IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 May 2025, accepted 4 June 2025, date of publication 13 June 2025, date of current version 20 June 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3579584

SURVEY

Multimodal Phishing Detection on Social Networking Sites: A Systematic Review

TANDIN WANGCHUK AND TAD GONSALVES, (Member, IEEE)
Department of Information and Communication Science, Faculty of Science and Technology, Sophia University, Chiyoda-ku, Tokyo 102-8554, Japan
Corresponding author: Tandin Wangchuk (tandin@eagle.sophia.ac.jp)

ABSTRACT Phishing is one of the most common cyberattacks, with the number of incidents increasing annually. Significant research interests have been generated in phishing emails, URLs, and websites over the past decade. Phishing attackers often target easy prey, and phishing on social networking sites (SNS) is increasing due to the popularity of easier communication using short text, images, and voice. Multimodal features can be used for phishing, yet there is relatively little research on phishing on SNS using multimodal features and models. This systematic review report, guided by the PRISMA and SEGRESS, aimed to uncover the techniques used for multimodal phishing detection. A comprehensive literature search in Scopus, Web of Science, IEEE Xplore, and ACM Digital Library was conducted, including studies published in English from 2018 to 2025. A total of 20 studies are included in the study out of 74 records returned. Twelve articles were included from the references of the selected studies. The study reveals that HTML content, URLs, and visual elements are used as multimodal features to classify phishing URLs and websites. These studies used deep learning models, including CNN, RNN, LSTM, and MLP, which produced promising results. However, challenges persist related to resource constraints, adversarial attacks, data quality and availability, false positives and negatives, and integration with existing SNS frameworks. These multimodal studies show promise, but require adaptation to SNS-based phishing attacks.

INDEX TERMS Machine learning, multimodal phishing, phishing detection, smishing, systematic review.

I. INTRODUCTION

Phishing scams are becoming more frequent despite extensive research being conducted to deter them. Scammers find it too easy to initiate contact with unsuspecting targets and trick them into relinquishing sensitive information. By exploiting the manipulation of human behavior known as ''social engineering'' [1], [2], people are forced to act in ways to ''leak'' information for financial gain. The most recognized form of social engineering is phishing through emails, uniform resource locators (URLs), and websites [3]. Some less widespread forms of phishing are gaining traction, such as SMS Phishing (Smishing) [4] and Voice phishing (Vishing) [5].

Phishing was the most common type of cyberattack in 2024. The Anti-Phishing Working Group (APWG) Phishing Activity Trends Report [1] noted that by the third quarter (Q3) of 2024, there were more phishing attacks than in the previous quarter. Likewise, the Information Systems Audit and Control Association (ISACA) has also reported in the State of Cybersecurity 2024 [6] that phishing, utilizing sophisticated social engineering methods, is the most common type of cyberattack. In the last decade, there has been increasing focus on studying phishing, especially phishing email and website detection, due to the possibility of new technologies in AI - Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP). On the flip side, new attack vectors are generated, increasing the attack surfaces.

With the increase in countermeasures against phishing emails and websites, attackers have switched to phishing using short messages and voice [7]. However, there is insufficient protection against these issues. Many phishing attacks are conducted through social networking services (SNS), and few studies have focused on combating these scams. Understanding this type of change in attack vectors is important so that countermeasures can be developed to

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

protect SNS users against these new emerging phishing threats.

Several review studies have been conducted on phishing detection over the last five years. Ige et al. [8] carried out a review of current machine learning and deep learning techniques for phishing detection by categorizing them into Bayesian, non-Bayesian, and deep learning methods. An empirical analysis was conducted to evaluate the performance of various classification methods and identify weaknesses in existing methods. The study focused primarily on phishing URLs and categorized URL features as controllable and uncontrollable. The two-stage prediction model was proposed to improve the performance of the underperforming algorithms in the model. In the paper, there is slight confusion on what controllable and uncontrollable features are, since the proposed solution uses one of these categories of features with web scraping. Kyaw et al. [9]'s systematic review of deep learning techniques focuses on detecting phishing emails. The paper provides a taxonomy of deep learning-based phishing detection methods, their limitations, and their effectiveness.

