ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.1124

Phishing in the Digital Era: A Systematic Review of the Most Promising Detection Techniques

Jhon Alonso Paucar Cordova[1],
[1]Universidad Tecnológica del Perú, Perú, *U21224301@utp.edu.pe*

Abstract*-- Phishing is a growing threat that surpasses traditional detection methods, requiring advanced approaches. The aim of this review was to identify the most promising emerging techniques for phishing detection and assess their effectiveness in overcoming the limitations of traditional methods. A systematic search was conducted using the PICO strategy in academic databases such as Scopus, Ridely, and SciELO. Out of a total of 293 identified articles, 30 studies published between 2020 and 2024 were selected after applying inclusion and exclusion criteria. Hybrid models in parallel execution (Random Forest, Naive Bayes, CNN, and LSTM) achieved accuracy rates above 99.97%, standing out as the most effective techniques in critical sectors such as finance and corporate environments. These technologies overcome traditional limitations by combining advanced Machine Learning and Deep Learning capabilities. Additionally, vulnerable groups were identified as users of banking and e-commerce services, due to their constant exposure to online transactions; users in governmental and financial sectors, who are key targets because of the sensitive information they handle; and young users (18-25 years old), particularly those with high interaction on social media and limited cybersecurity knowledge. Hybrid models in parallel execution represent a significant advancement in phishing detection, but their effectiveness depends on factors such as the computational load they require and their limited accessibility for everyone. Furthermore, these techniques will be fully effective if complemented with cybersecurity education that fosters user preparedness.

***Keywords*-- Phishing, Phishing Detection, Machine Learning, Deep Learning, Emerging Techniques.**

Phishing en la era digital: una revisión sistemática de las técnicas de detección más prometedoras

Jhon Alonso Paucar Cordova[1],
[1]Universidad Tecnológica del Perú, Perú, *U21224301@utp.edu.pe*

Resumen*-- El phishing es una amenaza creciente que supera los métodos tradicionales de detección, exigiendo enfoques avanzados. El objetivo de esta revisión fue identificar las técnicas emergentes más prometedoras en la detección de phishing y evaluar su efectividad para superar las limitaciones de los métodos tradicionales. Se realizó una búsqueda sistemática utilizando la estrategia PICO en bases de datos académicas como Scopus, Redalyc y SciELO. De un total de 293 artículos identificados, se seleccionaron 30 estudios publicados entre 2020 y 2024 tras aplicar criterios de inclusión y exclusión. Los modelos híbridos en ejecución paralela (Random Forest, Naive Bayes, CNN y LSTM) alcanzaron precisiones superiores al 99.97%, destacándose como las técnicas más efectivas en sectores críticos como el financiero y corporativo. Estas tecnologías superan las limitaciones tradicionales al combinar capacidades avanzadas de Machine Learning y Deep Learning. Además, se identificaron como grupos vulnerables a los usuarios de servicios bancarios y comercio electrónico, debido a su exposición constante a transacciones en línea; a usuarios en sectores gubernamentales y financieros, quienes son objetivos clave por la información sensible que manejan; y a usuarios jóvenes (18-25 años), especialmente aquellos con alta interacción en redes sociales y menor conocimiento en ciberseguridad. Los modelos híbridos en ejecución paralela representan un avance significativo para la detección de phishing, pero su efectividad depende de factores como la carga computacional que requieren y su limitada accesibilidad para todas las personas. Asimismo, estas técnicas serán plenamente efectivas si se complementan con educación en ciberseguridad que fomente la preparación de los usuarios.

***Palabras claves*-- Phishing, Detección de Phishing, Aprendizaje Automático, Aprendizaje profundo, Técnicas emergentes.**

I. INTRODUCCIÓN

En el ámbito de la ciberseguridad, el phishing se ha consolidado como una de las amenazas más persistentes y sofisticadas a nivel global. Este tipo de ataque se basa en el engaño para obtener información confidencial, como contraseñas y datos financieros, mediante técnicas de ingeniería social y la creación de sitios web falsos [1]. La creciente dependencia de Internet ha facilitado que los ataques de phishing se realicen de manera anónima y a gran escala, haciendo más grave aún el problema [1]. Esto resalta la necesidad urgente de enfoques más avanzados para identificar y prevenir eficazmente estos ataques [2]. El número de ataques de phishing ha aumentado drásticamente en el último año, y los métodos tradicionales de detección han demostrado ser insuficientes para contrarrestar la sofisticación de estos ataques [2].

Aunque se han logrado avances significativos en el desarrollo de nuevas técnicas de detección, muchas de ellas aún enfrentan limitaciones importantes [1]. En este contexto, tecnologías emergentes como el Machine Learning y el Deep Learning se presentan como alternativas prometedoras para diseñar técnicas más robustas frente al phishing [3]. No obstante, su implementación sigue enfrentando desafíos clave, como la adaptabilidad a amenazas en constante evolución y la reducción de falsos positivos [1]. Un análisis reciente sobre el rendimiento de algoritmos de clasificación en la detección de sitios web de phishing revela que, aunque estos métodos ofrecen perspectivas valiosas, su efectividad aún requiere evaluación en contextos diversos y realistas [2]. En esta misma línea, se ha propuesto un enfoque basado en word embeddings combinado con algoritmos de clasificación como Random Forest, logrando precisiones superiores al 99%; no obstante, se señala que estos modelos requieren ser entrenados con datos reales para alcanzar una mejor capacidad de generalización [15]. Asimismo, se ha evidenciado que técnicas tradicionales como las listas negras y los sistemas heurísticos siguen siendo ampliamente superadas por modelos supervisados como Support Vector Machines (SVM) y árboles de decisión, los cuales han demostrado un desempeño más robusto y preciso en distintos entornos de prueba [11]. Así, Las técnicas emergentes en la detección de phishing presentan un notable potencial para mejorar la precisión y minimizar los falsos positivos [3]. Sin embargo, aunque se ha evidenciado un incremento en las publicaciones que buscan consolidar estas técnicas, aún persiste la falta de una visión integral que abarque tanto sus avances como los desafíos pendientes. Esta carencia se debe, en parte, a la ausencia de estudios de revisión sistemática que sinteticen y analicen de manera estructurada los resultados obtenidos hasta la fecha. En respuesta a esta brecha, la presente revisión tiene como propósito analizar de forma crítica y comparativa las tecnologías y métodos emergentes utilizados en la detección de ataques de phishing, identificando los enfoques más prometedores.

A través de este análisis, se busca determinar cuáles técnicas poseen el mayor potencial para mejorar la precisión, reducir los falsos positivos y enfrentar eficazmente los ataques de phishing en diversos contextos. En tal sentido, el documento está organizado de la siguiente manera. La Sección II, presenta la metodología utilizada para realizar la RSL, detallando el enfoque PICO [4] y las estrategias de búsqueda empleadas. La Sección III presenta los resultados obtenidos, desglosados en

análisis bibliométrico y de contenido, respondiendo las preguntas planteadas en la metodología. En la Sección IV, se discuten los hallazgos clave, incluyendo las fortalezas y limitaciones de las técnicas emergentes analizadas. Finalmente, en la Sección V, se presentan las conclusiones del estudio, destacando su relevancia y las áreas futuras de investigación.

