A Novel Intelligent-Based Intrusion Detection System Approach Using Deep Multilayer Classification», Math. Probl. Eng., vol. 2022, pp. 1-10, may 2022, doi: 10.1155/2022/8030510.

[19] X. Zhang, J. Chen, Y. Zhou, L. Han, y J. Lin, «A Multiple-Layer Representation Learning Model for Network-Based Attack Detection», IEEE Access, vol. 7, pp. 91992-92008, 2019, doi: 10.1109/ACCESS.2019.2927465.

[20] N. O. Aljehane, H. A. Mengash, S. B. H. Hassine, F. A. Alotaibi, A. S. Salama, y S. Abdelbagi, «Optimizing intrusion detection using intelligent feature selection with machine learning model», Alex. Eng. J., vol. 91, pp. 39-49, mar. 2024, doi: 10.1016/j.aej.2024.01.073.

[21] K. Fotiadou, T.-H. Velivassaki, A. Voulkidis, D. Skias, S. Tsekeridou, y T. Zahariadis, «Network Traffic Anomaly Detection via Deep Learning», Information, vol. 12, n.o 5, p. 215, may 2021, doi: 10.3390/info12050215.

[22] D. Limon-Cantu y V. Alarcon-Aquino, «Multiresolution dendritic cell algorithm for network anomaly detection», PeerJ Comput. Sci., vol. 7, p. e749, oct. 2021, doi: 10.7717/peerj-cs.749.

[23] S. Rajagopal, P. P. Kundapur, y H. Katiganere Siddaramappa, «A predictive model for network intrusion detection using stacking approach», Int. J. Electr. Comput. Eng. IJECE, vol. 10, n.o 3, p. 2734, jun. 2020, doi: 10.11591/ijece.v10i3.pp2734-2741.

[24] LASTIMI Laboratory, Superior School of Technologies of Sale, Mohammadia School of engineering, Mohamed V University city of Rabat, Morocco et al., «Network Data Classification through Artificial Neural Networks and GenClust++ Algorithm», Int. J. Innov. Technol. Explor. Eng., vol. 8, n.o 10, pp. 2743-2750, ago. 2019, doi: 10.35940/ijitee.J9565.0881019.

[25] A. Das, S. N, N. M, G. Tiwary, y K. V, «An Enhanced Hybrid Deep Learning Model to Enhance Network Intrusion Detection Capabilities for Cybersecurity», J. Mach. Comput., pp. 472-486, abr. 2024, doi: 10.53759/7669/jmc202404045.

[26] Q. Zhou y C. Shi, «A Network Intrusion Detection Method for Various Information Systems Based on Federated and Deep Learning»:, Int. J. Semantic Web Inf. Syst., vol. 20, n.o 1, pp. 1-28, ene. 2024, doi: 10.4018/IJSWIS.335495.

[27] A. U. H. Qureshi, H. Larijani, M. Yousefi, A. Adeel, y N. Mtetwa, «An Adversarial Approach for Intrusion Detection Systems Using Jacobian Saliency Map Attacks (JSMA) Algorithm», Computers, vol. 9, n.o 3, p. 58, jul. 2020, doi: 10.3390/computers9030058.

[28] Y. Y. Aung y M. M. Min, «An Analysis of K-means Algorithm Based Network Intrusion Detection System», Adv. Sci. Technol. Eng. Syst. J., vol. 3, n.o 1, pp. 496-501, feb. 2018, doi: 10.25046/aj030160.

[29] Y. Hamid, M. Sugumaran, y L. Journaux, «A Fusion of Feature Extraction and Feature Selection Technique for Network Intrusion Detection», Int. J. Secur. Its Appl., vol. 10, n.o 8, pp. 151-158, ago. 2016, doi: 10.14257/ijsia.2016.10.8.13.

[30] T. A. Jasim Ali y M. M. Taher Jawhar, «Detecting network attacks model based on a convolutional neural network», Int. J. Electr. Comput. Eng. IJECE, vol. 13, n.o 3, p. 3072, jun. 2023, doi: 10.11591/ijece.v13i3.pp3072-3078.

[31] S. Alhassan, Dr. G. Abdul-Salaam, A. Micheal, Y. M. Missah, Dr. E. D. Ganaa, y A. S. Shirazu, «CFS-AE: Correlation-based Feature Selection and Autoencoder for Improved Intrusion Detection System Performance», J. Internet Serv. Inf. Secur., vol. 14, n.o 1, pp. 104-120, mar. 2024, doi: 10.58346/JISIS.2024.I1.007.

[32] «Improved feature selection method for features reduction in intrusion detection systems», Mesopotamian J. Cyber Secur., pp. 9-15, ene. 2021, doi: 10.58496/MJCS/2021/003.

[33] E. Tufan, C. Tezcan, y C. Acarturk, «Anomaly-Based Intrusion Detection by Machine Learning: A Case Study on Probing Attacks to an Institutional Network», IEEE Access, vol. 9, pp. 50078-50092, 2021, doi: 10.1109/ACCESS.2021.3068961.

[34] E. M. Maseno, Z. Wang, y H. Xing, «A Systematic Review on Hybrid Intrusion Detection System», Secur. Commun. Netw., vol. 2022, pp. 1-23, may 2022, doi: 10.1155/2022/9663052.

[35] A. A. Abdul Lateef, S. T. F. Al-Janabi, y B. Al-Khateeb, «Survey on intrusion detection systems based on deep learning», Period. Eng. Nat. Sci. PEN, vol. 7, n.o 3, p. 1074, ago. 2019, doi: 10.21533/pen.v7i3.635.

