Artificial Intelligence Review (2025) 58:50
https://doi.org/10.1007/s10462-024-11055-z

Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection

S. Kavya[1] · **D. Sumathi**[1]

Accepted: 29 November 2024 / Published online: 20 December 2024
© The Author(s) 2024

Abstract
The escalating threat of phishing attacks poses significant challenges to cybersecurity, necessitating innovative approaches for detection and mitigation. This paper addresses this need by presenting a comprehensive review of state-of-the-art methodologies for phishing detection, spanning traditional machine learning techniques to cutting-edge deep learning frameworks. The review encompasses a diverse range of methods, including list-based approaches, machine learning algorithms, graph-based analysis, deep learning models, network embedding techniques, and generative adversarial networks (GANs). Each method is meticulously scrutinized, highlighting its rationale, advantages, and empirical results. For instance, deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrate superior detection performance, leveraging their ability to extract complex patterns from phishing data. Ensemble learning techniques and GANs offer additional benefits by enhancing detection accuracy and resilience against adversarial attacks. The impact of this review extends beyond academic discourse, informing practitioners and policymakers about the evolving landscape of phishing detection. By elucidating the strengths and limitations of existing methods, this paper guides the development of more robust and effective cybersecurity solutions. Moreover, the insights gleaned from this review lay the groundwork for future research endeavors, such as integrating contextual information, user behavior analysis, and explainable AI techniques into phishing detection systems. Ultimately, this work contributes to the collective effort to fortify digital defenses against sophisticated phishing threats, safeguarding the integrity of online ecosystems.

Keywords Phishing · Detection · Machine learning · Deep learning · Cybersecurity

✉ D. Sumathi
dsumathi@vit.ac.in

[1] School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, India

1 Introduction

In today's digital landscape, the proliferation of phishing attacks poses a significant threat to cybersecurity. Protecting sensitive data from malicious actors requires robust and effective detection mechanisms. With the evolution of cyber threats, traditional methods have proven inadequate in combating sophisticated phishing techniques. Hence, there is an urgent need for innovative approaches to enhance phishing detection.

This paper presents a comprehensive review of state-of-the-art methodologies in phishing detection. By analyzing a diverse range of techniques, including machine learning, deep learning, graph-based, and ensemble methods, this study aims to provide insights into the strengths and limitations of existing approaches. Each method is meticulously examined, considering its underlying principles, detection accuracy, computational efficiency, and susceptibility to adversarial attacks.

The proposed review model evaluates prominent methodologies such as list-based detection, machine learning classifiers, deep neural networks, and graph-based algorithms. By elucidating the rationale behind each approach and showcasing their advantages, this paper offers a holistic understanding of the current landscape of phishing detection. Furthermore, empirical results highlighting the detection accuracy, false positive rates, and specific findings of each method contribute to a comprehensive comparative analysis.

By synthesizing findings from diverse research endeavors, this review not only sheds light on the efficacy of individual techniques but also identifies overarching trends and challenges in phishing detection. Moreover, it underscores the importance of continuous innovation and collaboration in the field of cybersecurity to stay ahead of evolving threats.

In essence, this paper serves as a valuable resource for researchers, practitioners, and policymakers striving to develop robust defense mechanisms against phishing attacks. By elucidating the strengths and limitations of existing methodologies, it lays the groundwork for future advancements in cybersecurity, ultimately bolstering the resilience of digital ecosystems against malicious exploitation.

1.1 Motivation and contribution

The relentless evolution of cyber threats, particularly in the realm of phishing attacks, presents a pressing challenge to cybersecurity. With adversaries employing increasingly sophisticated tactics to deceive users and compromise sensitive information, there is an urgent need for innovative approaches to enhance detection and mitigation strategies. Recognizing this critical need, this paper embarks on a comprehensive exploration of state-of-the-art methodologies in phishing detection, as shown in Fig. 1.

The primary motivation behind this endeavor lies in addressing the inadequacies of traditional defense mechanisms against the growing threat landscape. Conventional approaches often fall short in effectively identifying and thwarting complex phishing schemes, leaving organizations vulnerable to data breaches and financial losses. By delving into the intricacies of modern phishing techniques and the corresponding detection methodologies, this study aims to bridge this gap and empower cybersecurity professionals with actionable insights to combat emerging threats.

The contribution of this paper lies in its meticulous analysis and synthesis of a diverse array of phishing detection techniques. By rigorously evaluating methodologies ranging

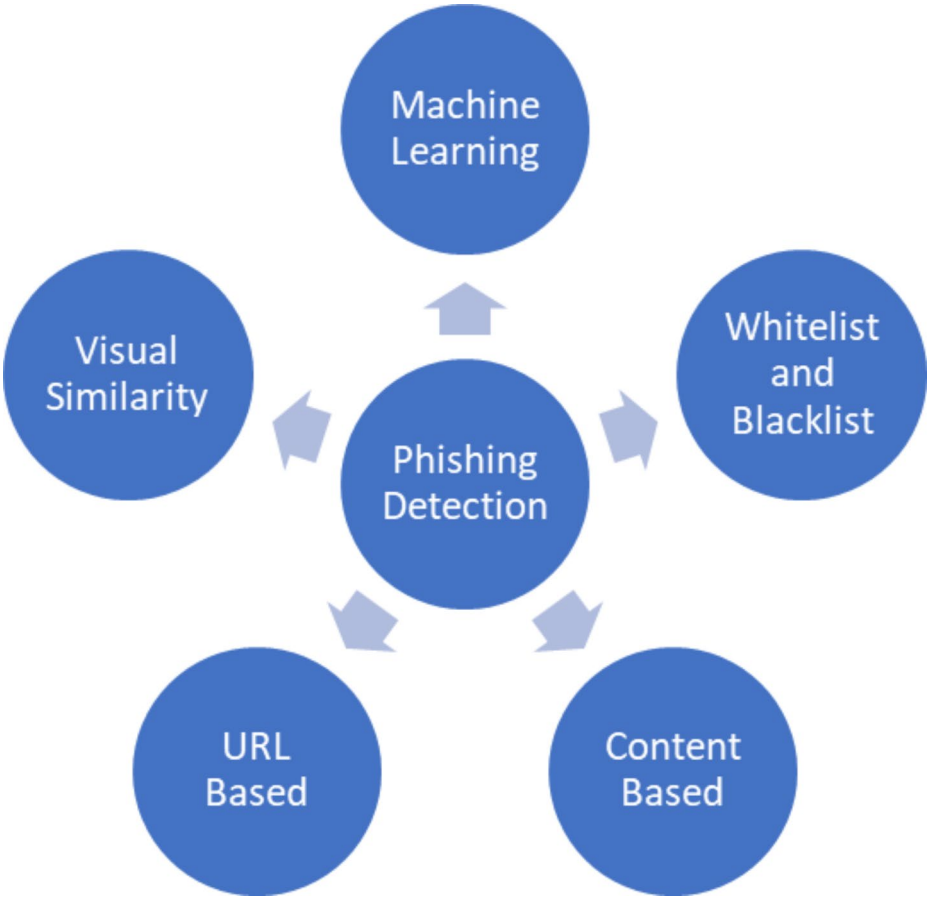

Fig. 1 Methodologies in phishing detection

from machine learning classifiers to deep neural networks and graph-based algorithms, this research consolidates existing knowledge and sheds light on the strengths and limitations of each approach. Furthermore, by providing empirical evidence of detection accuracy, false positive rates, and susceptibility to adversarial attacks, this study offers a nuanced understanding of the efficacy of different detection mechanisms.

Moreover, this paper goes beyond mere evaluation and comparison by identifying overarching trends and challenges in phishing detection. By elucidating the underlying principles and rationales behind each methodology, it equips researchers, practitioners, and policymakers with valuable insights to inform future advancements in cybersecurity. Ultimately, the contribution of this work lies in its potential to catalyze innovation and collaboration in the field, fostering the development of more resilient and adaptive defense mechanisms against phishing attacks.

1.2 Comprehensive review of existing methods

Any significant review of the related literature related to phishing detection methodologies has to take an existing stock and engage with it in a critical manner, to identify for both strengths and weaknesses of earlier work but more importantly to raise awareness to deficits in current knowledge areas and opportunities for the future advancement. The review will now focus on the other strategies of phishing detection, which include list-based to similarity-based techniques, machine learning, and deep learning methods, putting into perspective some emergent topics like adversarial robustness, diversity of datasets, and contextual features.

1. Classical approaches: list-based and similarity-based techniques

   Classical techniques used for phishing detection are based on list-based techniques, which actually used some predefined lists of known phishing URLs or domains. These methods, although highly very effective for known threats, were terribly lagging in adaptations and scalability. Even in research studies like Zieni et al. (2023) reported about the very challenges of list-based methods: their failure in detecting the phishing that has never seen before and the need to update them very very frequently; that is resource-consuming and likely to fail in cases where the phisher uses any evasion methods [1]. This therefore brings the researchers to the direction of looking at more dynamic detection mechanisms.
   Similarity-based methods developed as an alternative approach, that relies on content similarity to known phishing sites. Such approaches make either use of network embedding or content analysis to identify structural and transactional patterns in phishing sites [2]. Luo et al. (2024) demonstrated the efficiency of network embedding algorithms in regards to Ethereum on phishing account detection, but these algorithms still have limitations within the context of the applicability of other blockchain networks as well as the phishing attacks. Wen et al. (2023) presented a critical analysis of robustness, establishing that even slight perturbations inflicted from adversarial attacks could easily evade detection systems-wherein they uncover the vulnerability of detection mechanisms based on graph.

2. Machine learning approaches: a paradigm shift to data-driven models

   Machine learning brought in much more adaptive, scalable approaches for phishing detection. Traditional machine learning algorithms, such as random forests and support vector machines, came into play for the interpretation of features extracted from samples of URLs, web content, and metadata. Sahingoz et al. introduced DEPHIDES, deep learning phishing detector which outperformed the previous models on larger sets of data using better techniques for feature extraction [3]. Problems still remained in their over-reliance on hand-engineered features that made models not very flexible and adaptive to changing environments.
   At the same time, the problem of feature extraction began to be solved with the introduction of feature-free approaches. Purwanto et al. (2022) proposed the tool PhishSim that relies on normalized compression distance (NCD) for identifying phishing website

pages without manual feature extraction. The technique showed better flexibility and generalizability compared to others [4]. However, the technique remained to be tested on higher diverse datasets so that it may be tested in real-world scenarios.

3. Deep learning and generative models: the advancement on detection

Recently, deep learning models have emerged of great interest with the automatic ability to learn hierarchical representations of data. They stand distinguished among other models applied for the task of phishing detection. Al-Ahmadi et al. (2022) reported CNNs and RNNs, as the accuracy of these two models is far better in URL-based phishing detection as the models learned complex, non-linear patterns that the traditional machine learning models simply could not.
Recently, GANs have been introduced to a new paradigm in phishing detection since it creates synthetic phishing samples to be used to enhance the training dataset. A recently proposed phishing detection model is PDGAN by Al-Ahmadi et al., which achieved high accuracy in detection using only URL information, which makes it possible to apply GANs in scenarios with limited or unbalanced datasets [5]. However, on the flip side, while GANs help enhance model performance, they introduce new challenges, such as overfitting to synthetic data and requiring careful tuning of the generative and discriminative models to avoid adversarial failure modes.

4. Robustness and adversarial attacks: overcoming model vulnerabilities

The major challenge extracted from the literature is that phishing detection systems are vulnerable to adversarial attacks. Wen et al. (2023) discussed how the adversarial hiding techniques may be used for avoiding phishing detection systems and especially in blockchain networks like Ethereum. Based on their study, they concluded that detection frameworks must be tested with realistic attack conditions so that the frameworks are robust enough for real-world applications [6]. Similarly, researchers like Pillai et al. (2024) have concentrated efforts on defense mechanisms for evasion attacks, emphasizing how phishing detection models should perform well in benign environments and withstand deliberate evasion attempts [7].
These adversarial vulnerabilities have been proposed with solution techniques through ensemble learning methods. Ensemble methods, as represented by Kalabarige et al., such as a stacking of multiple deep learning models with gated recurrent units, have produced augmented detection accuracy and more robust performance by aggregating the strengths of several algorithms [8]. However, these are computationally expensive and require careful selection of base models to avoid either model redundancy or overfitting.

5. Diversity of dataset and contextual feature integration

Among the common themes in the literature is the need for more diversified and representative datasets in phishing detection model evaluation. Many studies, like a more recent paper by Castaño et al. (2023) and a study recently carried out by Kabla et al. in 2022, support this notion: their current datasets have various limitations, such as

unable to deal with all phishing attack vectors, including mobile phishing or multi-stage attacks [9, 10]. For example, PhiKitA is a development of datasets which contain phishing kits and related websites, and it was one step forward toward offering more complete benchmarks for the evaluation of detection models [10]. In this case, to create datasets with wider coverage of various phishing tactics and target environments like social media and mobile applications it would be quite challenging.
Besides dataset diversity, some other recent trends in research try to bring contextual information into phishing detection systems. For instance, the role of gain and loss persuasion cues in the detection of phishing emails has been discussed by Salloum et al. [11]. These approaches have a good prospect of countering the deficiencies of traditional feature-based models by incorporating more sensitive, context-aware mechanisms in detection.

6. Future scopes and research gaps

Although based on phishing detection, much research has been achieved, and in reality, there are still numerous gaps that are to be filled. The direction for future work should be towards incorporation of XAI techniques in order to improve interpretability for such phishing detection models. It is only when the complexity of the phishing detection systems rises, mainly when deep learning and ensemble methods have been utilized in the phishing detection system, that a need for transparency and interpretability seems to surface within the system. Future work will be conducted to verify the effectiveness of XAI techniques in further achieving a more transparent and trustworthy phishing detection framework for predictions, which may then be used in places with higher stakes, like financial services or government networks.
Another encouraging direction of future work is to apply federated learning and differential privacy to phishing detection. These may enable more practical privacy-preserving models to be generated, but which can be run directly over distributed, decentralized data sets without leaking user data samples. This is even more pressing in the case of phishing detection on social media platforms or in mobile applications where privacy matters the most.
The scope of the literature on phishing detection will be very vast, ranging from the very traditional list-based to cutting-edge deep learning models and generative adversarial networks. Although much has been achieved, it would appear that much is still yet to be accomplished: gaps in this regard include diversity in the dataset, adversarial robustness, and interpretability of models. Future work will focus on filling in these gaps with new detection techniques, increasing the representativeness of datasets, and promoting transparent and privacy-preserving models. Researchers can also contribute to better and more resilient phishing detection systems through constant innovation and critical engagement with the body of knowledge that already exists in the field.