II. METODOLOGÍA

Con el objetivo de identificar las técnicas más prometedoras en la detección de phishing, se llevó a cabo una revisión sistemática, organizando y dirigiendo la búsqueda de manera estricta. Se aplicó la estrategia PICO [4] como guía en el proceso, definiendo los pasos necesarios para obtener la información relevante que permitiera abordar la investigación de forma estructurada. Siguiendo esta estrategia, se formularon subpreguntas basadas en los componentes del acrónimo PICO y se seleccionaron las palabras clave más adecuadas. Posteriormente, se diseñaron ecuaciones de búsqueda, con la finalidad de obtener estudios de diversas bases de datos académicas.

El desarrollo de la estrategia ha sido el siguiente: en la Tabla I, se presenta la generación de la pregunta PICO, tomando como guía el tema de investigación junto a sus componentes específicos según el acrónimo. A partir de ello, se formularon las cubrejuntas correspondientes. En la Tabla II, se realizó la selección adecuada de las palabras clave basadas en esta información. Finalmente, en la Tabla III, se describen las ecuaciones de búsqueda, que incluyen los términos clave, tanto en inglés como en español, para su uso en distintas bases de datos académicas. A continuación, se presenta:

TABLA I – Descripción de la pregunta PICO

Tema de investigación: Avances en la Detección de Phishing: Evaluación de tecnologías emergentes y nuevas tendencias.	
Pregunta general: ¿Cuáles son las técnicas más prometedoras para detectar el phishing?	
Acrónimo y componente	Subpreguntas
P: Usuarios afectados por ataques de phishing.	¿Quiénes son los usuarios más afectados por el phishing?
I: Técnicas de detección de phishing.	¿Cuáles son las técnicas o métodos más efectivos para la detección de phishing?
C: Comparación con métodos tradicionales de seguridad.	¿En qué aspectos las técnicas emergentes superan a las estrategias tradicionales en la detección y prevención de ataques de phishing?
O: Efectividad de las técnicas emergentes en prevenir ataques.	¿Qué tan efectivas son las técnicas emergentes para evitar que los ataques de phishing se concreten?

TABLA II – Descripción de las palabras clave

Acrónimo y componente	Palabras clave
P: Usuarios afectados por ataques de phishing.	Phishing, victims, users.
I: Técnicas emergentes de detección de phishing.	Detection, methods, techniques.
C: Comparación con métodos tradicionales de seguridad.	Attacks, comparing, security, traditional.
O: Efectividad de las técnicas emergentes en prevenir ataques.	Evaluation, performance, efficient.

TABLA III – Descripción de las ecuaciones de búsqueda

Acrónimo	Ecuación de búsqueda
P	phishing AND victims OR users
I	detection AND methods OR techniques
C	attacks OR comparing AND security
O	evaluation OR performance OR efficient
EB1 - SCOPUS: (Phishing AND victims OR users) AND (detection AND methods OR techniques) AND (attacks OR comparing and security) (evaluation OR performance OR efficient).	
EB2 - REDALYC: phishing AND attacks OR methods	
EB3 - SCIELO: phishing AND attacks OR methods OR performance	

Se llevó a cabo la búsqueda de artículos utilizando una ecuación de búsqueda diseñada específicamente para las bases de datos elegidas como SCOPUS (EB1), REDALYC (EB2) y SCIELO (EB3). Se definieron criterios de elegibilidad que incluyeron tanto criterios de inclusión como de exclusión, con el objetivo de garantizar que los estudios considerados fueran pertinentes para el tema de detección de phishing.

TABLA IV – Descripción de los criterios de elegibilidad

N°	CRITERIOS DE INCLUSIÓN	N°	CRITERIOS DE EXCLUSIÓN
CI1	Artículos que se centren en técnicas emergentes o innovadoras de anti-phishing.	CE1	Artículos que no describan claramente la metodología utilizada.
CI2	Artículos que evalúen la efectividad de técnicas emergentes en comparación con métodos tradicionales.	CE2	Artículos publicados en cualquier idioma que no sea inglés o español.
CI3	Artículos que mencionen las ventajas y desventajas de las técnicas emergentes en la detección de phishing.	CE3	Artículos publicados antes del año 2020.
CI4	Artículos que discutan la importancia de incorporar nuevas técnicas de detección de phishing.	CE4	Artículos que no cuenten con acceso al texto completo.

A partir de las búsquedas realizadas, se identificaron 293 artículos en total: 212 en Scopus, 19 en Scielo y 62 en Redalyc. Tras eliminar 2 duplicados, se prosiguió con el cribado, donde 215 artículos fueron excluidos por no cumplir con la temática de la revisión sistemática.

Se encontraron 76 artículos pertinentes, pero 30 no estaban disponibles en texto completo, dejando 46 artículos para la evaluación de elegibilidad. Se aplicaron criterios de exclusión que resultaron en la eliminación de 16 artículos: aquellos que no describen claramente la metodología utilizada, los publicados en idiomas distintos al inglés o español, y los publicados hace más de 4 años. Finalmente, se incluyeron 30 artículos en la revisión sistemática, desglosándose en 26 de Scopus, 2 de Scielo y 2 de Redalyc.

A continuación, se presenta el diagrama PRISMA [5], que ilustra el proceso de selección de los 30 artículos restantes. Estos artículos, cuidadosamente seleccionados, servirán como base fundamental para elaborar esta revisión sistemática de la literatura, proporcionando una comprensión profunda de las técnicas emergentes en la detección de phishing y sus contribuciones al campo.

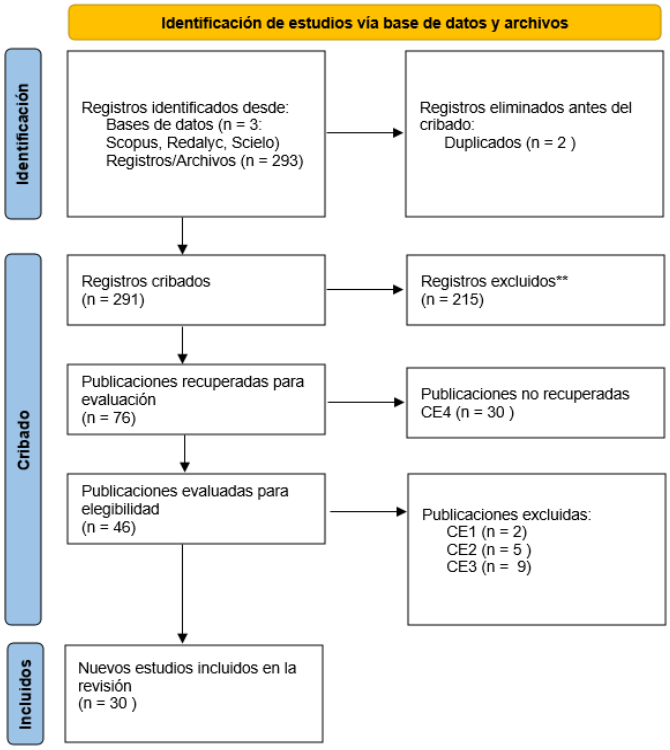

Fig. 1. Diagrama de flujo PRISMA del proceso de selección de artículos [5].

III. RESULTADOS

Para la elaboración de la fase de resultados, esta se ha dividido en dos apartados: datos bibliométricos y datos de contenido. El apartado de datos bibliométricos incluye una tabla con los artículos seleccionados, un gráfico que muestra la distribución de los artículos por año, el tipo de artículo, la metodología empleada y su distribución geográfica. En el apartado de datos de contenido, se responden las preguntas planteadas en la metodología PICO [4], utilizando la información extraída de los artículos seleccionados.