[36] D. N. Mhawi, A. Aldallal, y S. Hassan, «Advanced Feature-Selection-Based Hybrid Ensemble Learning Algorithms for Network Intrusion Detection

Systems», Symmetry, vol. 14, n.o 7, p. 1461, jul. 2022, doi: 10.3390/sym14071461.
[37]W. Guo, H. Qiu, Z. Liu, J. Zhu, y Q. Wang, «GLD-Net: Deep Learning to Detect DDoS Attack via Topological and Traffic Feature Fusion», Comput. Intell. Neurosci., vol. 2022, pp. 1-20, ago. 2022, doi: 10.1155/2022/4611331.
[38]M. Moukhafi, K. E. Yassini, y B. Seddik, «Intrusions detection using optimized support vector machine», Int. J. Adv. Appl. Sci., vol. 9, n.o 1, p. 62, mar. 2020, doi: 10.11591/ijaas.v9.i1.pp62-66.
[39]M. Arafah, I. Phillips, y A. Adnane, «Evaluating the impact of generative adversarial models on the performance of anomaly intrusion detection», IET Netw., vol. 13, n.o 1, pp. 28-44, ene. 2024, doi: 10.1049/ntw2.12098.
[40]J. L. Leevy, J. Hancock, R. Zuech, y T. M. Khoshgoftaar, «Detecting cybersecurity attacks across different network features and learners», J. Big Data, vol. 8, n.o 1, p. 38, dic. 2021, doi: 10.1186/s40537-021-00426-w.
[41]A. M. Basahel, M. Yamin, S. M. Basahel, y E. Laxmi Lydia, «Enhanced Coyote Optimization with Deep Learning Based Cloud-Intrusion Detection System», Comput. Mater. Contin., vol. 74, n.o 2, pp. 4319-4336, 2023, doi: 10.32604/cmc.2023.033497.
[42]A. Boukhalfa, N. Hmina, y H. Chaoni, «Parallel processing using big data and machine learning techniques for intrusion detection», IAES Int. J. Artif. Intell. IJ-AI, vol. 9, n.o 3, p. 553, sep. 2020, doi: 10.11591/ijai.v9.i3.pp553-560.
[43]觫car Mogoll髇-Guti開rez, J. Sancho N･ez, M. 羌ila Vegas, y A. Caro Lindo, «A Novel Ensemble Learning System for Cyberattack Classification», Intell. Autom. Soft Comput., vol. 37, n.o 2, pp. 1691-1709, 2023, doi: 10.32604/iasc.2023.039255.
[44]C. B. Cebi, F. S. Bulut, H. Firat, O. K. Sahingoz, y G. Karatas, «Deep Learning Based Security Management of Information Systems: A Comparative Study», J. Adv. Inf. Technol., pp. 135-142, 2020, doi: 10.12720/jait.11.3.135-142.
[45]G. Ketepalli y P. Bulla, «Review on Generative Deep Learning Models and Datasets for Intrusion Detection Systems», Rev. Intell. Artif., vol. 34, n.o 2, pp. 215-226, may 2020, doi: 10.18280/ria.340213.
[46]X. Tian, Z. Wu, J. Cao, S. Chen, y X. Dong, «ILIDViz: An Incremental Learning-Based Visual Analysis System for Network Anomaly Detection», Virtual Real. Intell. Hardw., vol. 5, n.o 6, pp. 471-489, dic. 2023, doi: 10.1016/j.vrih.2023.06.009.
[47]M. S p ń k , M. P w k , R. K k, y M. Ch r ś, « h App of Deep Learning Imputation and Other Advanced Methods for Handling Missing Values in Network Intrusion Detection», Vietnam J. Comput. Sci., vol. 10, n.o 01, pp. 1-23, feb. 2023, doi: 10.1142/S2196888822500257.
[48]C. M. N. C. M. Nalayini y J. K. C. M. Nalayini, «A New IDS for Detecting DDoS Attacks in Wireless Networks using Spotted Hyena Optimization and Fuzzy Temporal CNN», 網際網路技術學刊, vol. 24, n.o 1, pp. 023-034, ene. 2023, doi: 10.53106/160792642023012401003.
[49]R. J. Alzahrani y A. Alzahrani, «Security Analysis of DDoS Attacks Using Machine Learning Algorithms in Networks Traffic», Electronics, vol. 10, n.o 23, p. 2919, nov. 2021, doi: 10.3390/electronics10232919.
[50]M. Ramzan et al., «Distributed Denial of Service Attack Detection in Network Traffic Using Deep Learning Algorithm», Sensors, vol. 23, n.o 20, p. 8642, oct. 2023, doi: 10.3390/s23208642.
[51]I. AlSaleh, A. Al-Samawi, y L. Nissirat, «Novel Machine Learning Approach for DDoS Cloud Detection: Bayesian-Based CNN and Data Fusion Enhancements», Sensors, vol. 24, n.o 5, p. 1418, feb. 2024, doi: 10.3390/s24051418.
[52]V. Hnamte y J. Hussain, «An Efficient DDoS Attack Detection Mechanism in SDN Environment». 29 de diciembre de 2022. doi: 10.21203/rs.3.rs-2393388/v1.
[53]N. Lutsiv et al., «Deep Semisupervised Learning-Based Network Anomaly Detection in Heterogeneous Information Systems», Comput. Mater. Contin., vol. 70, n.o 1, pp. 413-431, 2022, doi: 10.32604/cmc.2022.018773.
[54]N. Z. Gorment, A. Selamat, L. K. Cheng, y O. Krejcar, «Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions», IEEE Access, vol. 11, pp. 141045-141089, 2023, doi: 10.1109/ACCESS.2023.3256979.
[55]H. Liu y B. Lang, «Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey», Appl. Sci., vol. 9, n.o 20, p. 4396, oct. 2019, doi: 10.3390/app9204396.