A comprehensive review [10] was conducted on state-of-the-art phishing detection methodologies, encompassing traditional machine learning techniques, deep learning frameworks, and advanced methods like GANs and network embedding techniques. Each technique is examined for its rationale, advantages, efficacy of the algorithms, and challenges. The authors also provide limitations and future research recommendations. Phishing in this review is not specific to phishing emails, URLs, or websites, but is discussed in general. The article is quite lengthy, and some of the points are repetitive. Another comprehensive review of phishing conducted by Ahmed et al. [11] mainly focused on phishing websites. They also covered machine learning and deep learning techniques, along with phishing datasets. The article also highlights the limitations, weaknesses, potential improvements, and a roadmap for future studies on phishing websites. None of these studies covers the multimodal features, multimodal models, or phishing on SNS.

Our objective with this research was to review how much phishing scams in a multi-modal form have been addressed using machine learning. The emphasis was on the detection of phishing scams employing multi-modal approaches. The following research questions were expected to be tackled by our study:

- **RQ1**: What are the existing techniques for detecting phishing scams in text, audio, and images?
- **RQ2:** How effective are these techniques in real-world scenarios?
- **RQ3**: What are the gaps and challenges in multimodal phishing scam detection?

The remaining part of this paper is structured as follows: Section II covers literature review, Section III presents the methodology employed in the study, Section IV Results, Section V Discussion, Section VI Limitations, and Section VII Conclusion.

II. LITERATURE REVIEW

With the increase in phishing attack instances, phishing detection has become crucial for protecting unsuspecting users from deceptive attempts to steal sensitive information. Understanding common phishing detection techniques is essential. Traditional detection techniques include list-based and content-based filtering, while contemporary detection techniques are based on machine learning, deep neural networks, and graph convolutional networks. List-based techniques involve blacklisting known phishing URLs and checking whether a URL matches any records in the block list [12]. This technique is widely used in popular web browsers such as Chrome, Edge, and Firefox [12], [13]. The downside of this technique is that the block list requires frequent updates to include newly created phishing URLs. This means that, this technique cannot detect phishing attacks that are not in the blacklists. One of the content-based filtering approaches is the visual similarity approach, which evaluates suspicious and authentic websites based on numerous visual characteristics [14]. Like the list-based approach, this technique cannot detect zero-day phishing attacks since it assesses against previously visited or saved sites.

Approaches based on machine learning, deep neural networks, and graph convolutional networks are found to be promising in addressing the above shortcomings. Generally, there are different types of phishing: phishing through emails, URLs, websites, SMS, and voice/video. Our preliminary scoping study found that phishing emails, phishing URLs, and phishing websites are the most studied approaches, which is corroborated by the systematic review [15]. Phishing through SMS and voice is gaining traction. Recently, social networking services (SNS) have emerged as the new phishing attack vector where phishing can be carried out using text, image, or voice [7], [16], [17]. The evolving nature of phishing attacks underscores the need for continuous improvement in detection techniques to protect users effectively.

The scoping study also reveals that web-based or URL-based phishing is the most researched category of phishing attacks. Among 430 journal articles retrieved from Scopus using the query string (''Phishing Detection'') published in English between 2018 and 2025 (inclusive), 246 (57%) focus on web-based or URL-based phishing. In contrast, only 52 (12%) articles address email-based phishing, while the remainder include SMS-based phishing (6 articles), voice-based phishing (6 articles), review articles (15 articles) and miscellaneous articles (105 articles). To combat phishing URLs or websites, a range of innovative solutions have been proposed using machine learning or deep learning. Traditional machine learning methods [18], [19], [20], [21] to more sophisticated machine learning [22], [23], [24], [25], [26] and deep learning approaches [27], [28], [29], [30], [31] have been suggested. Machine learning methods require fewer resources but hinge on careful feature selection, which is labor intensive. Unnecessary or noisy features can lead to overfitting and a decline in the accuracy of models. Deep learning approaches, while mitigating this