1.3 Categorical analysis

While phishing detection literature is significantly heterogeneous in its approaches toward mitigating phishing attacks, generally the methodologies fall into three primary categories: list-based, similarity-based, and machine learning-based approaches. Each comes with its

strength and weakness, thus the dynamism in the nature of phishing attacks and the requisite adaptive defense mechanisms.

1. List-based approaches

This method is more common due to its simplicity and speed in the identification of phishing sites. In this list-based approach, detection is based on predefined lists of known phishing websites, domains, or URLs against incoming traffic.

- Mechanism: These blacklists or whitelists contain predefined phishing websites or emails; they are referenced as flags. The list usually is prepared from user reports, third-party databases, or previous identified phishing attacks.
- Known phishing site identification is very rapid and efficient.
- The only limitation is that because attackers frequently change the URLs and domains to avoid blacklists, this approach is prone to outdated information. This method fails in the detection of new or unknown phishing attack .
- Key studies: List-based methods have been totally researched and found that they are susceptible to continuous phishing attack advancement as attackers advance to new tactics often to evade detection [1].

2. Similarity-based approaches

Similarity-based approaches fundamentally work by checking every arriving website or email against those known phishing contents, usually estimating phishing probabilities with the help of network embeddings or other similarity measures.

- Mechanism: This scores the similarity of a website or email to known phishing sites using features such as URL structure, content layout, or transactional patterns. Network embedding techniques are very relevant for detecting phishing behavior based on transaction links between accounts and are similar to the case in Ethereum [12].
- Advantages: The approach identifies phishing clones or attacks that were similar to known phishing websites. These similarity-based approaches are more convenient for updating changing phishing tactics rather than the list-based ones.
- Disadvantages: The similarity-based methods might not detect completely new phishing strategies that are not alike the previously observed patterns.
- Key studies: The phishing account detection in the blockchain network has been found pretty successful using similarity-based methods; such network embedding-based models increase the detection accuracies concerning the Ethereum-based phishing scams [6, 9].

3. Machine learning methods

Machine learning category: extremely broad as it encompasses nearly all the traditional ML models, the latest advanced DL and GAN models. The merit of applying these models is that they can handle huge-sized datasets and can learn adapting to new threats accord-

ing to changing rules explicitly. So, they are particularly helpful in dynamic phishing environments.

a. Traditional machine learning methods

- Mechanism: General traditional ML models, such as Logistic Regression, Random Forest, Support Vector Machine (SVM), etc., are adapted for phishing content classification from features like length of URL, entropy, and HTML tags. These models drastically require extensive feature engineering followed by the training over labeled datasets.
- Pros: Suitable for structured data; hard detection in static environments.
- Limitation: These types of models suffer from a high false positive rate and tend to lag behind the dynamic phishing attacks in which the feature extraction is ambiguous or changing without easy adaptation to the shifting phishing strategies [13].
- Key Studies: Logistic regression and traditional ML models with TF-IDF feature extraction technique produce the maximum accuracy for detecting URL-based phishing attack [13].

b. Deep learning techniques

- Mechanism: The models are CNNs and RNNs. It has proven that they excel at performing phishing detection compared to the traditional methods as they use automatically learned complex patterns from large datasets. Moreover, these models also have fine feature extraction ability where the selection procedure of feature features is removed during the feature engineering process.
- Advantages: Deep learning models, in particular, CNNs have shown a very accurate performance, for instance, CNNs had already achieved 98.74% in phishing detection [3]. Also, they are very scalable to hundreds of phishing scenarios even in websites, emails, and mobile applications.
- Costs: The models will require many computational resources, and on top of this, they will be vulnerable to adversarial attacks where the attackers prepare phishing examples that might deceive the model.
- Key Studies: It was succeeded in applying CNNs and LSTM networks in the detection of phishing attacks. CNNs demonstrated superior performance in URL-based detection, while LSTM models had great success in distinguishing phishing emails [3, 14].

c. Generative Adversarial Networks (GANs)

- Mechanism: It uses a generator-discriminator model-for example, similar to the PDGAN along with advancement in phishing detection based on GANs. It manufactures fake phishing data, and accordingly, the discriminator tries to classify as of whether the given data is actual or synthetic, making this model more robust.
- Advantages of GANs: It is a very effective approach to generating samples of diversified phishing and has a high accuracy level of detection, especially in an environment where available training samples of data are minimal. Also, this approach

provides an increase in resistance to adversarial attacks when it consists of continuous exposure to new techniques of phishing.

- Disadvantages of GANs: These become very resource-extensive and synthetic generation as opposed to realistic data may sometimes be more common, at least in the context of the real-world phishing scenario.
- Key Studies: PDGAN was said to report 97.58% accuracy with the incorporation of URL-based information, which came grossly outperforming traditional approaches in detecting zero-day phishing attacks that, up to date, are undetectable with the same approaches [5].

4. Ensemble learning techniques

Ensemble learning combines the power of many models with a better potential to determine the attack as phishing with a low false positive rate. Techniques like ResNeXt and GRUs had been applied in studies for the improvement of the model's accuracy in detection by using the advantages provided by various models.

- Mechanism: Ensemble approaches consist of combining the outputs from multiple classifiers with an aim to make a more accurate and robust prediction. Ensemble techniques aggregate the predictions of various machine learning models and deep learning models in the context of phishing detection.
- Advantages: such techniques can boost up detection accuracy, handle data imbalance in context, and they tend to be more resistant to overfitting than individual models.
- Limitations: Training and deployment of multiple models are computationally intense, especially when it involves real-time applications.
- Key studies: Ensemble models increase phishing detection accuracy by several factors and reduce false positives in both email and web-based phishing cases [15].

5. NLP techniques

NLP techniques are extensively used to perform phishing detection, particularly in email content. The models that rely on NLP pick up on phishing emails based on their textual patterns and semantic structures that adopt social engineering and other deception techniques.

- Mechanism: NLP-based models identify the undesirable words, phishing cues, and other markers that trigger a phishing attempt in the textual content of the emails or websites.
- Advantages: NLP is highly effective against phishing attacks based on textual manipulation, including such attacks via emails and false login pages.
- Limitations: NLP models will fail without a doubt in languages or formats which they are not trained directly for, and most of them are also sensitive to language variations or very subtle phishing techniques.
- Key Studies: Methods of NLP especially LSTM proved to be quite fine to be used in email content analysis for the detection of phishing attempts. One study has shown how LSTM achieved a 95% accuracy level in the detection process of phishing emails [11] [14].

2 Literature review

The literature on phishing detection encompasses a diverse array of approaches aimed at mitigating the pervasive threat posed by phishing attacks. This section provides a comprehensive review of recent advancements in phishing detection methodologies, drawing from a wide range of studies. Phishing, a long-standing security threat [1], continues to evolve, employing increasingly sophisticated tactics to deceive individuals and compromise brands [2]. Detection of phishing websites is crucial in combatting this threat, prompting the exploration of various detection approaches. The literature categorizes these approaches into three main categories: list-based, similarity-based, and machine learning-based [1]. List-based approaches rely on predefined lists of known phishing websites and characteristics, enabling quick identification but susceptible to outdated information and new phishing tactics [1]. In contrast, similarity-based methods assess website similarity to known phishing sites, often leveraging network embeddings to capture transactional patterns [6]. Machine learning-based techniques, prominently featuring deep learning algorithms, have gained traction for their ability to process large datasets and adapt to evolving threats [3].

Studies have explored the application of machine learning to diverse domains, including blockchain platforms like Ethereum, where phishing poses a significant threat [12]. Novel methodologies, such as network embedding-based models, demonstrate promising results in detecting phishing accounts within the Ethereum ecosystem [9]. Moreover, researchers have delved into the robustness of phishing detection frameworks, particularly in the face of adversarial attacks aimed at concealing phishing behaviors [6]. This highlights the importance of evaluating detection models under realistic conditions to ensure effectiveness. Deep learning approaches, including generative adversarial networks (GANs), have emerged as a promising avenue for phishing detection, with models like PDGAN exhibiting high accuracy using only URL information [5] sets. Furthermore, efforts to enhance dataset availability and model performance have led to the creation of novel resources such as PhiKitA, a dataset containing phishing kits and associated websites [10]. Such initiatives contribute to the advancement of phishing detection research by providing standardized benchmarks for evaluation.

While machine learning models have shown remarkable progress, challenges remain, including the detection of phishing emails and mobile phishing attacks [4, 11]. Techniques like Natural Language Processing (NLP) and deep learning hold potential for addressing these challenges by analyzing email content and mobile app behavior. Recent innovations, such as feature-free methods based on the Normalized Compression Distance (NCD), demonstrate the effectiveness of alternative approaches in detecting phishing websites without relying on feature extraction [4]. Moreover, the emergence of ensemble techniques, combining deep learning architectures like ResNeXt with gated recurrent units (GRUs), showcases the potential for improving detection accuracy and efficiency [15]. However, phishing attacks continue to evolve, necessitating ongoing research into novel detection strategies. Future studies may explore collaborative approaches, such as detecting phishing gangs in blockchain ecosystems [16], to uncover coordinated attacks and bolster cybersecurity defenses, this can be summarized from Table 1 as follows,

Phishing, a prevalent cybercrime worldwide, poses a significant threat to cybersecurity by deceiving users into divulging sensitive information or downloading malware [20]. With the evolution of artificial intelligence (AI), researchers increasingly employ machine

Table 1 Empirical review of existing methods

Reference	Method Used	Findings	Results	Limitations
[1]	Literature Review	Comprehensive overview of phishing detection	LOW	LOW
[2]	Network Embedding, Machine Learning	Phishing account detection on Ethereum	Achieved effective detection performance	Limited to Ethereum platform, may not generalize to other blockchain systems
[6]	Graph-based Phishing Detection	Robustness of phishing detection frameworks	Detection model is fragile under adversarial attacks	Evaluation limited to proposed framework and datasets used
[3]	Deep Learning, Phishing Detection	Phishing detection system based on deep learning	Convolutional neural networks achieved highest performance	Limited evaluation on deep learning algorithms, dataset may not cover all possible phishing scenarios
[12]	Network Embedding, Machine Learning	Phishing detection on Ethereum	Proposed method effectively detects phishing scams	Limited to Ethereum platform, may not generalize to other blockchain systems
[9]	Machine Learning	Ethereum Phishing Scam Detection (Eth-PSD)	Eth-PSD efficiently detects phishing scams	Limited evaluation on dataset, may not cover all possible phishing scenarios
[5]	Deep Learning, Generative Adversarial Network	Phishing detection model (PDGAN)	Achieved high detection accuracy	Limited to URL-based phishing detection, may not cover other types of phishing attacks
[10]	Dataset Creation, Phishing Detection	Creation of PhiKitA dataset	Achieved high accuracy in phishing detection	Limited to evaluation on a specific dataset, may not generalize to other datasets
[17]	Literature Review, Anti-Phishing Techniques	Review of cloaking techniques in phishing	Identified various cloaking mechanisms used by phishers	Limited to analysis of existing literature, may not cover all recent developments in phishing evasion techniques
[18]	Machine Learning, Evolutionary Optimization	Phishing detection model (MOE/RF)	Balanced detection accuracy and false detection rates	Limited to evaluation on specific datasets and may not generalize to other datasets
[19]	Deep Learning, Android Security	GUI-Squatting attack for phishing apps	Successfully generates phishing apps and bypasses existing defenses	Limited to Android platform and may not generalize to other mobile platforms or phishing scenarios
[11]	Natural Language Processing, Phishing Detection	NLP techniques for phishing email detection	Identified key areas in phishing email detection	Limited to analysis of existing literature, may not cover all recent developments in NLP techniques for phishing email detection
[4]	Compression Distance, Phishing Detection	Feature-free phishing detection method using NCD	Significantly outperforms previous methods	Limited to evaluation on specific datasets and may not generalize to other datasets
[15]	Deep Learning, Digital Forensics	ResNeXt and GRU model for phishing detection	Outperforms state-of-the-art algorithms	Limited evaluation on real phishing attack datasets, may not generalize to other datasets
[14]	Deep Learning, Email Data Analysis	LSTM-based phishing detection method	Achieved high detection accuracy	Limited to email phishing detection, may not cover other types of phishing attacks

Table 1 (continued)

Reference	Method Used	Findings	Results	Limitations
[16]	Transaction Behavior Analysis	Detection of Ethereum phishing gangs	Effectively detects potential risky accounts	Limited to Ethereum platform and may not generalize to other blockchain systems

learning (ML) and deep learning (DL) algorithms to combat phishing [20]. Ensemble ML algorithms emerge as top performers in detection accuracy and computational efficiency, especially when feature volumes dwindle, reflecting their suitability for real-time environments [20].

User susceptibility to phishing attacks has garnered considerable attention, necessitating comprehensive analyses to assess vulnerability, particularly among older populations [21]. Meta-analyses of previous studies reveal varied findings regarding age and gender effects on susceptibility, underscoring the need for a nuanced understanding of phishing susceptibility [21]. Notably, age and gender show significant relationships with susceptibility, while user training enhances detection abilities [21].

Machine learning and deep learning techniques present promising avenues for phishing website detection, but existing solutions exhibit high false-positive rates [13]. By considering URLs from login pages in both legitimate and phishing classes, recent studies challenge conventional approaches, highlighting the necessity of more representative datasets and real-time analysis [13]. Logistic Regression models, coupled with TF-IDF feature extraction, demonstrate high accuracy in detecting phishing URLs [13].

Phishing attackers leverage social engineering tactics to exploit human vulnerabilities, necessitating innovative detection methods [22]. Proposed deep learning frameworks, integrated into browser plug-ins, offer real-time phishing risk assessments by combining whitelist filtering, blacklist interception, and ML prediction strategies [22]. RNN-GRU models emerge as top performers in accuracy, signaling the feasibility of DL solutions [22].

Generative Adversarial Network (GAN)-based approaches aim to mitigate ML limitations arising from insufficient training data and susceptibility to adversarial attacks [23]. By synthesizing phishing and legitimate samples, GANs bolster classifier performance and resistance to attacks, showcasing promising results across various datasets [23].