DATOS BIBLIOMÉTRICOS

Respecto a las investigaciones publicadas sobre phishing entre 2020 y 2024, se observa un aumento en el número de publicaciones, comenzando con 4 en 2020, seguido de fluctuaciones con 6 en 2021, 5 en 2022, 7 en 2023, y alcanzando 8 en 2024. La línea refleja una tendencia ascendente a lo largo de estos años. Esta evolución se presenta en el gráfico a continuación.

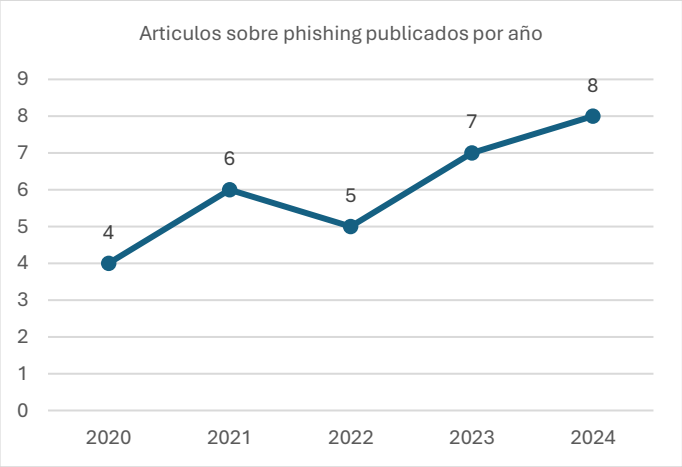

Fig. 2. Número de artículos sobre phishing publicados por año.

La distribución de los tipos de artículos seleccionados para esta revisión muestra que un 75% corresponde a artículos de investigación, mientras que el 25% restante son artículos de revisión. Se aprecia que la mayoría de los estudios en el campo del phishing están orientados hacia investigaciones empíricas o experimentales. A continuación, se presenta un gráfico que ilustra esta distribución.

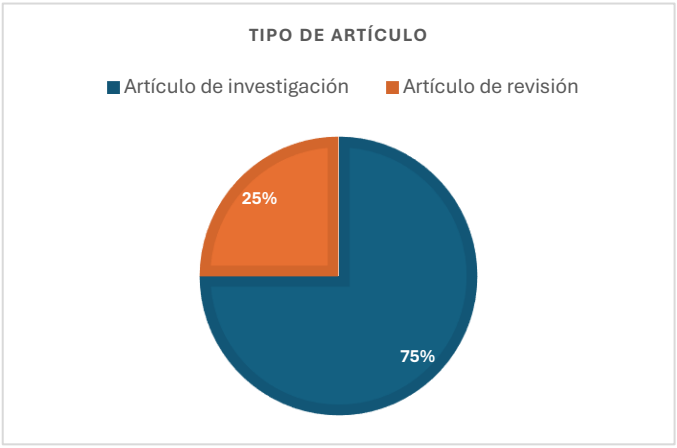

Fig. 3. Distribución cuantitativa de artículos seleccionados según el tipo de publicación.

Se observa que la mayoría de los estudios revisados sobre detección de phishing emplearon metodologías cuantitativas, alcanzando un total de 28 artículos, lo que refleja una clara predominancia de este enfoque. Por otro lado, solo 2 artículos utilizaron metodologías cualitativas, lo que evidencia una menor representación de este tipo de estudios en el campo. A continuación, se incluye un gráfico que representa esta distribución.

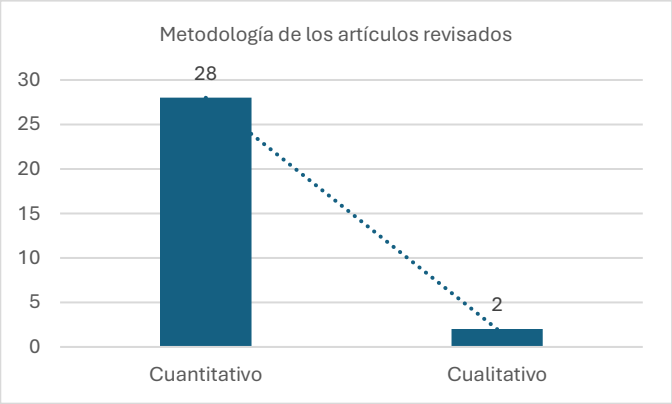

Fig. 4. Distribución de artículos según el tipo de metodología utilizada.

La distribución geográfica de los artículos incluidos en la Revisión Sistemática de la Literatura entre 2020 y 2024 muestra que India es el país con mayor cantidad de publicaciones, seguido por Arabia Saudita y Omán. Otros países como Pakistán, Reino Unido y Turquía también contribuyen, junto con otras naciones que tienen menor número de publicaciones. Esta información se detalla en la figura que se presenta continuación.

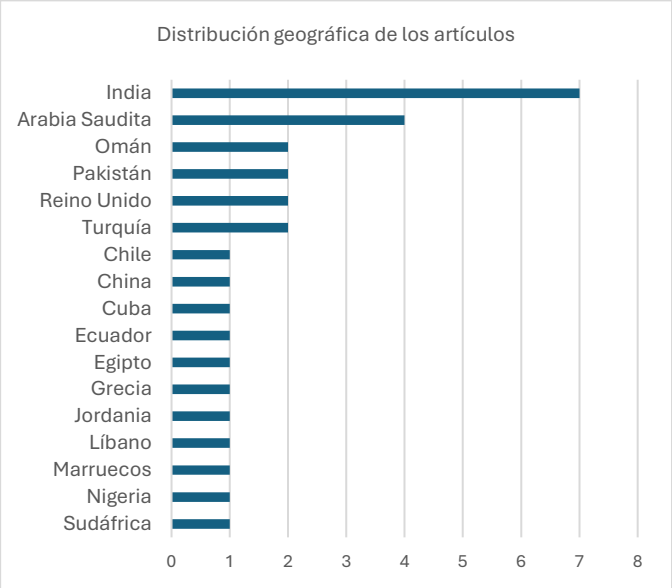

Fig 5. Distribución de artículos por país en la RSL.

DATOS DE CONTENIDO

A. ¿Quiénes son los usuarios más afectados por el phishing?

Según lo recopilado, los ataques de phishing afectan a una amplia variedad de usuarios, pero algunos grupos son particularmente propensos. Los usuarios de redes sociales son especialmente vulnerables al phishing, ya que estos entornos facilitan tácticas de ingeniería social que explotan la confianza de los usuarios [7].

En particular, los jóvenes entre 18 y 25 años, con menor conocimiento en ciberseguridad, están más expuestos debido a su frecuente interacción en estas plataformas y a su menor preparación en temas de seguridad digital [6, 7, 11]. Los usuarios de servicios bancarios, comercio electrónico y aquellos que realizan transacciones en línea son particularmente atractivos para los atacantes de phishing debido a la naturaleza sensible de la información financiera que manejan [8, 12, 13, 30].

Otros estudios señalan que los empleados de instituciones gubernamentales, organizaciones de servicios y sectores como las finanzas, servicios profesionales y manufactura representan objetivos valiosos debido al alto valor de los datos que manejan, convirtiéndolos en blancos recurrentes de ataques [10, 15]. Además, los usuarios de dispositivos móviles, especialmente aquellos que acceden a servicios en línea desde teléfonos inteligentes, tienen una exposición elevada debido a los enlaces y aplicaciones potencialmente maliciosas [16]. Finalmente, varios estudios concluyen que las personas con poco conocimiento en seguridad digital, que son propensas a cometer errores tipográficos o carecen de entrenamiento en ciberseguridad, también enfrentan un riesgo alto, ya que tienden a confiar fácilmente en remitentes desconocidos [19, 31, 32].