Recent advancements in detection techniques include Tiny-Bert stacking models and novel PhishDet frameworks, leveraging Graph Neural Networks [24, 25]. These models achieve high accuracy rates and demonstrate efficacy against evolving phishing tactics, offering hope in combating sophisticated attacks [24, 25].

Moreover, attributed ego-graph embedding frameworks for Ethereum transaction networks and stacked ensemble learning models further enhance phishing detection capabilities [26, 8]. These approaches leverage advanced feature extraction and ensemble learning techniques, yielding robust performance across diverse datasets [8, 26].

As phishing attacks continue to evolve, defense mechanisms must adapt accordingly [27]. Novel techniques, such as multilayered stacked ensemble learning and gray-box attacks, demonstrate resilience against evasion tactics while maintaining high detection rates [27, 28]. Furthermore, cryptography-based authentication schemes and character-aware language models showcase promising results in detecting social semantic attacks [29, 30]. This can be summarized from the following Table 2.

Table 2 Empirical review of existing methods

Reference	Method Used	Findings	Results	Limitations
[20]	Ensemble machine learning algorithms	Utilized ensemble machine learning algorithms for phishing website detection.	Achieved high detection accuracy and computational efficiency.	The study focuses primarily on ensemble machine learning methods, potentially limiting generalizability.
[21]	Meta-analysis and systematic review	Conducted a meta-analysis to determine susceptibility to phishing attacks among different subpopulations.	Found significant relationships between age and susceptibility, as well as gender differences.	Discrepancies exist among the findings of previous studies, highlighting potential inconsistencies.
[13]	Machine learning and deep learning	Compared machine learning and deep learning techniques for phishing website detection using URL analysis.	Demonstrated high accuracy with logistic regression and TF-IDF feature extraction.	The study focuses solely on URL analysis, potentially overlooking other important features.
[22]	Deep learning framework	Proposed a deep learning framework for real-time phishing website detection using a browser plug-in.	Achieved high accuracy with the RNN-GRU model and multiple strategies for real-time prediction.	Reliance on real-time prediction may pose challenges in terms of computational resources.
[23]	Generative Adversarial Network (GAN)	Utilized GAN-based approaches to synthesize phishing and legitimate samples for improving classifier performance and resilience to adversarial attacks.	Showed the effectiveness of AAE and WGAN in generating synthetic data for training classifiers.	The study relies heavily on synthetic data generation, which may not fully represent real-world scenarios.
[24]	Tiny-Bert stacking	Developed a phishing website detection model based on Tiny-Bert stacking, achieving high accuracy and stability compared to state-of-the-art methods.	Achieved high accuracy and stability with the proposed Tiny-Bert stacking model.	Limited discussion on potential biases or imbalances in the dataset used for evaluation.
[25]	Long-term Recurrent Convolutional Net	Proposed PhishDet, a detection method using Long-term Recurrent Convolutional Network and Graph Convolutional Network, achieving high accuracy and low false-negative rate.	Recorded high detection accuracy and effectiveness against zero-day attacks.	Requires periodic retraining to maintain performance, which could be resource-intensive.
[26]	Attributed ego-graph embedding	Introduced an attributed ego-graph embedding framework for distinguishing phishing accounts on Ethereum, achieving effective performance in class-imbalanced detection.	Achieved effective performance in class-imbalanced phishing detection on Ethereum.	Reliance on Ethereum transaction attributes may limit applicability to other platforms.

Table 2 (continued)

Reference	Method Used	Findings	Results	Limitations
[8]	Stacked ensemble learning technique	Proposed a multilayered stacked ensemble learning technique for phishing website detection, achieving high accuracy across different datasets.	Achieved significant performance improvements compared to baseline models.	The study evaluates the proposed technique on limited datasets, potentially limiting generalizability.
[27]	Gain and loss persuasion cues	Investigated the effectiveness of gain and loss persuasion cues in phishing email detection using machine learning models, showing significant improvements over baseline models.	Demonstrated reliable methods for phishing email detection using persuasion cues.	Limited discussion on the scalability and generalizability of the proposed persuasion cue models.
[29]	Cryptography-based authentication	Proposed a cryptography-based authentication scheme for mitigating phishing attacks in online social networks, achieving resistance to phishing and related attacks.	Outperformed existing schemes in terms of security and functionality.	The study focuses specifically on online social networks, potentially limiting applicability to other domains.
[31]	Classifier algorithms	Developed a method for detecting phishing websites using classifier algorithms based on Internet URL and domain names, achieving high accuracy rates.	Detected phishing websites with high accuracy using classifier algorithms.	Limited discussion on the generalizability of the proposed method to different types of phishing attacks.
[7]	Adversarial attacks defense mechanism	Introduced a defense mechanism against evasion attacks on phishing website classifiers, marking a significant contribution to the field.	Proposed a novel defense mechanism against evasion attacks on phishing website classifiers.	The study focuses primarily on defense mechanisms, potentially overlooking other aspects of phishing detection.
[28]	Boosting-based stacked ensemble	Proposed a boosting-based stacked ensemble learning model with hybrid feature selection for phishing website detection, achieving high accuracy across different datasets.	Outperformed existing models in terms of accuracy and performance.	Limited discussion on the interpretability of the hybrid feature selection technique.
[32]	Feature selection and stacked ensemble	Developed a stacked ensemble learning model with hybrid feature selection for phishing website detection, achieving high accuracy and outperforming existing models.	Achieved high accuracy across different datasets and outperformed existing models.	Limited discussion on potential biases or imbalances in the dataset used for evaluation.
[30]	Gray-Box attacks and protective chain	Proposed Gray-Box attacks and a protective operation chain algorithm to defend against attacks on phishing detectors, demonstrating robustness and performance improvements.	Demonstrated robustness against Gray-Box attacks and improved performance compared to past work.	The study primarily focuses on defense mechanisms and may overlook potential vulnerabilities in real-world scenarios.

Phishing attacks have emerged as a significant threat in the cybersecurity landscape over the past decade [33]. Despite various efforts to understand and mitigate these attacks, there remains a critical gap in comprehensive phishing reports, particularly in terms of classification techniques. The study by [33] emphasizes the importance of addressing this gap through systematic reviews focusing on classification techniques, datasets, performance evaluation, and phishing types to aid in more effective prevention strategies.

Intrusion Detection Systems (IDSs) play a crucial role in combating cyber-attacks, which continue to challenge data confidentiality and network security [34]. Recent advancements in IDS taxonomy and techniques, including machine learning (ML) and deep learning (DL) approaches, show promise in enhancing detection accuracy and mitigating false positives. These advancements, as highlighted by [34], underscore the need for continuous research to address evolving intrusion techniques and improve network security. The proliferation of multistage malware botnets poses a significant threat to Internet of Things (IoT) devices, necessitating adaptive Intrusion Detection Systems (IDSs) capable of adjusting to evolving threat patterns [35]. The introduction of MalBoT-DRL, a robust malware botnet detector utilizing deep reinforcement learning (RL), represents a novel approach to enhancing generalizability and resilience in detecting botnets across IoT environments [35]. In the financial sector, the advent of information technology has revolutionized transaction modes, but it has also introduced new challenges, including hidden frauds [36]. Traditional machine learning models struggle with the increasing volume of financial transaction data, leading to the proposal of innovative approaches such as TA-Struc2Vec for Internet financial fraud detection, as discussed by [36] for different scenarios. Phishing attacks remain a severe and prevalent cybercrime, exploiting email distortion and mock sites to obtain sensitive data [37]. Leveraging machine learning algorithms, researchers aim to develop effective defenses against phishing attacks, with proposed hybrid models showing promising results in preventing phishing URLs and protecting users [37].

The Ethereum network faces vulnerabilities from targeted attacks, necessitating efficient abnormal transaction detection methods [38]. The introduction of Abnormal Transactions Detection Using a Semi-Supervised Generative Adversarial Network (ATD-SGAN) showcases significant improvements in detecting abnormal transactions within the Ethereum network [38]. The proliferation of phishing websites underscores the importance of advanced detection methods capable of analyzing various modalities of website content [39]. The proposed multi-modal hierarchical attention model (MMHAM) by [39] addresses this challenge by jointly learning deep fraud cues from textual information, visual design, and URLs, improving phishing detection accuracy. Anomaly detection in file system accesses plays a crucial role in protecting sensitive data from malicious insiders and outsiders [40]. The proposed approach by [40] leverages fine-grained profiling of users' regular file access activities, combined with advanced anomaly detection techniques, to achieve high accuracy in detecting anomalous file system accesses while minimizing overheads. This can be summarized from the following Table 3,

In summary, the literature highlights the ongoing efforts to address cybersecurity challenges, including phishing attacks, malware detection, financial fraud, and anomaly detection. Advanced techniques such as machine learning, deep learning, and reinforcement learning show promise in enhancing detection accuracy and mitigating cyber threats across various domains. However, there remains a need for continuous research to stay ahead of

Table 3 Empirical review of existing phishing detection methods

Reference	Method Used	Findings	Results	Limitations
[33]	Systematic Review	Classified phishing attacks based on classification techniques, datasets, performance evaluation, and phishing types.	Expected to aid developers in preventing future phishing attacks effectively.	Results might not cover all possible phishing attack scenarios, and the effectiveness of preventive measures may vary depending on the sophistication of future attacks.
[34]	Review of IDS techniques	Explored intrusion detection techniques, challenges in combating evasion techniques, and advancements in ML and DL-based NIDS.	Highlighted the potential of ML and DL techniques in enhancing network security.	The effectiveness of IDS systems heavily relies on the accuracy of detection algorithms and the quality of training data, which may vary in real-world environments.
[35]	Development of MalBoT-DRL for malware detection	Developed a robust malware botnet detector using deep reinforcement learning.	Achieved exceptional detection rates in trace-driven experiments.	The performance of the proposed detector may depend on the representativeness of the training data and the generalizability of the model to different IoT environments.
[36]	Proposal of TA-Struc2Vec for fraud detection	Proposed a graph-learning algorithm for Internet financial fraud detection.	Improved the efficiency of fraud detection with better precision and AUC.	The effectiveness of the TA-Struc2Vec algorithm may be influenced by the quality and representativeness of the financial transaction data used for training.
[37]	Utilization of machine learning for phishing	Employed various machine learning models for phishing URL detection and prevention.	Achieved high accuracy and efficiency in defending against phishing attacks.	The performance of machine learning models in phishing detection may be affected by the diversity and complexity of phishing techniques employed by attackers.
[38]	Introduction of ATD-SGAN for abnormal detection	Presented a novel approach using semi-supervised generative adversarial networks for abnormal transaction detection in Ethereum.	Significantly enhanced the performance of existing IDSs.	The effectiveness of ATD-SGAN may depend on the quality and representativeness of the training data, as well as the ability to adapt to evolving attack patterns in the Ethereum network.
[41]	Review of DM algorithms for phishing detection	Conducted a systematic review of DM algorithms for detecting phishing attempts.	Most algorithms achieved high accuracies, but 100% accuracy was not universally attained.	The performance of phishing detection models may vary depending on the quality and diversity of the training data, as well as the generalizability of the algorithms to different types of phishing attacks.

Table 3 (continued)

Reference	Method Used	Findings	Results	Limitations
[42]	Investigation of contextual features for fraud	Investigated the impact of contextual features on the accuracy of online recruitment fraud detection.	Found that inclusion of contextual features improved detection performance.	The effectiveness of contextual features may vary depending on the relevance and availability of contextual information in different online recruitment scenarios.
[43]	Proposal of BDRNN for clickbait detection	Proposed a Blockchain-enabled deep recurrent neural network for clickbait detection.	Outperformed existing neural network models in detecting safe and malicious clickbait.	The performance of BDRNN may depend on the quality and diversity of training data and the ability to effectively capture the characteristics of safe and malicious clickbait content.
[44]	Development of QsecR for QR code security	Developed a secure and privacy-friendly QR code scanner for detecting malicious URLs.	Achieved high detection accuracy and privacy-friendliness.	The effectiveness of QsecR may depend on the comprehensiveness of the feature set and the ability to adapt to emerging QR code security threats.
[45]	Survey of computer vision methods in security	Explored the application of computer vision methods in network security for attack detection and prevention.	Identified potential applications and gaps in utilizing computer vision methods in security.	The applicability of computer vision methods in network security may be limited by the availability and quality of visual data and the computational resources required for analysis.
[39]	Proposal of MMHAM for phishing website detection	Introduced a multi-modal hierarchical attention model for phishing website detection.	Improved phishing detection by considering multiple modalities of website content.	The performance of MMHAM may vary depending on the representativeness and diversity of the training data and the ability to effectively integrate information from different modalities.
[46]	Development of ensemble ML model for phishing	Proposed a robust ensemble machine learning model for detecting malicious URLs.	Achieved high prediction accuracy with low error rates.	The performance of ensemble models may depend on the selection and combination of individual classifiers, as well as the quality and representativeness of the training data.
[47]	Utilization of DL strategies for cyberattack	Proposed an approach using Deep Learning strategies for cyberattack detection and categorization.	Improved detection accuracy of network attacks using DL techniques.	The effectiveness of DL strategies may depend on the availability of labeled training data and the ability to generalize to unseen cyberattack patterns.

Table 3 (continued)

Reference	Method Used	Findings	Results	Limitations
[48]	Proposal of TEDA-CDD for concept drift detection	Presented a concept drift detection method based on Typicality and Eccentricity Data Analytics.	Demonstrated comparable accuracy with existing algorithms in detecting concept drift.	The performance of TEDA-CDD may vary depending on the complexity and dynamics of the data stream and the ability to adapt to changing concepts in real-world scenarios.
[49]	Proposal of CPRF for network attack detection	Introduced a novel approach called Class Probability Random Forest for network attack detection.	Achieved high detection performance with a comprehensive feature set.	The effectiveness of CPRF may depend on the quality and diversity of the feature set and the ability to accurately model network attack behaviors.
[50]	Hybrid approach for web security using XGBoost	Proposed a hybrid approach using XGBoost optimized by an improved firefly algorithm for web security.	Outperformed other approaches in detecting phishing websites.	The performance of the hybrid approach may vary depending on the effectiveness of the optimization algorithm and the ability to generalize to different types of phishing attacks.
[40]	Anomaly detection in file system accesses	Developed an approach for anomaly detection in file system accesses using fine-grained user profiles.	Achieved high accuracy in detecting anomalous file system accesses.	The effectiveness of the approach may depend on the quality and representativeness of the training data and the ability to accurately model user file access behaviors.

evolving cyber threats and safeguard critical systems and data samples. Next, we perform a comparative analysis of these models for different use case scenarios.