En particular, los usuarios de servicios financieros y plataformas de pago son un grupo de alto riesgo, como lo confirma una investigación de la APWG, que los identifica como objetivos clave para los atacantes de phishing [2].

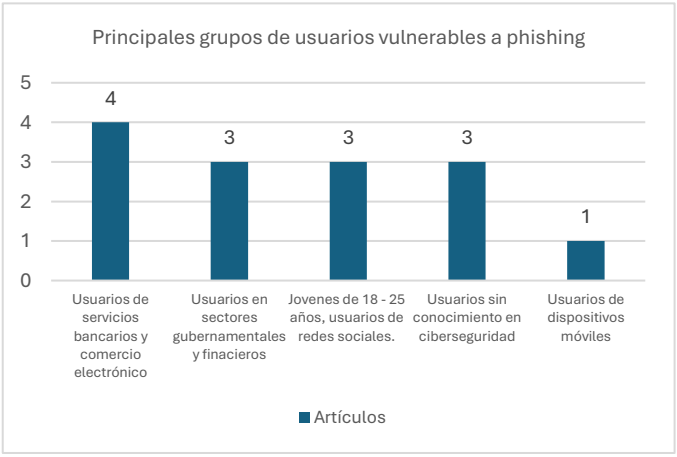

Fig 6. Grupos de usuarios más afectados por ataques de phishing según el número de estudios.

B. ¿Cuáles son las técnicas o métodos más efectivos para la detección de phishing?

Respecto a las técnicas más efectivas para la detección de phishing, diversos estudios destacan el uso de métodos de machine learning y deep learning [10, 11, 12, 19, 22, 29, 34]. Es importante mencionar que se han considerado las técnicas más mencionadas en los artículos revisados, ya que su recurrencia sugiere relevancia y efectividad reportada en diferentes contextos, sin especificar si son técnicas emergentes o consolidadas.

Entre dichas técnicas se encuentran las redes neuronales y algoritmos de clasificación como el Random Forest y las Máquinas de Vectores de Soporte (SVM, por sus siglas en inglés), que han mostrado alta precisión en la detección de sitios de phishing [10, 11, 15, 19]. Adicionalmente, el aprendizaje profundo se aplica en modelos como las redes neuronales convolucionales (CNN, Convolutional Neural Networks), redes neuronales recurrentes (RNN, Recurrent Neural Networks) y técnicas avanzadas de codificación a nivel de caracteres y palabras, logrando mejorar la capacidad de detección al analizar patrones de URLs y textos asociados a phishing [17, 29, 34].

Métodos específicos como el uso de listas negras y análisis de tráfico también se mencionan, aunque su efectividad puede ser limitada en comparación con algoritmos avanzados de meta-aprendizaje como los Meta-Learners (modelos que aprenden a aprender y adaptarse rápidamente a nuevas tareas) combinados con el Algoritmo de Árboles Extra (Una variante de los árboles de decisión que mejora la precisión mediante la inyección de aleatoriedad) [7, 9].

Asimismo, técnicas de detección basadas en inteligencia de amenazas cibernéticas y en aprendizaje de múltiples vistas con Transformadores optimizados, como el modelo TB-DBN (Red de Creencias Profundas Basada en Transformadores), son frecuentemente mencionadas [14, 24, 28]. También aparecen enfoques híbridos como el uso de LSTM-CNN (Memoria a Largo y Corto Plazo combinada con Redes Neuronales Convolucionales) con el Algoritmo de Optimización del Buitre Africano (AVO, por sus siglas en inglés) y SqueezeNet optimizado con FDHPO (Optimización de Hiperparámetros de Decisión Difusa) que han demostrado ser altamente efectivos al incrementar la precisión y reducir el tiempo de procesamiento en la detección de ataques de phishing [30, 33].

Las técnicas de Machine Learning y Deep Learning han ganado protagonismo por su alta precisión y capacidad de adaptación. La siguiente tabla sintetiza los métodos más destacados.

TABLA V – Técnicas de Machine Learning y Deep Learning destacadas

Técnica principal	Métodos destacados
Machine Learning	Random Forest, SVM (Máquinas de Vectores de Soporte)
Deep Learning	CNN (Redes Neuronales Convolucionales), RNN (Redes Neuronales Recurrentes)

Complementando la información presentada en la anterior tabla, el siguiente gráfico ofrece una visión más amplia al incluir una mayor variedad de técnicas y destacar la frecuencia con la que han sido mencionadas en los estudios revisados.

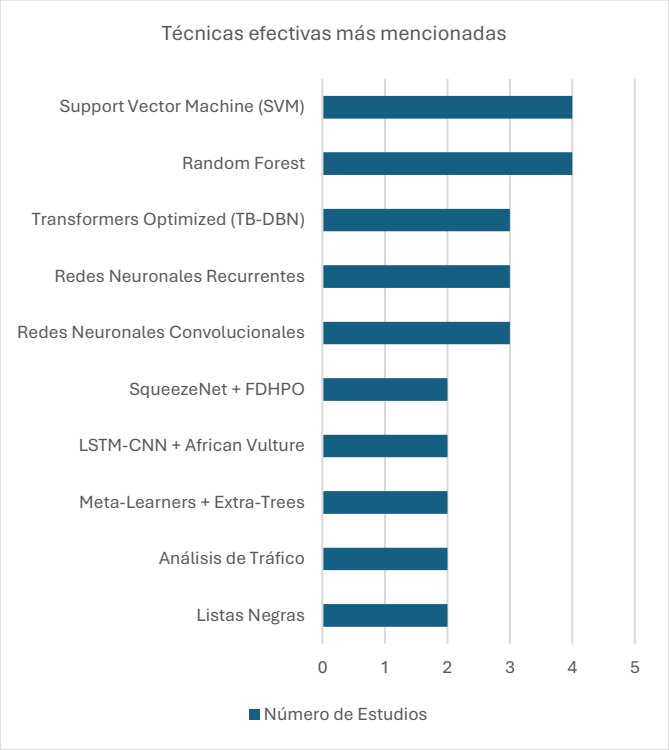

Fig. 7. Número de estudios que respaldan cada técnica para la detección de phishing.

C. *¿En qué aspectos las técnicas emergentes superan a las estrategias tradicionales en la detección y prevención de ataques de phishing?*

Para responder a la pregunta sobre los aspectos en los que las técnicas emergentes superan a las estrategias tradicionales en la detección y prevención de ataques de phishing, se han clasificado las técnicas emergentes en tres categorías: Machine Learning, Deep Learning y Técnicas Híbridas. Esta clasificación permite identificar las principales ventajas de cada enfoque frente a los métodos tradicionales. Las técnicas de machine learning destacan en la detección de URLs no reportadas, superando las limitaciones de las listas negras al ofrecer una mayor precisión y reduciendo la tasa de falsos positivos [7, 9, 10]. Modelos avanzados como ABET (Herramienta de Mejora Automatizada de Listas Negras) y LBET (Herramienta de Evaluación Basada en Aprendizaje) destacan en esta categoría, ya que logran reducir los falsos positivos y mejorar la precisión [9]. También, la técnica de lista blanca automatizada permite una mejor gestión de datos confiables, minimizando los falsos positivos y manteniendo una base de URLs seguras [13]. En enfoques específicos, FastText junto con Random Forest es eficaz en la detección de phishing en correos electrónicos, mejorando la precisión [14, 15].