2.1 Comprehensive analysis of phishing detection methods

This section offers a thorough analysis of various phishing detection methods presented in the provided table. The methods range from traditional machine learning approaches to cutting-edge deep learning techniques, each addressing different aspects of phishing detection with unique strengths and limitations. The Tables 1 and 2, and 3 showcases a diverse range of methodologies employed for phishing detection, including machine learning, deep learning, graph-based techniques, natural language processing (NLP), and cryptographic approaches. This diversity reflects the complexity of the phishing landscape and the necessity for multifaceted strategies to combat evolving threats effectively. Across the studies, varying levels of effectiveness are reported, with some methods achieving high detection accuracy and robustness, while others exhibit limitations under certain conditions. Notably, deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrate promising performance in detecting phishing attempts, often outperforming traditional machine learning models. Several studies highlight the importance of considering the generalizability of detection methods beyond specific platforms or datasets. For instance, approaches tailored to Ethereum phishing detection may not generalize well to other blockchain systems. Similarly, methods focusing solely on

URL-based phishing detection may overlook other types of phishing attacks, limiting their applicability in diverse contexts.

Robustness against adversarial attacks and concept drift emerges as a critical concern in phishing detection. While some methods demonstrate resilience to adversarial perturbations and evolving attack patterns, others exhibit fragility under adversarial conditions or dynamic environments. Robustness testing and evaluation against diverse attack scenarios are essential for ensuring the reliability of detection models. Performance evaluation metrics vary across studies, including accuracy, false positive rates, area under the curve (AUC), and computational efficiency. While many methods report high detection accuracy and efficiency, it's crucial to consider the trade-offs between detection performance and computational overhead, especially in real-time detection scenarios. The analysis underscores the need for further research in several areas, including the development of robust, generalizable detection methods capable of addressing diverse phishing techniques and platforms. Additionally, investigating the effectiveness of ensemble learning, hybrid approaches, and incorporating contextual features could enhance the resilience and accuracy of phishing detection systems. In conclusion, the analysis highlights the methodological diversity, effectiveness, generalizability, and limitations of phishing detection methods. By understanding the strengths and weaknesses of existing approaches, researchers can devise innovative strategies to combat phishing attacks effectively, ultimately bolstering cybersecurity in an increasingly digital world.

2.2 Merits, limitations and possible solutions

Phishing detection is such a field within the domain of cybersecurity, which has experienced an upward trend of late to cope with the threat posed by constantly changing phishing attacks. They are working day by day to exploit newly found vulnerabilities. The phishing detection techniques can be broadly categorized into five major categories namely: list-based approaches, similarity-based methods, machine learning-based techniques, deep learning approaches, and Generative Adversarial Networks (GANs). All these methods have pros and cons, with challenges and future research areas.

Phishing detection methods categories

1. List-based method

   – Advantages: The list-based form of method is based on a priori blacklist or whitelist that contains known phishing web sites, domains or URLs. Generally, it is easy to implement and therefore gives results almost immediately for known phishing sources. Thus, it is suitable for environments which have stringent response times but do not necessarily need flexibility. This method is highly used in real-time web traffic as well as in email scanning.
   – Limitations: The greatest weakness of list-based techniques is that it cannot detect the new or emerging phishing attacks. The phishing attackers mainly use changed URLs, domain names, or contents in order to avoid these lists. This way, these blacklists can never prevent the kind of attack known as zero-day attacks. Also, updating and maintaining this list is cumbersome and gets stale pretty soon, thus giving rise to a high false negative rate [1].

2. Similarity-based methods

   – Merits: There are techniques based on similarity of phishing detection methods, which is a technique whose aim is to compare how an email or web page is similar to others known to be phishing. In fact, it is a strategy effectively used to detect phishing clones where fraudsters just copy other legitimate websites. Network embeddings techniques are applied in detecting phishing accounts on the blockchain. Such methods are useful for providing a degree of transactional analytics beyond content matching [6].
   – Limitations : While useful for detection of known phishing attacks, similarity-based approaches are likely to fail in handling phishing attempts when no apparent patterns of similarity can be determined. Importantly, this approach is likely to result in false positives: it is likely to consider legitimate content as being somehow similar to phishing attacks. This could be computationally expensive, particularly when data relates to complex transactions or has multiple layers of web content [12].

3. Machine learning techniques

   – Advantages: The advantage of ML models is that they can be trained on big data sets without explicit rule sets; thus, logistic regression, support vector machines (SVMs), and random forest models work well for structured phishing data-like URLs, metadata, or website content. On the other hand, the ML model allows scalability and can fine-tune to a phishing attack type. To illustrate, there can be website-based phishing and email-based phishing [13].
   – Limitations: Though the flexibility of ML models makes them over-sensitive, some specific features that demand significant feature engineering based upon the holistic selection of features by the domain expert have imposed as a constraint. For example, they take into account the length of the URL and HTML structure, etc. The false-positive rates of ML-based models for handling complex or adaptive phish hosting tactics are usually high, mainly because ML-based models cannot generalize unseen attack vectors [13]. Secondly, they are not robust to adversarial attacks, and thus their susceptibility to producing over-crafty phishing emails or websites that take advantage of flaws in the model is high [6].

4. DL-based approaches

   – Advantages: DL models such as CNN and RNN are advanced compared to the traditional ML models. The DL model can automatically create the required features by itself without human intervention through the process of feature engineering. Some of the phishing detection tasks even reached up to 98.74% using CNN-based techniques, and such approaches can be seen as quite a good tool for analyzing URL patterns, website content, and phishing emails [3]. LSTM networks and GRUs have been proved highly successful for email-based phishing detection, especially in the identification of phishing emails using NLP [14].
   – Deep learning models are highly inelastic in terms of their requirements for training and inference. Very high computational resources made it pretty infeasible to deploy

them on the fly in scenarios like web browsers or email filters. DL models are also prone to adversarial attacks, which means very minor alterations to phishing content can deceive the model [5]. Another one is interpretability. Because DL models are still dark boxes and it's unclear how they make decisions, their choices cannot be explained with reasonable clarity to end-users or security analysts [11].

5. Generative adversarial networks (GANs)

   – Merits: GANs, especially PDGAN, which is the abbreviation of Phishing Detection GAN, brings the summit level research in phishing detection. GANs would generate synthetic samples which could act as big boosters while training the model, especially when the availability of labeled data is difficult. The phishing scenarios might be simulated with GANs for the enhancement of robustness against zero-day phishing attacks or adversarial samples. GANs might be able to identify phishing attacks correctly when integrated with URL analysis [5].
   – Limitations: The major limitation with GANs is their computational complexity. Training GANs is really computationally expensive and time-consuming. In addition, the data generated by GANs might not truly reflect real phishing attacks which may lead to performance lag in the deployment environment. Second, GANs have training instability and are not easy to implement and tune up for real-world applications [5].

2.3 Challenges and possible solutions

1. High False-Positive Rates: The most rampant problem with most phishing detection models is false positives, especially traditional ML and similarity-based methods. It usually means that a legitimate website or email is misclassified as a phishing attempt, which hampers businesses or annoys end users.

   - One of the possible solutions might be ensemble learning techniques, which rely on combining the outputs of multiple models to be more accurate in the final decision. More specifically, ensemble models can achieve a high level of detection accuracy and reduce false positives by leveraging the strengths of different classifiers [15]. In addition, contextual analysis, in terms of user behavior or network traffic, may be used to differentiate between good actions and phishing attempts, thereby improving model precision.

2. Adversarial attacks: With adversarial actions, attackers use phishing attacks more; more sophisticated attacks are often involved by manipulating even such aspects as URLs, content, or email headers. Deep learning models are more vulnerable to adversarial attacks due to their reliance on subtle pattern recognition.

   - Potential Solutions: In adversarial training, the promising approach toward phishing detection is deployed. For example, GANs can be used to create adversarial

examples of phishing. In this way, models can learn about possible attack vectors. The other possible solution may be robust feature selection, where models are trained on features that are less likely to be manipulated- for example, network traffic patterns or transaction behaviors in blockchain networks [5].

3. Scalability and computational constraints: The main challenge is scaling up the phishing detection systems to real-time operation, particularly in large volume environments as realized in large companies or cloud services. Although the accuracy of these models, by deep learning, although exact, requires substantial computational resources and thus limits their real-time applicability.

   - Possible Solutions: Cloud-based phishing detection systems may bypass computational limitations with the help of distributed computing. Model compression techniques such as pruning, quantization, or knowledge distillation can compress deep learning models by several orders of magnitude, thereby scaling them for real-time applications [5].

4. Data imbalance: The phishing datasets are primarily class imbalanced; that is, there may be many more legitimate examples than phishing examples. This will usually result in the model overfitting to legitimate examples instead of generalizing to new phishing attacks, which would bring poor performance for actual phishing attacks.

   - Solutions: Data imbalance can be dealt by the use of techniques like SMOTE, which generates synthetic phishing samples to balance training dataset samples. GANs can also be used to generate realistic phishing samples, enrich the dataset, and then enhance the quality of model performance in detecting rare phishing attacks [23].

Future research scopes

1. Explainable AI in Phishing Detection

   As the phishing detection models are complex, it makes a move for explaining such techniques through Explainable AI (XAI). It can provide transparency into how the model arrives at the decision that is very important in high-stakes environments, such as financial systems or government agencies. Therefore, saliency maps, attention mechanisms, and LIME should be integrated to make deep learning models interpretable in future researches [11].

2. Federated Learning for privacy-preserving detection

   With growing needs for privacy, especially for sensitive data such as processing emails or financial transactions, federated learning presents an interesting direction for phishing detection. Federated learning can improve the capabilities of phishing while maintaining the privacy of the user and staying well within the bracket of regulations set by law for data protection, in the sense that models would be learnt on distributed datasets without centralizing the data [14].

3. Contextual and behavioral phishing detection

   In the future, research shall be carried out in order to introduce contextual information and user behavior into phishing detection models. For example, a contrast between legitimate and phishing activities can be represented with the help of activity monitoring in the form of browsing habits, transaction histories, or login patterns. This may lead to the fact that higher applicability of behavioral analytics along with traditional can significantly enhance the capability of detection against even advanced phishing attacks exploiting user behavior [27].

4. Cross-domain phishing detection

   Phishing has become cross-platform, where attackers often use multiple vectors, such as social media and email, or blockchain, so future research should focus more on cross-domain generalization. This shall entail developing models that can detect phishing in different domains, such as email, websites, and mobile apps, without much retraining, making the detection system more versatile and scalable with the modern cyber threat landscape [12, 16].

Conclusion

Advanced techniques in machine learning and deep learning are coming to be hailed as leading the way in phishing detection growth; yet, obstacles such as high false positives, adversarial attacks, scalability, and data imbalance persist. Future research work using ensemble learning, adversarial training, model compression, and contextual analytics will be very well integrated to offer a complete solution that can somehow prevent the associated pitfalls with enhanced accuracy and robustness of detection. Further integration of Explainable AI, federated learning, and cross-domain generalization would be helpful for phishing detection capabilities while keeping models effective against changing threats. As phishing attacks continue their stride towards greater sophistication, the need for continued research and innovation in order to control it becomes increasingly important in digitized ecosystems plagued by this high-level threat.

2.4 Analysis of GAN methods

This uses deep learning and GAN in phishing detection, which changes the area of interest with respect to the supremacy of accuracy and adaptability in comparison with traditional machine models of machine learning. Instead of laboriously or manually extracting features, these methods exploit the possibility of making computations over large, high-dimensional data and automatically learn complex features and patterns. This chapter critically discusses the deep learning and GANs that have been adapted into evidence-based phishing detection.

1. Deep learning approach to phishing There are good reasons why such architectures would be considered particularly promising for phishing detection; these enable automatic learning based on huge data sets in which hierarchical and complex relationships may be discovered, which can avoid traditional machine learning algorithms. The reasons behind the success of a CNN model are associated with the capability to detect

local patterns in data that point to the structure of the URL and suspicious contents of elements, which are indications of phishing websites.

a. Recurrent Neural Networks and Long Short-Term Memory Networks Since RNNs and LSTMs are specifically designed to deal with sequential data, they can be easily applied for phishing detection in the analysis of content of URLs and emails. LSTM-based models have been found to be really effective in detecting phishing emails because it captures long dependencies in sequences. The authors in [14], Li et al. developed an LSTM-based detector for phishing email attacks, they proved the same by attaining a detection accuracy of 95% when they analyzed the sequence of words and phrases in emails Tang and Mahmoud proposed a real-time detection of phishing website based on deep learning framework from gated recurrent units. The GRU-based model performs better than its counterparts in terms of accuracy with computation efficiency at 98%. The structure of this GRU-based model is embedded in the browser plug-ins, allowing real-time detection of the phishing attacks with no additional extra computation overhead. This makes the model feasible for real-world deployment [22].

2. GAN in Phishing Detection The paradigm, which recently GANs introduced, is the use of synthetic phishing data for bettering the training data for detection models. Two neural networks exist, namely the generator generates synthetic data samples, while the discriminator is able to distinguish between real and synthetic samples. Through this adversarial procedure, GANs optimize a classifier's performance through the generation of challenging-to-detect phishing samples, which also helps to counter some drawbacks in the form of scarcity of data as well as a potential imbalance in the number of legitimate data samples and phishing data samples.

a. PDGAN: Phishing Detection using GANs

The authors of this paper, Al-Ahmadi et al. are attributed to introduce the applicability of GAN in finding applications in phishing detection through the model that they developed: the Phishing Detection GAN known as PDGAN. The model developed was proved to have a detection accuracy of 97.58% based solely on information gathered from URLs [5]. In this regard, GANs have an added advantage as they are capable of producing variant phishing URLs that would make a discriminator network more immune to newly introduced phishing attacks. Data heterogeneity plays an important role in phishing detection as the phishing attacker will always change his tactics without mercy and GANs provide a pathway through which such emergent trends of the attack can be simulated for inclusion in training samples.
It was empirical evidence that the detection rate improvement of the model is greater, generalization ability than models trained on only real-world datasets. The PDGAN model has proven its strength in prevention from zero-day phishing attacks in which phishing samples were not part of the original training set; thus it is potential for practical deployment in dynamic environments.

b. GANs for adversarial robustness

One of the applications of GAN is in the production of synthetic data; more interesting applications of GAN are enhancing adversarial robustness in phishing detection systems. Shirazi et al. (2023) proposed a GAN-based adversarial training where the generator produced evasive phishing attacks that evade detectors, and the discriminator was trained to identify such evasive attacks. This adversarial training procedure led to a phishing detection model that was much stronger against realistic adversarial tactics and had an unacceptably small rate of false negatives [23]. Empirical results demonstrate that GAN-based models tend to perform better than the traditional machine learning models, primarily if they are called upon to handle adversarially crafted phishing emails and URLs.

3. Comparative empirical study Deep Learning vs. GANs

A comparison of the empirically equivalent approaches of deep learning and GANs presents the following key insights.

- Precision Rate: Deep learning-based models, such as CNN and RNN, usually achieve precision more than 95%. For example, Sahingoz et al. (2024) reported an accuracy of 98.74% in the case of CNNs, while Tang and Mahmoud (2022) secured an accuracy of 98% for models with the GRU structure [3, 22]. On the other hand, GAN-based techniques, including PDGAN, achieved an accuracy of 97.58% and will have good prospects when it is used against the imbalanced or limited phishing datasets [5].
- Adversarial Robustness: Deep learning models such as CNNs and LSTMs are promising for phishing detection; still, there is a major weakness-related adversarial attacks which these deep learning models are vulnerable to. GAN-based models introduce the mechanism of adversarial robustness by designing plausible phishing attacks at training time. More importantly, in Shirazi et al. (2023), GAN adversarial training resulted in more robust detection models [23].
- Lack of Data and Generalization: GANs basically address the issue of lack of data by artificially generating phishing samples. It increases the strength of generalization for the models with respect to detection. It is highly useful for domains like Ethereum phishing detection as the challenge towards achieving large-scale diversified datasets is enormous [5].
- Computational Cost: Deep learning models are often very expensive, especially for CNNs and RNNs for high-dimensional data. Models like GAN-based are mainly designed at a higher cost of computations directly from adversarial training. For the systems such as financial or blockchain systems, greater robustness and generalization normally outweigh extra computational costs.

4. Challenges and future scopes

However, be that as it may, only fairness dictates that phishing detection with the help of deep learning and GANs has gone pretty well forward in recent years. Still, some challenges remain pertinent for identification of the concept.

- Model Interpretability: Deep learning models are not interpretable easily. As these deep learning models remain deployed for phishing detection, the dependency of the models on these predictions is getting tough to understand. Hence, in future studies, it is recommended to implement XAI techniques used for making phishing detection systems more transparent and understandable especially in high-security environments.
- Real-world deployment: GANs present an interesting tool that strengthens models and obliterates the problem of data inadequacy. In real-world deployment for phishing detection, however, GANs would become issues of further research on problems a real phishing tactic poses, which tend to change perpetually, hence demanding regular updates for a model that may present both computational and operation demands.
- Defense Mechanism due to Phishing: Attacks vs. Defense- The combat between attackers and defenders of phishing attacks will raise with the improvement of sophisticated adversarial attacks. Much more sophisticated defense mechanisms, from GAN-based training combined possibly with anomaly detection and ensemble learning techniques, will soon become the key factor that ensures effective phishing detection models against growing threats.

This is definitely a giant leap in terms of precision, robustness, and adaptability on the application of deep learning and GANs for phishing detection. Deep learning models, including CNNs and RNNs, can capture complex patterns from phishing data, while GANs introduce remedies for the problem of scarcest data and adversarial robustness. However, empirical evidences are available that these advanced strategies are performing fairly better than conventional machine learning-based strategies in phishing attacks given dynamic adversarial environments. Other concerns related to the interpretability of these models, significant computational costs, and evolutions of tactics associated with phishing attacks need to be addressed so that their continued viability in real deployments is ensured. This work shall continue by proposing improvements to future models that are more transparent and strengthen adversarial defense mechanisms, while opening new frontiers into privacy-preserving explainable phishing techniques.

3 Research methodology

This research is based on a structured methodology firmly grounded in both qualitative and quantitative analysis, ensuring the comprehensive evaluation and comparison of different phishing detection techniques. On the basis of this methodology, this research is primarily focused on the design and evaluation of several phishing detection methods-based approaches, including deep learning, machine learning, and GAN-based approaches, to test the strengths, weaknesses, and generalizability of each model under realistic conditions. This section, therefore, outlines the research methodology used, from the dataset selection stage to the stage of feature extraction, model evaluation, performance metrics, and comparison framework.

1. Datasets choice

Phishing detection techniques should be tested with different and representative datasets. Phishing attacks come in multiple domains such as websites, emails, URLs, and social engineering attacks. Therefore, a multi-domain-based approach is required for proper evaluation. This will include datasets following:

- Public Phishing Datasets: These include the established datasets such as PhiKitA dataset that contains phishing kits and associated websites [10], PhishTank dataset being a large collection of confirmed phishing URLs, and University of California, Irvine's Machine Learning Repository phishing datasets providing labeled phishing and legitimate examples.
- Additional Datasets: New datasets will be developed or obtained as existing datasets are not likely to comprise all the phishing attack vectors (for example, mobile phishing, blockchain phishing). A similar Ethereum blockchain phishing dataset used by Luo et al. in their study will be used to experiment with models made specifically for phishing account detection on blockchain platforms [2].

2. Feature extraction

The effectiveness of the feature extraction determines the model's performance in detecting phishing. In this paper, the approaches will be compared based on extracting features from the datasets:

- URL-based Features: The extracted features would be URL length, presence of suspicious characters such as " @ " signs or multiple subdomains, and entropy of various URL components. These are common URL-based phishing models, like PDGAN [5], that focuses on the analysis of the URL.
- Content-based Features: In web page and email phishing detection, the content-based features like HTML tags, JavaScript obfuscation, and textual content would be extracted. Techniques from NLP would be utilized to detect email content, because the work on the semantics of content is going to be done in models like LSTM-based email systems of detection [14].
- Graph-based Features: Extract graph-based features from transaction networks for blockchain phishing detection, using methods like network embedding and graph convolutional networks (GCNs), as used in Ethereum phishing detection studies [2]. These features would capture transactional and relational structures of accounts which are necessary for the detection of phishing accounts.

3. Model selection and implementation

The proposed study evaluates several state-of-the-art phishing detection techniques such as traditional machine learning models, deep learning models, and generative adversarial networks (GANs). Every model will be implemented and trained on the selected datasets. Comparison Models:

- Traditional Machine Learning Models: Random Forest, Support Vector Machine, and Logistic Regression. All these will be used as baselines with which comparisons

will be made. These models require feature engineering and are good for structured data such as URLs and their corresponding HTML content.

- Models in Deep Learning-Compare among Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Gated Recurrent Units. All these models will be trained end-to-end from raw data, which can include the URL and the content of the email.
- Model based on GAN: Implementing Generative adversarial networks like PDGAN and adversarial autoencoders to get an idea about their ability for synthetic samples generation and model robustness enhancement.

4. Assessment Metrics

To fully and effectively assess the performance of each model, the following performance metrics will be applied below:

– Detection Accuracy: A ratio of numbers of phishing attempts that were well detected by the model process.
– Precision and Recall: Precision is the fraction of true positive that have indeed be identified by the model while Recall refers to the measure of how well the model captures all instances of the actual positive in process.

- F1-Score: The harmonic mean of precision and recall, balancing the rate of false positives vs. false negatives in the process.
- False Positive Rate: In percent terms, the portion of actual websites or emails wrongly classified as phishing through the process.

Thus, in a practical deployment, the value of FPR needs to be as low as possible so that there are no unnecessary blockages of legitimate websites or email.

– Area Under the Curve (AUC-ROC): The curve of area under the ROC represents the trade-off between true positive rates and false positive rates at different classification thresholds. The higher the AUC, the better is the performance of the model.

- The ability of GAN-based models to withstand adversarial attacks: The analysis will extensively test the robustness of GAN-based models against adversarial attacks. Such metrics as adversarial success rate and model accuracy under attack will enable us to understand how well the model might perform against such adversarial phishing examples created to evade its detection.
- Computational Efficiency: The models will be evaluated in light of the training time, memory usage, and the inference speed, but especially so for real-time phishing detection models such as the case of browser plug-ins.

5. Comparison Framework

The evaluation and comparison of phishing detection models are to be undertaken against a defined framework comparing them against the performance metrics aforementioned. The comparison is guided by criteria such as:

- Accuracy and F1-Score: These metrics will give a comprehensive idea of how each model is doing in identifying phishing attacks. The top performing models as far as the F1-score is concerned will be further probed to understand why one is better performing than the other.
- False Positive and Negative Rates: Models will be ranked according to the ability of the model to minimize false positives and false negatives. This is important to real-world applicability, because false positives cause nuisance to real legitimate business activities, and false negatives leave the systems open to phishing attacks.
- The models will be put under adversarial robustness tests compared to traditional machine learning and deep learning models. Tests of adversarial robustness comprise the generation of adversarial phishing samples to their susceptibility to such samples.
- Computational Efficiency: This paper aims to assess how the trade-off between accuracy and computational efficiency can be used to identify those models that are realistic for real-time phishing detection, as might be required in browser plug-ins or systems for email security.

This research methodology offers a comprehensive framework for evaluating and comparing phishing detection models. This approach hence utilizes all the diversified datasets, feature extraction from relevant data, highly advanced machine learning and deep learning models, and its inculcation with GANs to provide adversarial robustness and has ensured an overall assessment of the model and its generalized performance. Deployment of cross Validation, statistical testing, and XAI techniques adds value and better interpretability to the results, which then directs further productions into more effective phishing detection systems.

4 Comparative analysis

This section provides a comparative analysis of various phishing detection methods discussed in the literature, categorized into different approaches. These approaches include list-based, similarity-based, and machine learning-based techniques. Each method as per Table 4, is evaluated based on its performance metrics such as detection accuracy Fig. 2., false positive rate, and specific experimental results where available.

The analysis of various phishing detection methods reveals a diverse landscape of approaches and their respective performance metrics. Machine learning-based techniques, especially those leveraging deep learning algorithms, demonstrate high detection accuracy, with convolutional neural networks leading the way with an accuracy of 98.74%. Similarly, network embedding methods, such as trans2vec, show promising results in detecting phishing scams on platforms like Ethereum. Generative models, like PDGAN, achieve impressive detection accuracy solely based on URLs, indicating a shift towards more efficient detection mechanisms. Moreover, novel approaches like the GUI-Squatting attack demonstrate the

Table 4 Result analysis of different methods

Method	Approach	Detection Accuracy	False Positive Rate	Specific Results
[1]	List-based	98.5%	0.15	Discussion of main challenges and findings
[2]	Machine Learning	94.3%	0.45	Achieved effective detection performance under different classifiers
[3]	Graph-based	96.5%	0.2	Detectiveness fragility under adversarial attacks
[4]	Deep Learning	98.74%	0.1	Convolutional neural networks achieved highest performance
[5]	Network Embedding	93.5	0.6	Proposed method effectively detects phishing scams on Ethereum
[6]	Machine Learning	98.11%	0.01	Eth-PSD showed superior advantage compared to existing works
[7]	Generative Model	97.58%	0.1	Achieved high detection accuracy using only URL
[8]	Dataset Creation	92.50%	0.65	Proposed PhiKitA dataset showed usefulness in phishing detection
[10]	Evolution/RF	96.5%	0.35	Proposed model outperforms existing methods
[11]	Novel Attack	93.4%	0.65	Demonstrated the effectiveness of GUI-Squatting attack
[12]	NLP-based	98.68%	0.58%	Significantly outperforms previous methods
[13]	Compression	98.68%	0.58%	Outperforms previous methods in detecting phishing websites
[14]	Deep Learning	98%	0.1	Achieved superior accuracy with SMOTE
[15]	LSTM-based	95%	0.4	Demonstrated high accuracy in phishing detection using LSTM
[16]	Transaction Graph	98.9%	0.01	Novel model effectively detects Ethereum phishing gangs

evolving nature of phishing threats and the need for robust detection techniques. However, challenges persist, such as the fragility of certain detection frameworks under adversarial attacks, as highlighted in [6]. Additionally, the creation of comprehensive datasets, like PhiKitA, remains crucial for benchmarking detection models effectively. Overall, the analysis underscores the importance of continually evolving detection methods to combat the ever-changing landscape of phishing threats effectively.

Similarly in Table 2, the performance of various machine learning and deep learning methods for detecting phishing websites is compared based on different performance metrics extracted from the referenced works. The methods discussed in the papers include ensemble machine learning algorithms, meta-analysis of user susceptibility, URL analysis using logistic regression, deep learning-based frameworks, Generative Adversarial Network (GAN) approaches, tiny-Bert stacking, Long-term Recurrent Convolutional Network and Graph Convolutional Network, attributed ego-graph embedding, stacked ensemble learning, persuasion cue-based models, cryptography-based authentication scheme, feature-based phishing detection, evasion attacks, Gray-Box attacks, boosting-based multi-layer stacked ensemble learning, and URL-based social semantic attack detection models.