Por otro lado, las técnicas de deep learning aportan ventajas en términos de velocidad y capacidad para analizar patrones complejos en tiempo real, especialmente al trabajar con grandes volúmenes de datos. Técnicas como RNN-LSTM (Red Neuronal Recurrente - Memoria a Largo y Corto Plazo) permiten identificar URLs maliciosas en tiempo real, mejorando así la respuesta ante amenazas de phishing [12]. Además, modelos optimizados como LSTM-GRU (Memoria a Largo y Corto Plazo - Unidad de Recurrencia con Puerta) y SqueezeNet con FDHPO (Optimización de Hiperparámetros impulsada por Características) sobresalen por su capacidad de detección rápida y precisa, incluso en escenarios de phishing de gran escala y constante cambio [30], [34]. Estos modelos permiten una adaptación superior frente a amenazas que evolucionan rápidamente [26, 27, 28, 29, 31, 33].

Respecto a las técnicas híbridas combinan lo mejor de ambos enfoques, logrando adaptabilidad y robustez frente a nuevas amenazas. Los modelos en ejecución paralela (Random Forest, Naive Bayes, CNN y LSTM) han alcanzado una precisión superior al 99.97%, integrando capacidades avanzadas para maximizar la detección de phishing [23]. Asimismo, herramientas como HELPHED han mostrado ser prometedoras en entornos corporativos [22], mientras que combinaciones como BMEO + k-NN y Transformers destacan por su capacidad de reducir la necesidad de actualizaciones manuales y adaptarse rápidamente a patrones cambiantes [17, 18, 21].

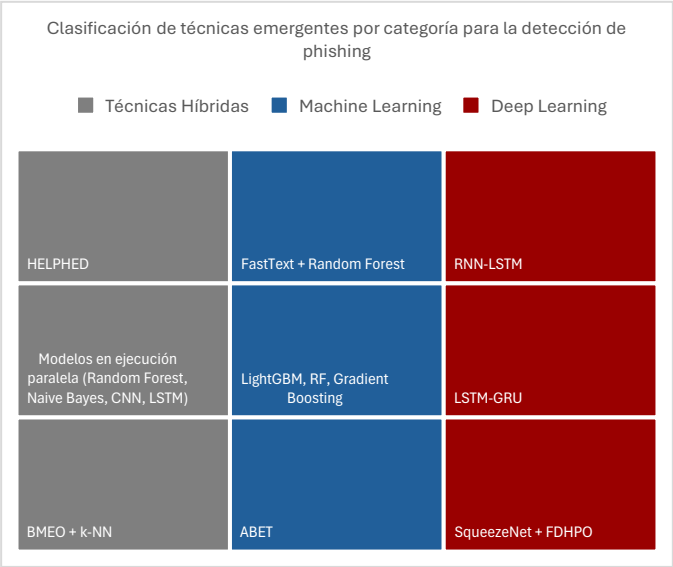

Fig. 8. Clasificación de técnicas emergentes.

A continuación, se muestran los aspectos clave en los que estas técnicas avanzadas ofrecen un mejor rendimiento comparado con los métodos tradicionales.

TABLA VI – Aspectos en los que las técnicas emergentes superan a las estrategias tradicionales.

Técnica	Aspecto 1	Aspecto 2	Aspecto 3
Machine Learning	Detección de URLs no reportadas con alta precisión.	Reducción de falsos positivos.	Actualización dinámica de bases de datos de phishing.
Deep Learning	Análisis de patrones complejos en tiempo real.	Capacidad para gestionar grandes volúmenes de datos.	Adaptación a tácticas de phishing en evolución.
Técnicas Híbridas	Adaptabilidad ante ataques nuevos y avanzados.	Integración de múltiples enfoques para mayor robustez.	Reducción de la necesidad de reentrenamiento constante.

No obstante, a pesar de sus ventajas frente a los métodos tradicionales, las técnicas emergentes también presentan limitaciones relevantes que condicionan su aplicabilidad. Por ejemplo, requieren una infraestructura tecnológica avanzada y capacidad de procesamiento elevado, lo que puede ser una barrera para organizaciones con recursos limitados o para dispositivos móviles [16, 30]. Además, muchas de estas técnicas dependen de grandes volúmenes de datos etiquetados y actualizados para mantener su efectividad, lo cual no siempre es factible en escenarios reales [25]. Algunos modelos también muestran dificultades para generalizar ante ataques novedosos o variaciones no vistas durante su entrenamiento, lo que puede afectar su desempeño frente a amenazas en constante evolución [9]. Finalmente, su complejidad algorítmica puede dificultar su adopción en entornos donde se requiere transparencia, interpretabilidad o implementación rápida, como en pequeñas empresas o instituciones educativas [16, 25].

D. ¿Qué tan efectivas son las técnicas emergentes para evitar que los ataques de phishing se concreten?

Las técnicas emergentes para la detección de phishing han demostrado ser altamente efectivas, superando en muchos casos a los métodos tradicionales, según estudios que muestran precisiones superiores al 95% en diversos enfoques, como el uso de modelos de aprendizaje profundo y algoritmos avanzados [9, 10, 11]. Modelos avanzados como los Meta-Learners combinados con el Algoritmo de Árboles Extra alcanzan una precisión de hasta el 97.5%, manteniendo una baja tasa de falsos positivos [9]. En general, los modelos de aprendizaje profundo presentan una efectividad superior al 95% en la identificación y bloqueo de ataques de phishing [10, 11]. Un caso destacado es el modelo RNN-LSTM (Red Neuronal Recurrente con Memoria a Largo y Corto Plazo), que logra una precisión superior al 97% en la detección de URLs maliciosas en tiempo real [12]. Métodos específicos como la lista blanca automatizada también muestran alta efectividad, con una precisión promedio de 96.17% y una tasa de verdaderos positivos del 95% [13].

La plataforma basada en inteligencia de amenazas cibernéticas alcanza un 97% de precisión, con tasas de verdaderos positivos que oscilan entre el 96% y el 98% en la detección de sitios sospechosos [14].

En cuanto a enfoques híbridos, sin duda alguna el que predomina es el modelo que combina técnicas de aprendizaje automático y profundo en ejecución paralela (Random Forest, Naive Bayes, CNN y LSTM) llegando a tener una precisión del 99.97%, también tenemos el uso de FastText (modelo de aprendizaje de palabras eficiente que permite una representación rápida y precisa de texto) con Random Forest (algoritmo de ensamble basado en múltiples árboles de decisión para mejorar la precisión) ha demostrado una precisión del 99.5% y una tasa de falsos positivos muy baja en la detección de correos electrónicos de phishing [15].

En el caso de phishing en dispositivos móviles, la técnica APuML (Advanced Phishing URL Machine Learning, un enfoque avanzado de machine learning especializado en la detección de URLs de phishing) con Random Forest alcanza un 93.85% de precisión en la detección de phishing, con una tasa de verdaderos positivos del 93.22% [16]. Técnicas avanzadas como el modelo DNN (Red Neuronal Profunda) a nivel de codificación de caracteres logran una precisión del 98.13% en la detección de URLs de phishing [17].

El sistema BMEO (Optimización basada en Entornos de Multiobjetivos) con k-NN (k-Nearest Neighbors) alcanza una efectividad del 98%, lo cual demuestra una alta capacidad para identificar sitios maliciosos [18]. Las redes neuronales profundas también mantienen un 97% de precisión en la detección de correos de phishing, evidenciando su efectividad en conjuntos de datos complejos [19]. La optimización de modelos a través del ajuste de hiperparámetros permite alcanzar hasta un 98.27% de precisión en la detección de URLs de phishing [20].

El uso de modelos optimizados con técnicas como LightGBM (Light Gradient Boosting Machine, un algoritmo de boosting que es eficiente en la gestión de grandes volúmenes de datos y de menor tiempo de entrenamiento), Random Forest y Gradient Boosting (técnica de ensamble que combina modelos débiles secuencialmente para crear un modelo fuerte) permite alcanzar una precisión promedio de 99.7% en la identificación de ataques de phishing [21]. Otros enfoques híbridos, como HELPHED, muestran una precisión del 99.43% y un F1-score de 0.9942 en la detección de correos electrónicos de phishing [22].

La combinación de técnicas de machine learning y deep learning en ejecución paralela logra precisiones de hasta el 99.97%, permitiendo una detección altamente efectiva de URLs maliciosas [23].

Por su parte, el enfoque basado en Transformadores optimizados con mezcla de expertos mantiene una precisión superior al 96%, incluso después de varios meses de cambios en los datos [24]. El aprendizaje continuo (continual learning) también resulta altamente efectivo, con técnicas como "learning without forgetting" (LWF) y consolidación de pesos elásticos (EWC, Elastic Weight Consolidation), manteniendo precisiones de entre 93% y 95% a lo largo del tiempo, con una disminución mínima del rendimiento [25].

Modelos como Gradient Boosting (técnica de ensamble que construye un modelo fuerte a partir de múltiples modelos débiles) y Random Forest, alcanzan precisiones de 97.2% y 97.1%, respectivamente, en la detección de sitios de phishing [26].

PhiUSIIL, un framework basado en índice de similitud y aprendizaje incremental (diseñado para actualizarse continuamente sin necesidad de reentrenar desde cero), logra una precisión de hasta 99.79% en preentrenamiento y 99.24% en entrenamiento incremental, manteniendo su efectividad frente a técnicas de phishing avanzadas [27]. El modelo TB-DBN (Red de Creencias Profundas Basada en Transformadores) alcanza una precisión del 99.4% y altos puntajes de recall y F1-score, demostrando su eficacia en la prevención de ataques de phishing [28].

Las redes convolucionales empleadas en el sistema DEPHIDES alcanzan una precisión del 98.74%, lo que subraya su capacidad en la detección de URLs de phishing [29]. La técnica SqueezeNet optimizada con FDHPO (Optimización de Hiperparámetros de Decisión Difusa) alcanza una precisión del 93.05% y una sensibilidad del 94.26% [30]. El algoritmo de clustering mejorado de K-Means logra una precisión del 89.2% en conjuntos de datos pequeños, mostrando altos niveles de efectividad en diferentes tamaños de conjuntos de datos [31].

La combinación de LSTM-CNN (Memoria a Largo y Corto Plazo y Redes Convolucionales) con el Algoritmo de Optimización del Buitre Africano (AVOA) alcanza una precisión de hasta el 99.37%, superando otros métodos en la detección de phishing en múltiples conjuntos de datos [33]. El modelo basado en LSTM-GRU (Memoria a Largo y Corto Plazo y Unidad de Puerta Recurrente) logra una precisión del 98.89%, demostrando ser uno de los enfoques más efectivos en la clasificación de URLs de phishing [34]. Las técnicas emergentes más precisas en la detección de phishing han alcanzado eficacias superiores al 99%, demostrando su superioridad frente a los métodos tradicionales [9, 21, 23].

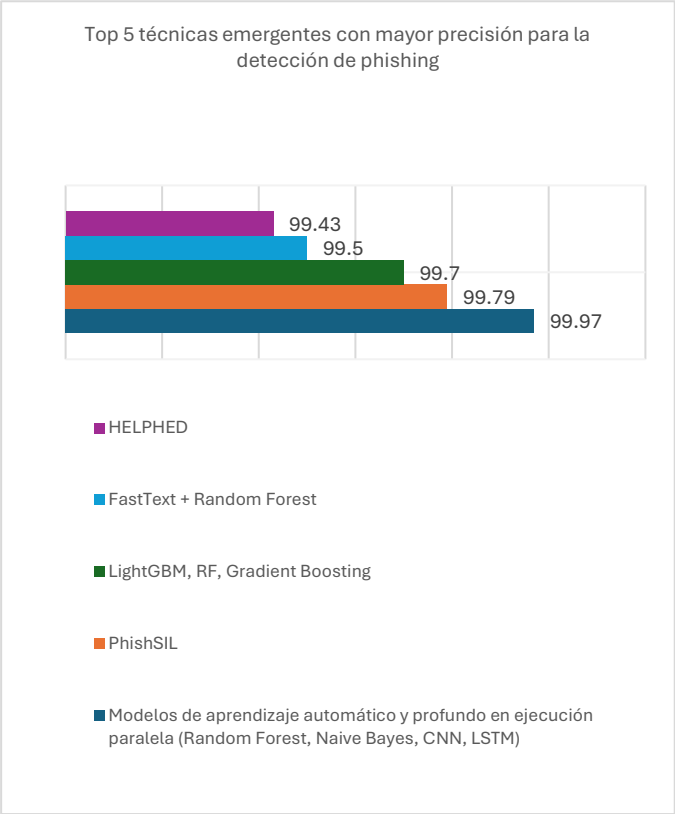

Fig. 9. Técnicas emergentes con mayor precisión

Es importante mencionar que una técnica de detección de phishing se considera efectiva con precisiones superiores al 90%-95% [22]. En sectores críticos, como el financiero, se espera superar el 97%-99%, con bajas tasas de falsos positivos, para garantizar tanto la precisión como la confiabilidad del sistema [23, 24]. Para los usuarios comunes, es importante que las técnicas mantengan una precisión superior al 90% [25].

Después del análisis de las cinco técnicas emergentes más precisas, se detalla a continuación una tabla que recopila un total de 23 técnicas, junto con sus porcentajes de precisión reportados.

Tabla VII – Efectividad de técnicas emergentes para la prevención de ataques de phishing

N.º	Técnica	Precisión (%)	Aplicación Principal
1	Modelos de aprendizaje automático y profundo en ejecución paralela (Random Forest, Naive Bayes, CNN, LSTM)	99.97	Correos electrónicos y URLs
2	PhishUSILL	99.79	URLs
3	LightGBM, RF, Gradient Boosting	99.70	URLs
4	FastText + Random Forest	99.50	Correos electrónicos
5	HELPHED	99.43	Correos electrónicos
6	TB-DBN	99.40	URLs
7	LSTM-CNN + AVOA	99.37	Correos electrónicos
8	LSTM-GRU	98.89	URLs
9	DEPHIDES (CNN)	98.74	URLs
10	Optimización de modelos	98.27	Correos electrónicos y URLs
11	Modelo RNN-LSTM	97	Correos electrónicos
12	Plataforma de inteligencia de amenazas	97	Correos electrónicos y URLs
13	Redes neuronales profundas	97	Correos electrónicos
14	Gradient Boosting	97.2	URLs
15	Random Forest	97.1	URLs
16	Modelo DNN a nivel de codificación de caracteres	98.13	Correos electrónicos
17	BMEO con k-NN	98	URLs
18	PhiUSIIL (framework de similitud e incremental)	99.79	URLs
19	Lista blanca automatizada	96.17	Correos electrónicos
20	Transformer con mezcla de expertos	96	URLs
21	APuML con Random Forest	93.85	Correos electrónicos
22	SqueezeNet optimizado con FDHPO	93.05	URLs
23	K-Means (clustering mejorado)	89.2	URLs