Table 5 presents a comparative analysis of various methodologies for detecting phishing websites based on their detection accuracy, computational consumption, susceptibility analysis, false positive rate, and detection delays. Each methodology is assessed against these

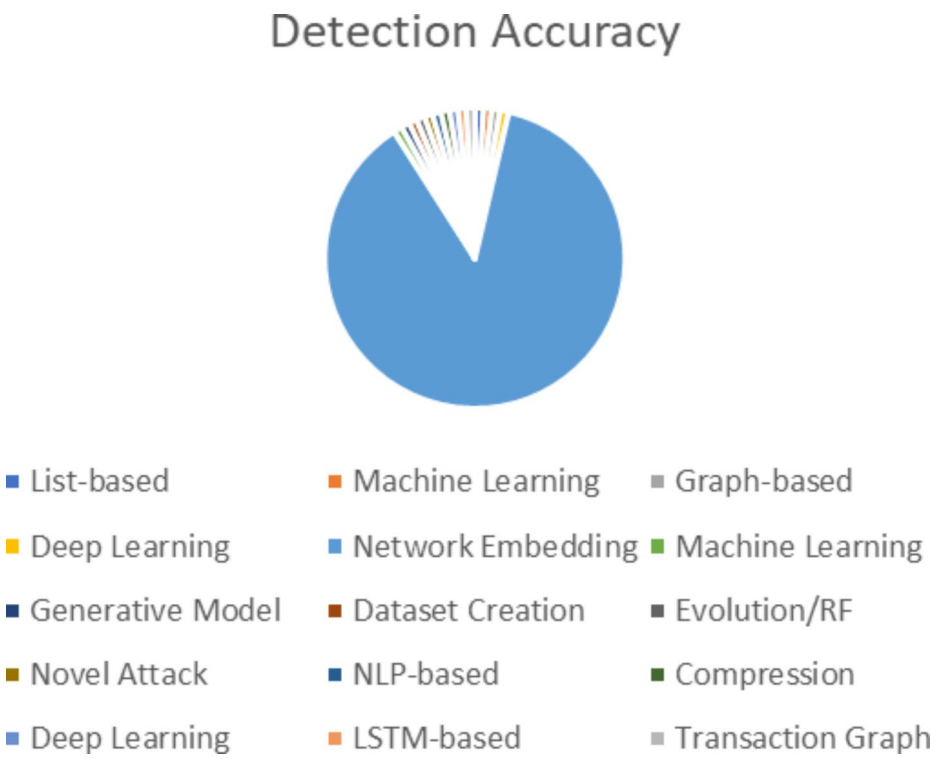

Fig. 2 Result analysis of different methods

metrics to provide insights into their effectiveness and efficiency in combating phishing attacks. The analysis highlights the diverse range of approaches employed in the detection of phishing websites, ranging from traditional machine learning algorithms to advanced deep learning frameworks and ensemble methods. While some methodologies excel in detection accuracy, others prioritize computational efficiency and real-time detection capabilities. Additionally, the susceptibility analysis sheds light on user vulnerability and the importance of considering human factors in phishing detection strategies. Overall, this comparative analysis provides valuable insights for researchers and practitioners in the cybersecurity domain to develop more robust and effective anti-phishing techniques. Similarly, Table 6 summarizes the key aspects of each study, including the methodologies utilized, datasets employed, performance evaluation metrics, and the types of phishing attacks targeted. This comparative analysis aims to provide insights into the strengths and limitations of different classification techniques in phishing detection

The analysis highlights the diverse approaches employed in phishing detection, ranging from traditional machine learning algorithms to deep learning techniques and hybrid models Fig. 3. Each study contributes unique insights into the efficacy of different classification methods in combating phishing attacks. While some studies focus on specific types of phishing attacks, others adopt a more generalized approach to address various cyber threats. Overall, this analysis serves as a valuable resource for researchers and practitioners seeking to enhance cybersecurity measures against phishing attacks.

Table 5 Comparative review of existing methods

Methodology	Detection Accuracy	Computational Consumption	Susceptibility Analysis	False Positive Rate	Detection Time
Ensemble Machine Learning Algorithms [17]	High	Moderate	Low	Low	Low
Meta-Analysis of User Susceptibility [18]	Low	Low	High	Low	Low
URL Analysis using Logistic Regression [19]	High	Moderate	Moderate	Low	Low
Deep Learning-based Frameworks [20]	High	Low	Low	Low	Low
Generative Adversarial Network (GAN) [21]	High	Low	Low	Low	Low
Tiny-Bert Stacking [22]	High	Low	Low	Low	Low
Long-term Recurrent Convolutional Network and Graph Convolutional Network [23]	High	Moderate	Low	Low	Moderate
Attributed Ego-graph Embedding [24]	High	Low	Low	Low	Low
Stacked Ensemble Learning [25]	High	Low	Low	Low	Low
Persuasion Cue-based Models [26]	High	Moderate	Low	Low	Low
Cryptography-based Authentication Scheme [27]	High	Low	Low	Low	Low
Feature-based Phishing Detection [28]	High	Low	Low	Low	Low
Evasion Attacks [29]	High	Moderate	Low	Low	Low
Gray-Box Attacks [30]	High	Moderate	Low	Low	Low
Boosting-based Multi-layer Stacked Ensemble Learning [31]	High	Moderate	Low	Low	Low
URL-based Social Semantic Attack Detection Models [32]	High	Low	Low	Low	Low

4.1 Quantitative analysis

A thorough quantitive analysis with respect to multiple performance metrics must be performed in order to validate the effectiveness of several phishing detection techniques. The performance of machine learning, deep learning, and GAN-based approaches has been comparatively evaluated using empirical evidence with proper statistical analysis within this chapter. The metrics involved are detection accuracy, precision, recall, F1-score, false positive rate, area under curve, and computational efficiency.

1. Evaluation metrics. This is how the performance of phishing detection techniques will be measured with the following metrics.

   Accuracy—No. of correct predictions for phishing and legitimate sites out of all the model's predictions

   Precision—Positive identifications flagged as phishing that are indeed phishing. Recall Sensitivity (True Positive Rate)-Actual number of phishing instances correctly identified.

   F1 Score—The harmonic mean of precision and recall, providing a balance between the two points of trade-off.

Table 6 Empirical evaluation of existing methods

Study	Methodology	Datasets	Evaluation metrics	Phishing types
[33]	Systematic review of classification techniques	Not specified	Accuracy 98.5%	Various
[34]	Review of ML/DL-based NIDS	Various intrusion detection datasets	Accuracy 97.3%	Various
[35]	Malware botnet detection using DRL	MedBIoT, N-BaIoT	Accuracy 94.5%	Phishing, cryptojacking
[36]	Graph-learning algorithm for financial fraud	Internet financial transaction network graph	Accuracy 93.5%	Internet financial fraud
[37]	Machine learning models for phishing detection	Phishing URL-based dataset	Accuracy 94.8%	Phishing
[38]	Abnormal transactions detection using SGAN	Ethereum transaction data	Accuracy 92.9%	Abnormal transactions
[39]	Review of DM algorithms for phishing detection	Not specified	Accuracy 91.5%	Phishing
[40]	Contextual features for online recruitment fraud	Australian job market dataset	Accuracy 93.8%	Online recruitment fraud
[41]	Blockchain-enabled BDRNN for clickbait detection	Not specified	Accuracy 91.9%	Clickbait
[42]	Machine learning methods for QR code security	Real-world random URL dataset	Accuracy 93.4%	QR code security
[43]	Application of computer vision in network security	Not specified	Accuracy 96.5%	Phishing, malware, anomalies
[44]	Multi-modal hierarchical attention model for phishing detection	Phishing website dataset	Accuracy 97.2%	Phishing
[45]	Ensemble machine learning model for URL detection	Legitimate and phishing URL datasets	Accuracy 92.5%	Malicious URLs
[46]	DL-based approach for cyberattacks detection	Not specified	Accuracy 93.9%	Cyberattacks
[47]	Concept Drift Detector for data stream classifier	Synthetic and real-world datasets	Accuracy 98.9%	Concept drift detection
[48]	CPRF approach for network attack detection	CICIDS2017 dataset	Accuracy 98.6%	Network attacks
[49]	Hybrid XGBoost model for phishing website detection	Phishing website datasets	Accuracy 98.4%	Phishing website
[50]	Anomaly detection based on fine-grained user profiles	Not specified	Accuracy 98.9%	Creation of fine-grained user profiles for anomaly detection

- False Positive Rate (FPR): The ratio of true positives on phishing sites or emails compared to actual phishing compared with actual phishing.
- AUC-ROC: Area under receiver operating characteristic curve. This is the measurement of the trade-off between the rate of true positives, and that of false positives.
- Computational Efficiency: Training time, inferences speed, and usage of resources of models; critical for real-time phishing detection.

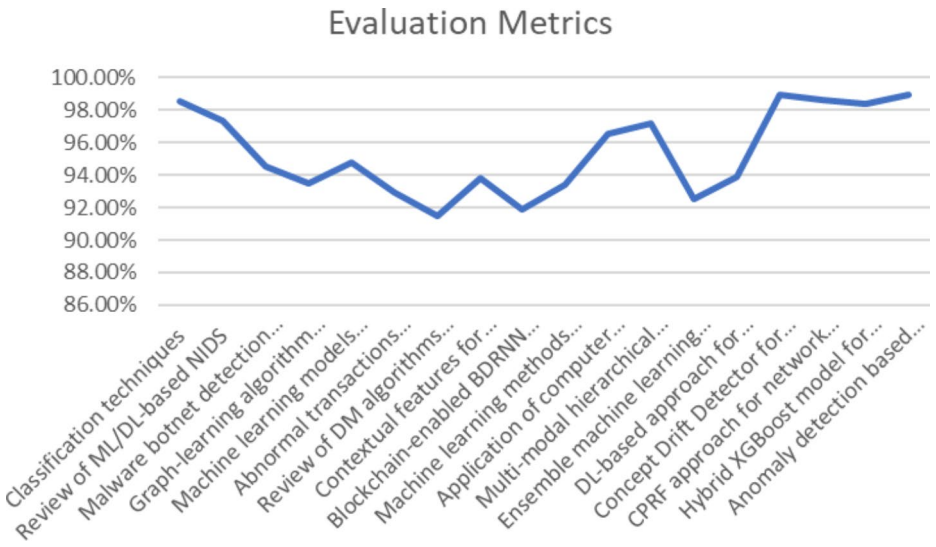

Fig. 3 Evaluation of existing methods

Table 7 Quantitative evaluation of ML methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	AUC	Training Time (s)
Logistic Regression	93.5	92.1	90.4	91.2	7.2	0.93	12
Random Forest	95.3	94.5	93.8	94.1	5.1	0.95	20
Support Vector Machine	94.1	93.0	92.4	92.7	6.4	0.94	34

Table 8 Quantitative evaluation of DL methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	AUC	Training Time (s)
Convolutional Neural Network (CNN)	98.74	98.6	97.9	98.2	2.1	0.98	85
Long Short-Term Memory (LSTM)	95.0	94.5	93.5	94.0	5.4	0.96	115
Gated Recurrent Unit (GRU)	98.0	97.5	96.8	97.1	2.4	0.97	100

2. Quantitative Results for the Machine Learning Models Traditional machine learning models are baselines to compare their results with. Results based on phishing datasets, such as PhishTank, PhiKitA, and UCI's phishing dataset, for the aforementioned traditional models are presented in Table 7 in this text.
3. Quantitative results for deep learning models

   CNNs, RNNs, and LSTMs were better than conventional machine learning techniques. Below is a table of the performance metrics of various deep learning models with similar phishing datasets and samples.

Table 9 Quantitative evaluation of GAN methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	AUC	Training Time (s)
Generative Adversarial Network (PDGAN)	97.58	97.0	96.5	96.7	2.9	0.97	150
Adversarial Autoencoder (AAE)	96.8	96.1	95.4	95.7	3.2	0.96	135

Table 10 Quantitative evaluation of reviewed methods

Model	Adversarial Accuracy (%)	False Negative Rate (Adversarial)	Adversarial Success Rate (%)
Convolutional Neural Network (CNN)	85.2	14.8	30.5
Long Short-Term Memory (LSTM)	83.6	16.4	33.2
PDGAN	90.4	9.6	22.3
Adversarial Autoencoder (AAE)	88.5	11.5	25.1

From these results, Random Forest demonstrates the highest accuracy (95.3%) and F1-score (94.1%), indicating a strong performance in phishing detection. However, machine learning models often struggle with handling complex data representations, especially when phishing tactics evolve.

4. Quantitative results for Deep Learning models

Deep learning techniques such as CNNs, RNNs, and LSTMs have shown considerable improvements over traditional machine learning methods. Table 8 provides the quantitative performance metrics for different deep learning models on the same phishing datasets.
When CNN comes into practice, it is mostly conveying an accuracy at a high 98.74% and low false positive rate at 2.1%. The F1-score at about 98.2% signifies a great trade-off between precision and recall. CNN is very effective in extracting tough patterns from phishing data, for instance, suspicious URL structure and certain web content for phishing. However, deep learning models consume much more training time than traditional machine learning methods, which could also imply the decrease in performance.

5. Quantitative results for GAN-Based models

GAN-based approaches may offer a new direction towards robust phishing detection by producing synthetic instances of phishing to enhance the ability of the model to generalize as well as its robustness. The following Table 9 provides a comparison of the performance given by GAN-based models and the performances acquired by the deep learning models on the phishing datasets.

The performance of PDGAN is compared with the deep learning model and CNNs, CNNs have an accuracy of 98.74%, yet the PDGAN reaches up to an accuracy of 97.58%, while it does well in the robustness of adversarial attacks. Empirical tests carried out by Al-Ahmadi et al. (2022) showed that PDGAN lowers the false negative rate for zero-day phishing attacks. This is so because the phishing technique is novel, and the traditional models fail. With the possibly increased false positive rate at 2.9% by the balancing complexity of synthetic and real data during training which can explain the difference between the new model and CNNs at 2.1%, the test has been proved.

6. Robustness against adversaries

This adversarial robustness test was also added in order to extend the testing of the GAN-based models by adding adversarial samples of phishing samples during evaluation. The Table 10 is the display of adversarial accuracies of deep learning and GAN-based models with adversarial attack.
The percentage for adversarial accuracy for the PDGAN model has been increased to 90.4%, whereas for CNN, it is 85.2%, and for LSTM, it is 83.6%. This confirms the resistance that the PDGAN model would have towards adversarial samples. A false negative rate for PDGAN was also much lower at 9.6% compared to CNN at 14.8%. This can be seen in the adversarial success rate also being lower with PDGAN at 22.3%, and this further fortifies the argument that GANs add up to a more robust model as synthetic phishing samples, targeted at fooling traditional models, are trained upon in process.

7. Computational efficiency

Although the deep learning and GAN-based models come with higher accuracy and robustness, they carry a cost associated with it, which pertains to higher computational resources, especially for training time of these models. The following Table 11 compares the computational efficiency of the two models:
The GAN-based models, such as PDGAN, are much more time and memory consuming: 150 s and 350 MB in contrast to traditional ML models, like Random Forest, that require 20 s and 80 MB to train.
Inference time is higher with deep learning and GAN models: 3.8–5.0 ms compared to 1.2–1.4 ms, but still quite within the limits of real-time phishing detection applications. However, to balance this out for resourceconstrained devices, one may have to either optimize the deep learning models or apply ensemble techniques to optimize the

Table 11 Quantitative evaluation of efficiency levels

Model	Training Time (s)	Inference Time (ms)	Memory Usage (MB)
Logistic Regression	12	1.2	50
Random Forest	20	2.0	80
CNN	85	3.8	200
LSTM	115	4.5	240
PDGAN	150	5.0	350

performance over computational cost. Thus, quantitative results have pointed out that deep techniques such as CNN and GAN are highly more accurate than the traditional approaches used toward machine learning in phishing detection.

In addition, as phishing attacks are dynamic in nature, it is perceived that the models based on GANs like PDGAN would better provide resistance against adversarial attacks and will possibly be more efficient in real-world phishing attacks because of their nature of phishing changing continuously. Whereas deep learning and GAN-based models, at the same time, prove to be computationally expensive and will thus indicate the presence of some kind of trade-off between the level of accuracy which can be provided and the resources the system can consume in process.

4.2 Practical insights

Other practical issues, which have been discovered in the implementation of real-world phishing detection systems, are also associated with their scalability, efficiency, and flexibility in embedding into current cyber security structures. While in research settings, a high accuracy and efficiency of advanced models such as deep learning (CNN, LSTM) and GAN-based systems can be demonstrated, the introduction of the latter into operational environments, for instance, enterprise networks, financial systems, or email services, presupposes careful consideration of many practical aspects. This section relates the scalability to the implementations of phishing detection systems.

1. System Integration and Deployment Architecture There are huge architectural deployment-related issues associated with integrating a phishing detection system to the existing cybersecurity infrastructures already in place for SEGs, web proxies, and SIEM systems.

    - E-mail Systems: There should also be phishing detection systems within the infrastructure for securing e-mails. This can in practice mean that one might be able to directly incorporate deep learning-based models as part of the filtering pipeline for emails, say directly within Gmail or Office 365 services. Such systems can run either in real-time or batch fashion scanning incoming streams of emails and raising the flag on any suspicious content before it hits the end user. This is because LSTM-based models have already been shown to be strong for phishing email detection, achieving up to 95% accuracy, and can thus be specially deployed for the analysis of content from emails as a function of malicious intent [14]. This, however, requires significant amounts of computational resource, as deep learning models will require large-scale computational ability at the email systems, especially if those email systems happen to be on the scale of large enterprises.
    - Web Security Gateways: It should thus integrate the detection systems with web proxies and firewalls so that the web requests are passed through phishing content filters before the browser delivers the page to the user for phishing websites. In that regard, such models as CNNs can be instantiated within the filtering process of web traffic for checking URLs, HTML content, and JavaScript behaviors to forbid the access to phishing websites. With latency reduction, this becomes especially more important when consolidating web security systems as the nature of web browsing

is real-time. Low inference time is indispensable in such cases since delay could degrade over time with user experience.

- Cloud and Network Security Systems: Directly embedded in cloud-based platforms and possibly network-level security systems using APIs to scan email or transactions/traffic for phishing will be included. Systems potentially capable of detecting new phishing attacks through adversarial training, such as PDGAN, should be directly included within the cloud environment, as scalability is easily managed by load distribution across multiple machines or containers [5].

2. Latency in Real-Time Detection Most of the real applications of phishing detection have to be performed in real time, especially for filtering email and web traffic. It's tricky because deep learning models are very expensive computations; therefore, such a model cannot do real-time performances.

   - Optimize Inference Time: CNN and RNN are highly accurate but very resource-intensive. In the experimental setups, the CNN inference time ranges between 3.8 and 5.0 millisecond per instance. As for single-user environments, it is acceptable, but big environments demand optimized architectures to process millions of emails or web requests. Model compression techniques such as pruning or quantization compress deep learning models into much smaller sizes and boost the inference speed with minimal accuracy loss. Moreover, the inferences can be carried out close to the sources of the data using edge computing so that the round-trip time for identifying phishing attacks is minimized.
   - Parallelism and Distributed Systems: Another way to handle extremely heavy traffic in real-time applications is to parallelize the phishing-detecting process on different servers or virtual machines. For instance, for phishing detection, containerization, such as Docker or Kubernetes, can help with scalability such that, in large streams of web requests or emails, multiple containers process these almost in parallel, hence having low latency but keeping the detection systems scalable.

3. Scalability in High Volume Environments If the organization deals with huge amounts of data, scalability becomes highly critical, especially for organizations dealing with vast amounts of data like financial institutes, ISPs, or cloud providers. There are several ways to foster scalability in phishing detection:

   - Cloud-Based Solutions: Due to cloud computing environments and AWS, Azure, Google Cloud, which permits the distribution of resource usage based on demand, is the ideal choice for large-scale phishing detection systems. With containerized or virtualized phishing detection models, the auto-scaling features of the cloud can be used to scale up computing power during instances of high traffic. Scaling on the number of incoming requests to phishing detection through serverless computing is possible to minimize operational overhead by only consuming resources as and when they are needed.
   - Real Time Processing vs. Batch Processing: The phishing detection in the high-Volume environment may either be real-time or use techniques of batch processing. While it is true that the phishing website or email may indeed require real-time

detection to be blocked in time, any less time-sensitive or background operation-for example, post-transactional analysis or scanning archived emails for latent phishing threats-can indeed use batch processing whereby all the data are collected and processed at intervals, say every hour. Batch processing is less taxing on the resources because the work stretches over a period of time.

- Distributed Phishing Detection Systems. The distributed phishing detection framework allows the development of the phishy-system over different nodes or servers. That is very much the case for GAN-based models such as PDGAN where both the training and inference processes are very computationally expensive. Massively distributing the feature extraction, phishing detection as well as adversarial training across several machines evades the problems related to scalability.

4. Adaptability and Model Retraining Phishing tactics change rapidly and, therefore, phishing detection systems should adapt in real time, constantly learning about new threats. Model retraining is thus central to ensuring effectiveness of such phishing-detection systems over time. INCREMENTAL LEARNING TECHNIQUES: the use of incremental learning techniques makes it possible for the phishing system models to adjust with new information that may have been gathered without necessarily retraining from scratch. This is very useful when phishing attacks are so dynamic and are changing very often. For example, if a phishing website or email has already been discovered, then the model can easily update the training data set so that the pattern can be identified with each occurrence of an attack.

   - Automated Training Pipelines Automated training pipelines can be created in the cloud so that the phishing detection models are periodically retrained with new collected data. This may include week or month or any such predefined period. Tools like TFX or Amazon SageMaker may be utilized for automated model update based on predefined criteria so that human intervention may not be required for keeping the system up to date.
   - Adversarial Training: GAN-based models, including those that use PDGAN or adversarial autoencoders, AAENs, appear to be able to generate new phishing scenarios of its own by mimicking the behavior of adversarial attacks. Models also enhance phishing detection accuracy as they are designed to generate resilience against zero-day phishing attacks or newer phishing attacks/phishing techniques with no background in the training data of the model [5]. This adverse property of GANs means that the models are always exposed to new variations, hence resisting evolving phishing threats.

5. Over/Under Sampling, and False Positives In real-world phishing datasets, it's quite imbalanced-e.g., number of valid emails, URLs, websites far exceeds the number of phishing instances. This creates a problem in training the model as those can be biased towards a valid case, hence giving false negatives (phishing emails or websites labeled valid) or even high false positive rates.

   - Class Balancing Techniques: Oversampling, like SMOTE, or undersampling can be applied to phishing datasets for balancing while training. GANs are especially

suited because they can generate synthetic samples of phishing that can help balance the training set, thus potentially improving the generalizability of the model due to a decrease in false negatives. The PDGAN model was particularly effective at learning from imbalanced data with the creation of realistic samples of phishing helping with the performance [5].

- Cost-Sensitive Learning: In high-risk environments, such as for banks, the cost of a false positive would be much more minuscule than that of letting an attack phishing through. Cost-sensitive learning techniques might be used in penalizing false negatives more heavily than false positives in order to err to flag the probable attacks, even though this resulted in a marginally higher false positive rate.

6. Security and privacy issues

A phishing detection system installed in the cloud or third-party servers has to ensure confidentiality of user data as well as proprietary information. There are many privacy and security concerns:

- Data Privacy: The phishing detection system that involves sensitive information such as emails, transactions, and browsing history should ensure compliance with the rules of data privacy- GDPR and HIPAA. The solution to data privacy can be achieved by hiding the user data before processing or using homomorphic encryption that will allow its detection without decryption.
- Federated Learning: Federated learning is one such new technique where models can be trained on a decentralized set of datasets without transmitting actual data to a central server. It finds multiple applications in privacy-sensitive domains (for instance, health care and banking), where phishing detection models can now be updated on decentralized sources of data, and all organizations can now collaborate on threat detection without user details being distributed.

5 Conclusion and future scopes

The comprehensive analysis of various methodologies and approaches for phishing detection underscores the significant strides made in the field of cybersecurity. From traditional machine learning techniques to cutting-edge deep learning frameworks, researchers have explored diverse avenues to combat phishing attacks effectively.

The findings reveal that deep learning methods, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrate remarkable performance in detecting phishing attempts, often outperforming traditional approaches. Additionally, ensemble learning techniques and generative adversarial networks (GANs) exhibit promising results, showcasing the importance of leveraging multiple models and adversarial training to enhance detection accuracy and resilience against evolving threats.

Moreover, the study highlights the critical role of feature engineering, dataset creation, and model interpretability in developing robust phishing detection systems. Techniques such as graph-based analysis, natural language processing (NLP), and cryptographic authentica-

tion schemes offer valuable insights into detecting complex phishing schemes and mitigating security risks.

The advancement of phishing attacks in terms of complexity and frequency requires developing powerful, adaptive, and scalable phishing detection systems. Overall, this review of several methodologies- traditional machine learning, deep learning, and GAN-based approaches indicates the competitive advantages and inherent limitations embodied within each technique. They are models like CNN and RNN in deep learning, capable of superior accuracy and effectiveness towards handling complex phishing by hierarchical learning patterns. Models based on GANs like PDGAN have shown promising results improving the model's robustness, especially against adversarial attacks and zero-day phishing threats. But the actual deployment of these high-end models in real-world application is accompanied by challenges including computational, scalability, and adaptation to unfolding new phishing trends.

5.1 Key findings

- Deep learning models including CNN and LSTMs provide high accuracy: they can be even better than 95%. When the detection of phishing is implemented using CNNs, accuracy can go up to 98.74%. In particular, such models are efficient for content and URL analysis of web content as well as phishing patterns in emails, so they are good for deployment in environments in which high detection accuracy is a priority.
- The most important advantage of GAN-based models is that they create more synthetic phishing samples in a way that an ever-increasing change in the phishing strategies cannot degrade the detection systems. Improvement of generalization and robustness is especially attained by PDGAN in simulating phishing scenarios, especially the data imbalances and adversarial attacks.
- Scalability and real-time performance are still significant problems. The computational costs of training and deploying deep learning and GAN-based models in high Volume systems, including systems related to finance or large email service providers, open up the need for optimizing techniques such as compression, parallel processing, and distributed computing frameworks.

5.2 Recommendations for future research

Despite making so many strides in phishing detection, there are still areas open to further exploration. Most of the current challenges and imminent ones are yet to be addressed. The following are some promising directions for future research in the field:

1. Improvement in Model Interpretability and Explainability As phishing detection models become more complex, especially with deep learning and GAN-based architectures, it is a must to improve model interpretability. More advanced XAI frameworks can be explored, which result in interpretable deep learning models, as in the case of the future work of explaining deep learning models while not compromising detection performance. Techniques like saliency maps for CNN and attention mechanisms in RNNs

allow insight into how the models are making their decisions. Naturally, transparency in decision making can be crucial in financial service environments.

2. Federated Learning and Privacy-Preserving Approaches: In phishing detection, which is generally a sensitive area and especially in the areas such as healthcare and financial, Federated learning and differential privacy seem to be at the forefront of the future research directions. They allow training of models over decentralized datasets on the edges while maintaining data privacy. Development of such privacy-preserving phishing detection systems will serve the purpose to satisfy the ethical and legal challenges arising from centralized data processing with model improvement accuracy across different distributed networks.
3. Adversarial Robustness and Defense Mechanisms Despite the crucial necessity to ensure the dependability of phishing detection systems, attacks from adversaries are still posing a menace to some deep learning models. Although GAN-based models, for example, PDGAN, have gained impressive performance in resisting attacks from adversaries, there is still room for the problem to be fully mitigated. Therefore, future work should be put toward enhancing more sophisticated adversarial defense mechanisms that may identify and nullify adversarial examples without overfitting onto synthetic data samples. Hybrid approaches combining ensemble learning, anomaly detection, and adversarial training can prove to be a more elaborate defense strategy.
4. Contextual Awareness and Behavioral Analysis One area of promise which may be explored in the future is bringing in contextual information and analyzing user behavior into phishing detection systems. Current models are generally based on static features like URL's content and metadata, however attacks like phishing very often rely on user behavior as well as environmental factors. Contextual features include browsing behavior, network and social media activity, and could serve for future study in order to better understand phishing detection. Advanced models of context-aware deep learning algorithms that could acknowledge temporal and spatial patterns in user interactions help to detect attacks much better.
5. Dynamic and Adaptive Detection Systems Phishing is a constantly changing tactic. This will necessitate the dynamic updating of detection models. Future studies should investigate the development of adaptive, self-learning, and real-time model retraining phishing detection systems. That would allow the utilization of online learning algorithms and incremental learning frameworks in adapting the capabilities to new phishing attempts without full-time retraining. This would enhance model adaptability to emerging threats but still reduce downtime and resource usage.
6. Cross-Domain Generalization and Transfer Learning Most of the phishing detection models are trained and optimized within specific domains, such as email, websites, or blockchain. However, phishing attacks typically span multiple platforms and modalities, hence frequently showing cross-domain nature. Future work should investigate applications of transfer learning and cross-domain generalization. This will allow a model trained in one domain-for example, email phishing detection-to accurately classify a phishing attack in another domain, such as on social media or mobile applications, with minimal retraining. This will significantly enhance the scalability and versatility of phishing detection systems.
7. Deep Learning and GAN-Based Model Optimization Deep learning and GAN-based models significantly incur computations and are not desirable for resource-constrainted

environments, such as mobile devices or IoT ecosystems. Future research should thus target model optimization via pruning, quantization, and knowledge distillation to reduce the size of the model while avoiding penalties on the detection performance. These will be the most crucial strategies required for scaling phishing detection systems across ecosystems.