El modelo que encabeza la tabla, con una precisión destacada del 99.97%, está diseñado específicamente para entornos financieros y gubernamentales, donde la seguridad es crítica. Este enfoque híbrido combina técnicas de aprendizaje automático y profundo en ejecución paralela (Random Forest, Naive Bayes, CNN y LSTM), cada una aportando un valor clave: Random Forest clasifica patrones sospechosos en grandes volúmenes de datos, Naive Bayes identifica probabilísticamente términos y patrones lingüísticos recurrentes en mensajes fraudulentos, CNN detecta patrones complejos en datos estructurados y visuales, y LSTM analiza dependencias temporales en datos secuenciales [23].

Cuando se busque la implementación, es recomendable seguir una estrategia escalonada. Primero, identificar los escenarios de mayor riesgo (por ejemplo, departamentos que gestionan datos financieros o personales) [2]. Luego, integrar modelos híbridos optimizados (como Random Forest + LSTM) en sistemas de correo electrónico y navegación mediante APIs o plugins específicos [21]. A nivel formativo, las organizaciones deben diseñar campañas periódicas de concienciación, incluir simulaciones de ataques de phishing y fomentar una cultura de alerta digital. No solo se debe priorizar desplegar tecnología avanzada, sino también fortalecer el eslabón humano, que sigue siendo clave en la cadena de ciberseguridad [25, 27, 28].

IV. DISCUSIÓN

En esta revisión, las técnicas de Machine Learning y Deep Learning, integradas en modelos híbridos en ejecución paralela (Random Forest, Naive Bayes, CNN y LSTM), obtuvieron una precisión casi perfecta del 99.97% [23]. Estos modelos, que combinan lo mejor de ambos enfoques, se destacan como soluciones altamente eficaces para entornos financieros y gubernamentales [10, 15]. Su capacidad para procesar grandes volúmenes de datos y analizar patrones complejos reduce significativamente el riesgo de ataques sofisticados. En comparación con los métodos tradicionales, como las listas negras, estas técnicas no solo incrementan la precisión, sino que también minimizan las tasas de falsos positivos al adaptarse dinámicamente a nuevos patrones de phishing [9, 12]. Según [9], las listas negras son ineficaces frente a amenazas desconocidas, mientras que los modelos híbridos en ejecución paralela ofrecen una capacidad superior para manejar amenazas emergentes.

Al mismo tiempo, diversos estudios también resaltan que la educación en ciberseguridad es clave para prevenir ataques de phishing. Los usuarios jóvenes, especialmente aquellos entre 18 y 25 años, son particularmente vulnerables debido a su menor conocimiento en seguridad digital y su frecuente interacción en redes sociales, donde los atacantes explotan la confianza mediante tácticas de ingeniería social [6, 7, 11]. Para este grupo demográfico, estudios destacan dos técnicas efectivas. PhishUSIIL, con una precisión del 99.79% [27], sobresale por incorporar aprendizaje incremental, lo que le permite actualizarse automáticamente frente a nuevas amenazas, superando así las limitaciones de los enfoques tradicionales. Además, ofrece una protección robusta y adaptada a dispositivos personales con recursos limitados. Por otro lado, APuML (Advanced Phishing URL Machine Learning), con una precisión del 93.85% [16], representa un avance prometedor en la detección de phishing en entornos móviles. Sin embargo, según [16], su precisión y escalabilidad aún no igualan a la de modelos más avanzados, lo que limita su implementación en escenarios más amplios.

Es necesario reconocer que estas técnicas emergentes aún enfrentan desafíos importantes que limitan su aplicación generalizada. Entre estas limitaciones destacan la alta demanda computacional que requieren, la necesidad de grandes volúmenes de datos etiquetados para su entrenamiento, y su menor rendimiento frente a ataques no conocidos o modificados respecto a los datos de entrenamiento [9], [16], [25], [30]. Asimismo, su implementación en dispositivos con recursos limitados o en contextos sin infraestructura tecnológica adecuada puede resultar inviable [18, 19, 20]. Se reafirma que la adopción efectiva de estas tecnologías dependerá tanto de avances técnicos como del contexto en el que se implementen [22, 23].

Por ultimo pero no menos importante, resulta pertinente considerar la viabilidad económica de aplicar estas soluciones en escenarios reales [12, 13]. Aunque los modelos híbridos ofrecen niveles de precisión sobresalientes, su implementación puede representar una inversión considerable en términos de infraestructura tecnológica, procesamiento de datos y personal especializado [22, 23, 27]. Esta realidad puede limitar su adopción en pequeñas y medianas organizaciones, especialmente en contextos con recursos limitados. Para estos casos, es recomendable explorar alternativas más accesibles [16, 20].

V. CONCLUSIÓN

En esta Revisión Sistemática de la Literatura, se han identificado las técnicas más prometedoras para la detección de ataques de phishing, cumpliendo así con la pregunta general planteada en la metodología PICO. Tal como se anticipaba, las tecnologías basadas en Machine Learning y Deep Learning destacan como las soluciones más prometedoras, superando las limitaciones de los métodos tradicionales al mejorar la precisión y reducir significativamente los falsos positivos. Aunque estas técnicas han demostrado ser efectivas por separado, la clave reside en su combinación, ya que cuando se implementan de manera conjunta y adaptativa, permiten mayor capacidad de respuesta ante amenazas en evolución, lo cual refuerza su relevancia en contextos dinámicos como el entorno digital en el que vivimos.

Sin embargo, también se ha evidenciado que la efectividad de estas tecnologías está condicionada por el contexto y las limitaciones actuales. Por ejemplo, las altas demandas computacionales y las restricciones en dispositivos móviles plantean desafíos significativos. Además, vale la pena mencionar que estas técnicas solo son plenamente efectivas si los usuarios tienen una conciencia adecuada sobre los riesgos del phishing y aplican buenas prácticas de ciberseguridad en su vida diaria. Los tiempos han cambiado, y en la actualidad, las amenazas de phishing no solo afectan a adultos o profesionales, sino que también impactan a niños y jóvenes, quienes desde edades tempranas están expuestos al uso de dispositivos conectados a internet. La proliferación de aplicaciones educativas, plataformas de entretenimiento y redes sociales destinadas a públicos jóvenes ha abierto nuevas puertas para los atacantes. Adicionalmente, se recomienda fomentar el desarrollo de interfaces accesibles para usuarios no técnicos, implementar sistemas de alerta temprana en navegadores y plataformas educativas, así como promover simulaciones interactivas de phishing como herramienta pedagógica. La combinación de modelos tecnológicamente avanzados con una ciudadanía digital preparada será clave para enfrentar de manera plena las amenazas de phishing en un entorno digital cada vez más complejo. De cara al futuro, las investigaciones deberán priorizar la escalabilidad y accesibilidad de estas técnicas, asegurando que cualquier usuario, independientemente de su nivel técnico o su edad, pueda beneficiarse de una protección efectiva frente a las amenazas digitales.