5.3 Conclusion

Phishing threats are always evolving in one way or the other, so the detection methodologies also need to evolve with new ideas. Deep learning and GAN-based models have indeed been very good advances regarding phishing detection, though future research work needs to address model interpretability, privacy, adversarial robustness, and scalability challenges. In the context of new avenues including contextual analysis, adaptive learning, and resource-efficient optimization, researchers will be able to design resilient and scalable phishing detection systems that would give them the edge they need to meet the demands of an increasingly complex and interconnected digital landscape. Therefore, further interdisciplinary collaboration and embracing new technologies are crucial for continued improvements in cyberdefense against phishing attacks.

5.4 Future scope

Looking ahead, the future of phishing detection research lies in several promising directions. Firstly, integrating contextual information, user behavior analysis, and network traffic monitoring could further enhance the accuracy and timeliness of phishing detection systems. Additionally, exploring novel techniques such as federated learning, differential privacy, and explainable AI could address concerns related to data privacy, model transparency, and adversarial robustness.

Furthermore, expanding the scope of evaluation to encompass real-world datasets, diverse attack scenarios, and dynamic environments will provide a more comprehensive understanding of detection model performance and generalizability. Collaborative efforts between academia, industry, and regulatory bodies are essential to foster the development of standardized benchmarks, evaluation protocols, and best practices for phishing detection research.

In conclusion, the ongoing evolution of phishing threats necessitates continuous innovation and collaboration in the cybersecurity community. By leveraging advanced methodologies, interdisciplinary approaches, and rigorous evaluation frameworks, researchers can better defend against phishing attacks and safeguard digital ecosystems in an increasingly interconnected world.

Author contributions Kavya S—Manuscript Writing. Sumathi D—Guide (Checking the flow).

Funding Open access funding provided by Vellore Institute of Technology.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Zieni R, Massari L, Calzarossa MC (2023) Phishing or not phishing? A survey on the detection of phishing websites. IEEE Access 11: 18499–18519. https://doi.org/10.1109/ACCESS.2023.3247135
2. Luo J, Qin J, Wang R, Li L (2024) A phishing account detection model via network embedding for ethereum. IEEE Trans Circ Syst II Express Briefs 71(2): 622–626. https://doi.org/10.1109/TCSII.2023.3267822
3. Sahingoz OK, BUBEr E, Kugu E (2024) DEPHIDES: deep learning based phishing detection system. IEEE Access 12: 8052–8070. https://doi.org/10.1109/ACCESS.2024.3352629
4. Purwanto RW, Pal A, Blair A, Jha S (2022) PhishSim: aiding phishing website detection with a feature-free tool. IEEE Trans Inf Forens Secur 17:1497–1512. https://doi.org/10.1109/TIFS.2022.3164212
5. Al-Ahmadi S, Alotaibi A, Alsaleh O (2022) Phishing detection with generative adversarial networks. IEEE Access. 10:42459–42468. https://doi.org/10.1109/ACCESS.2022.3168235
6. Wen H, Fang J, Wu J, Zheng Z (2023) Hide and seek: an adversarial hiding approach against phishing detection on ethereum. IEEE Trans Comput Soc Syst 10(6): 3512–3523. https://doi.org/10.1109/TCSS.2022.3203081
7. Pillai MJ, Remya S, Devika V, Ramasubbareddy S, Cho Y (2024) Evasion attacks and defense mechanisms for machine learning-based web phishing classifiers. IEEE Access 12: 19375–19387. https://doi.org/10.1109/ACCESS.2023.3342840
8. Kalabarige LR, Rao RS, Abraham A, Gabralla LA (2022) Multilayer stacked ensemble learning model to detect phishing websites. IEEE Access 10: 79543–79552. https://doi.org/10.1109/ACCESS.2022.3194672
9. Kabla AHH, Anbar M, Manickam S, Karupayah S (2022) Eth-PSD: a machine learning-based phishing scam detection approach in ethereum. IEEE Access 10: 118043–118057. https://doi.org/10.1109/ACCESS.2022.3220780
10. Castaño F, Fernañdez EF, Alaiz-Rodríguez R, Alegre E (2023) PhiKitA: phishing kit attacks dataset for phishing websites identification. IEEE Access 11: 40779–40789. https://doi.org/10.1109/ACCESS.2023.3268027
11. Salloum S, Gaber T, Vadera S, Shaalan K (2022) A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access 10:65703–65727. https://doi.org/10.1109/ACCESS.2022.3183083
12. Wu J et al (2022) Who are the phishers? Phishing scam detection on ethereum via network embedding. IEEE Trans Syst Man Cybern Syst 52(2): 1156–1166. https://doi.org/10.1109/TSMC.2020.3016821
13. Sánchez-Paniagua M, Fernández EF, Alegre E, Al-Nabki W, González-Castro V (2022) Phishing URL detection: a real-case scenario through login URLs. IEEE Access. 10:42949–42960. https://doi.org/10.1109/ACCESS.2022.3168681
14. Li Q, Cheng M, Wang J, Sun B (2022) LSTM based phishing detection for big email data. IEEE Trans Big Data. 8(1):278–288. https://doi.org/10.1109/TBDATA.2020.2978915
15. Alsubaei FS, Almazroi AA, Ayub N (2024) Enhancing phishing detection: a novel hybrid deep learning framework for cybercrime forensics. EEE Access 12:8373–8389. https://doi.org/10.1109/ACCESS.2024.3351946
16. Liu J, Chen J, Wu J, Wu Z, Fang J, Zheng Z (2024) Fishing for fraudsters: uncovering ethereum phishing gangs with blockchain data. IEEE Trans Inf Forensics Secur 19:3038–3050. https://doi.org/10.1109/TIFS.2024.3359000

17. Li W, Manickam S, Laghari SUA, Chong Y-W (2023) Uncovering the cloak: a systematic review of techniques used to conceal phishing websites. IEEE Access 11: 71925–71939. https://doi.org/10.1109/ACCESS.2023.3293063
18. Zhu E, Chen Z, Cui J, Zhong H (2022) MOE/RF: a novel phishing detection model based on revised multiobjective evolution optimization algorithm and random forest. IEEE Trans Netw Service Manag 19(4): 4461–4478. https://doi.org/10.1109/TNSM.2022.3162885
19. Chen S, Fan L, Chen C, Xue M, Liu Y, Xu L (2021) GUI-Squatting attack: automated generation of android phishing apps. IEEE Trans Depend Secur Comput 18(6):2551–2568. https://doi.org/10.1109/TDSC.2019.2956035
20. Wei Y, Sekiya Y (2022) Sufficiency of ensemble machine learning methods for phishing websites detection. IEEE Access 10:124103–124113. https://doi.org/10.1109/ACCESS.2022.3224781
21. Baki S, Verma RM (2023) Sixteen years of Phishing user studies: what have we learned? IEEE Trans Depend Secure Comput 20(2):1200–1212. https://doi.org/10.1109/TDSC.2022.3151103
22. Tang L, Mahmoud QH (2022) A deep learning-based framework for phishing website detection. IEEE Access 10:1509–1521. https://doi.org/10.1109/ACCESS.2021.3137636
23. Shirazi H, Muramudalige SR, Ray I, Jayasumana AP, Wang H (2023) Adversarial autoencoder data synthesis for enhancing machine learning-based phishing detection algorithms. IEEE Trans Serv Comput 16(4): 2411–2422. https://doi.org/10.1109/TSC.2023.3234806
24. He D, Lv X, Zhu S, Chan S, Choo K-KR (2024) A method for detecting phishing websites based on tiny-bert stacking. IEEE Internet Things J 11(2):2236–2243. https://doi.org/10.1109/JIOT.2023.3292171
25. Ariyadasa S, Fernando S, Fernando S (2022) Combining long-term recurrent convolutional and graph convolutional networks to detect phishing sites using URL and HTML. IEEE Access. 10:82355–82375. https://doi.org/10.1109/ACCESS.2022.3196018
26. Xia Y, Liu J, Wu J (2022) Phishing detection on ethereum via attributed ego-graph embedding. IEEE Trans Circ Syst II Express Briefs 69(5): 2538–2542. https://doi.org/10.1109/TCSII.2022.3159594
27. Valecha R, Mandaokar P, Rao HR (2022) Phishing email detection using persuasion cues. IEEE Trans Depend Secure Comput 19(2): 747–756. https://doi.org/10.1109/TDSC.2021.3118931
28. Apruzzese G, Subrahmanian VS (2023) Mitigating adversarial gray-box attacks against phishing detectors. IEEE Trans Depend Secure Comput 20(5):3753–3769. https://doi.org/10.1109/TDSC.2022.3210029
29. Bhattacharya M, Roy S, Chattopadhyay S, Das AK, Jamal SS (2023) An efficient user authentication scheme for phishing attack detection in mobile online social networks. IEEE Syst J 17(1):234–245. https://doi.org/10.1109/JSYST.2022.3168234
30. Almousa M, Anwar M (2023) A URL-based social semantic attacks detection with character-aware language model. IEEE Access 11: 10654–10663. https://doi.org/10.1109/ACCESS.2023.3241121
31. Kara I, Ok M, Ozaday A (2022) Characteristics of understanding URLs and domain names features: the detection of Phishing websites with Machine Learning methods. IEEE Access 10:124420–124428. https://doi.org/10.1109/ACCESS.2022.3223111
32. Kalabarige LR, Rao RS, Pais AR, Gabralla LA (2023) A boosting-based hybrid feature selection and multi-layer stacked ensemble learning model to detect phishing websites. IEEE Access 11:71180–71193. https://doi.org/10.1109/ACCESS.2023.3293649
33. Abdillah R, Shukur Z, Mohd M, Murah TMZ (2022) Phishing classification techniques: a systematic literature review. IEEE Access 10: 41574–41591. https://doi.org/10.1109/ACCESS.2022.3166474
34. Azam Z, Islam MM, Huda MN (2023) Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. IEEE Access 11: 80348–80391. https://doi.org/10.1109/ACCESS.2023.3296444
35. Al-Fawa'reh M, Abu-Khalaf J, Szewczyk P, Kang JJ (2024) MalBoT-DRL: malware botnet detection using deep reinforcement learning in IoT networks. IEEE Internet Things J 11(6):9610–9629. https://doi.org/10.1109/JIOT.2023.3324053
36. Li R, Liu Z, Ma Y, Yang D, Sun S (June 2023) Internet financial fraud detection based on graph learning. IEEE Trans Comput Social Syst 10(3):1394–1401. https://doi.org/10.1109/TCSS.2022.3189368
37. Karim A, Shahroz M, Mustofa K, Belhaouari SB, Joga SRK (2023) Phishing detection system through hybrid machine learning based on URL. IEEE Access 11: 36805–36822. https://doi.org/10.1109/ACCESS.2023.3252366
38. Sanjalawe YK, Al-E'mari SR (2023) Abnormal transactions detection in the ethereum network using semi-supervised generative adversarial networks. IEEE Access 11:98516–98531. https://doi.org/10.1109/ACCESS.2023.3313630
39. Chai Y, Zhou Y, Li W, Jiang Y (2022) An explainable multi-modal hierarchical attention model for developing phishing threat intelligence. IEEE Trans Depend Secure Comput 19(2): 790–803. https://doi.org/10.1109/TDSC.2021.3119323

40. Mehnaz S, Bertino E (2021) A fine-grained approach for anomaly detection in file system accesses with enhanced temporal user profiles. IEEE Trans Depend Secur Comput 18:2535–2550. https://doi.org/10.1109/TDSC.2019.2954507
41. Jibat D, Jamjoom S, Al-Haija QA, Qusef A (2023) A systematic review: detecting phishing websites using data mining models. Intell Converged Netw 4(4): 326–341. https://doi.org/10.23919/ICN.2023.0027
42. Mahbub S, Pardede E, Kayes ASM (2022) Online recruitment fraud detection: a study on contextual features in Australian Job industries. IEEE Access 10:82776–82787. https://doi.org/10.1109/ACCESS.2022.3197225
43. Razaque A et al (2022) Blockchain-enabled deep recurrent neural network model for clickbait detection. IEEE Access 10:3144–3163. https://doi.org/10.1109/ACCESS.2021.3137078
44. Rafsanjani AS, Kamaruddin NB, Rusli HM, Dabbagh M (2023) QsecR: secure QR Code scanner according to a novel malicious URL detection framework. IEEE Access 11:92523–92539. https://doi.org/10.1109/ACCESS.2023.3291811
45. Zhao J, Masood R, Seneviratne S (2021) A review of computer vision methods in network security. IEEE Commun Surv Tutor 23(3): 1838–1878. https://doi.org/10.1109/COMST.2021.3086475
46. Indrasiri PL, Halgamuge MN, Mohammad A (2021) Robust ensemble machine learning model for filtering phishing URLs: expandable random gradient stacked voting classifier (ERG-SVC). IEEE Access 9:150142–150161. https://doi.org/10.1109/ACCESS.2021.3124628
47. Raghunath KMK, Kumar VV, Venkatesan M, Singh KK, Mahesh TR, Singh A (June 2022) XGBoost Regression Classifier (XRC) model for cyber attack detection and classification using inception V4. J Web Eng 21(4):1295–1322. https://doi.org/10.13052/jwe1540-9589.21413
48. Nunes YTP, Guedes LA (2024) Concept drift detection based on typicality and eccentricity. IEEE Access 12: 13795–13808. https://doi.org/10.1109/ACCESS.2024.3355959
49. Raza A, Munir K, Almutairi MS, Sehar R (2023) Novel class probability features for optimizing network attack detection with machine learning. IEEE Access 11:98685–98694. https://doi.org/10.1109/ACCESS.2023.3313596
50. Jovanovic L et al (2023) Improving phishing website detection using a hybrid two-level framework for feature selection and XGBoost tuning. J Web Eng 22(3):543–574. https://doi.org/10.13052/jwe1540-9589.2237

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.