Solo mediante una estrategia integral que contemple avances tecnológicos, educación temprana y accesibilidad universal será posible enfrentar el phishing en todas sus formas y proteger a las generaciones actuales y futuras.

REFERENCIAS

[1] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," Electronics (Switzerland), vol. 12, no. 21, p. 4545, 2023.

[2] N. H. M. Ariffin, M. I. M. Iqbal, M. Yusoff, and N. A. M. Zulkefli, "A Study on the Best Classification Method for an Intelligent Phishing Website Detection System," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 48, no. 2, pp. 197–210, 2024.

[3] G. Varshney, R. Kumawat, V. Varadharajan, U. Tupakula, and C. Gupta, "Anti-phishing: A comprehensive perspective," Expert Systems with Applications, vol. 238, p. 122199, 2024.

[4] A. Nishikawa-Pacher, "Research Questions with PICO: A Universal Mnemonic," Publications, vol. 10, no. 3, p. 21, 2022.

[5] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, y The PRISMA Group, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," PLoS Med, vol. 6, no. 7, e1000097, 2009.

[6] H. J. Parker y S. V. Flowerday, "Contributing factors to increased susceptibility to social media phishing attacks," South African Journal of Information Management, vol. 22, no. 1, Art. no. a1176, Jun. 2020.

[7] E. Benavides-Astudillo, W. Fuertes-Díaz, y S. Sánchez-Gordon, "Un experimento para crear conciencia en las personas acerca de los ataques de Ingeniería Social," Revista Ciencia Unemi, vol. 13, núm. 32, pp. 27-40, 2020.

[8] L. Mayer Lux y G. Oliver Calderón, "El delito de fraude informático: Concepto y delimitación," Revista Chilena de Derecho y Tecnología, vol. 9, no. 1, pp. 151-184, 2020.

[9] Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, y A. K. Alazzawi, "AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites," IEEE Access, vol. 8, pp. 142532-142542, Aug. 2020.

[10] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, y K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," Telecommunication Systems, vol. 76, pp. 139–154, 2021.

[11] A. Hernández Dominguez y W. Baluja García, "Principales mecanismos para el enfrentamiento al phishing en las redes de datos," Revista Cubana de Ciencias Informáticas, vol. 15, esp., pp. 413-441, 2021.

[12] A. K. Dutta, "Detecting phishing websites using machine learning technique," PLoS ONE, vol. 16, no. 10, Art. no. e0258361, Oct. 2021.

[13] N. A. Azeez, S. Misra, I. A. Margaret, L. Fernandez-Sanz, y S. M. Abdulhamid, "Adopting automated whitelist approach for detecting phishing attacks," Computers & Security, vol. 108, Art. no. 102328, May 2021.

[14] A. M. Elmisery y M. Sertovic, "Modular Platform for Detecting and Classifying Phishing Websites Using Cyber Threat Intelligence," Electronic Communications of the EASST, vol. 80, 2021.

[15] S. M. Somesha y A. R. Pais, "Classification of Phishing Email Using Word Embedding and Machine Learning Techniques," Journal of Cyber Security and Mobility, vol. 11, no. 3, pp. 279–320, Apr. 2022.

[16] A. K. Jain, N. Debnath, y A. K. Jain, "APuML: An Efficient Approach to Detect Mobile Phishing Webpages using Machine Learning," Wireless Personal Communications, vol. 125, pp. 3227-3248, 2022.

[17] M. Alshehri, A. Abugabah, A. Algarni, y S. Almotairi, "Character-level word encoding deep learning model for combating cyber threats in phishing URL detection," Computers & Electrical Engineering, vol. 100, Art. no. 107868, Mar. 2022.

[18] S. Minocha y B. Singh, "A novel phishing detection system using binary modified equilibrium optimizer for feature selection," Computers and Electrical Engineering, vol. 98, Art. no. 107689, Jan. 2022.

[19] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, y E. Abu Elsoud, "An intelligent cyber security phishing detection system using deep learning techniques," Cluster Computing, vol. 25, pp. 3819-3828, 2022.

[20] S. R. Abdul Samad, S. Balasubaramanian, A. S. Al-Kaabi, B. Sharma, S. Chowdhury, A. Mehbodniya, J. L. Webber, y A. Bostani, "Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection," Electronics, vol. 12, no. 7, Art. no. 1642, Mar. 2023.

[21] A. R. Omar, S. Taie, y M. E. Shaheen, "From Phishing Behavior Analysis and Feature Selection to Enhance Prediction Rate in Phishing Detection," International Journal of Advanced Computer Science and Applications, vol. 14, no. 5, pp. 1033–1044, 2023.

[22] P. Bountakas y C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," Journal of Network and Computer Applications, vol. 210, Art. no. 103545, 2023.

[23] N. Nagy, M. Aljabri, A. Shaahid, A. A. Albin, F. Alnasser, L. Almakramy, M. Alhadab, y S. Alfaddagh, "Phishing URLs Detection Using Sequential and Parallel ML Techniques: Comparative Analysis," Sensors, vol. 23, Art. no. 3467, 2023.

[24] Y. Wang, W. Ma, H. Xu, Y. Liu, y P. Yin, "A Lightweight Multi-View Learning Approach for Phishing Attack Detection Using Transformer with Mixture of Experts," Applied Sciences, vol. 13, Art. no. 7429, 2023.

[25] A. Ejaz, A. N. Mian, y S. Manzoor, "Life-long phishing attack detection using continual learning," Scientific Reports, vol. 13, Art. no. 11488, 2023.

[26] K. Omari, "Comparative Study of Machine Learning Algorithms for Phishing Website Detection," International Journal of Advanced Computer Science and Applications, vol. 14, no. 9, pp. 417-425, 2023.

[27] A. Prasad y S. Chandra, "PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning," Computers & Security, vol. 136, Art. no. 103545, 2024.

[28] A. B. Majgave y N. L. Gavankar, "Automatic phishing website detection and prevention model using transformer deep belief network," Computers & Security, vol. 147, Art. no. 104071, Aug. 2024.

[29] O. K. Sahingoz, E. Buber, y E. Kugu, "DEPHIDES: Deep Learning Based Phishing Detection System," IEEE Access, vol. 12, pp. 8052-8068, Jan. 2024.

[30] N. Kamble y N. Mishra, "Hybrid optimization enabled squeeze net for phishing attack detection," Computers & Security, vol. 144, Art. no. 103901, May 2024.

[31] A. Al-Sabbagh, K. Hamze, S. Khan, y M. Elkhodr, "An Enhanced K-Means Clustering Algorithm for Phishing Attack Detections," Electronics, vol. 13, Art. no. 3677, Sep. 2024.

[32] Z. Alkhalil, C. Hewage, L. Nawaf, y I. Khan, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," Frontiers in Computer Science, vol. 3, Art. no. 563060, Mar. 2021.

[33] M. A. Elberri, Ü. Tokeşer, J. Rahebi, y J. M. Lopez-Guede, "A cyber defense system against phishing attacks with deep learning game theory and LSTM-CNN with African vulture optimization algorithm (AVOA)," International Journal of Information Security, vol. 23, pp. 2583–2606, May 2024.

[34] N. Subhashini, A. Banerjee, A. Kumar, S. Muthulakshmi, y S. Revathi, "Deep learning based phishing website detection," TELKOMNIKA Telecommunication Computing Electronics and Control, vol. 22, no. 1, pp. 113-121, Feb. 2024.