challenge [32], [33] are resource intensive during training and require high-quality and substantive datasets. Alternative strategies, such as feature selection methods [24], [34], [35], [36], [37], aim to remove noisy features to improve model accuracy but suffer from scalability problems. However, for phishing email detection, deep learning coupled with Natural Language Processing (NLP) performs better than machine learning methods. In the study by Melendez et al. [38] reported RoBERTA (99.43% accuracy), XLNET (98.84% accuracy), and BERT (99.11% accuracy) performed better than traditional machine learning models for phishing email detection. Most of these solutions are based on URL features, Document Object Model (DOM), markup language, JavaScript, textual content, headers, or visual representations.

III. REVIEW METHODS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework [39] and Software Engineering Guidelines for Reporting Secondary Studies (SEGRESS) [40] guided this systematic review report. The following shows the detailed procedures and resources used for the study.

A. ELIGIBILITY CRITERIA

The review aims to identify the existing machine-learning techniques for detecting text, audio, and image-based phishing scams, which are tackled with multimodal approaches. The information is to be obtained from several relevant and popular academic databases. Therefore, the following inclusion criteria are identified:

- The primary studies aim to discuss or investigate phishing attacks perpetrated online through text, image, or audio, alongside interventions using machine learning, deep learning, or multimodal analysis techniques.
- Techniques explored should include processing texts (both web content and URLs) and images, specifically phishing based on emails or URLs.
- Studies should also focus on smishing or phishing.
- Studies must be published in English and should be published from 2018 (inclusive) to 2025.

The identified exclusion criteria are:

- Primary studies reporting blockchain-, metaverse, or cryptocurrency-based phishing attacks that are not conducted through email, SMS, voice, or messaging platforms.
- Cybersecurity studies that are not relevant to phishing, such as malware detection or anomaly detection.
- Papers that do not have full-text access.
- Articles that are available only as a book section without abstracts or full paper access.
- Articles that are published as conference proceedings, and studies that were published before 2018.

B. INFORMATION SOURCES

We decided to limit our search to Scopus, Web of Science, IEEE Xplore, and ACM Digital Library, given the numerous publications available during our preliminary scoping of sources. Moreover, we required articles published in high-quality journals to address our research questions. These academic databases are renowned for indexing high-impact research in computer science. The search on Scopus was conducted through the advanced search interface using the search string in the document title, abstract, and keywords. Similarly, the advanced search interface was utilized for Web of Science, the ACM Digital Library on Computing Literature, and Command Search on IEEE Xplore. We also searched for related articles cited in those included articles within these databases and on Google Scholar. The database searches were conducted between February 9 and 10. Any full texts that were not accessible were requested from the respective authors through ResearchGate.

C. SEARCH STRATEGY

Our search strategy identified relevant literature by tailoring searches to four academic databases. The searches were conducted in Scopus, Web of Science, IEEE Xplore, and ACM Digital Library. The following keywords were identified to construct search query strings for the respective databases: ''artificial intelligence,'' ''machine learning,'' ''deep learning,'' ''phishing,'' ''multimodal,'' and ''social engineering.'' These keywords are derived from the research idea and questions. The search strings are developed from the keywords and their synonyms, such as ''AI,'' ''ML,'' and ''DL,'' using Boolean expressions. Search filters were applied to some databases to maintain consistency in the search according to the inclusion and exclusion criteria, as indicated in Table 1. Searches were limited to 2018-2025 published journal articles in English to screen out duplicates and low-quality articles, given the numerous publications over the last decade on phishing. Non-English articles were excluded due to the lack of expertise of the authors. Additional articles were found through the references of the selected articles.

The searches were conducted between February 9 and 10, 2025. The selection of search dates was random and has no significance. Each database supports its search queries. Customized search queries must be constructed based on the features supported by each database. For example, Scopus supports searches for document titles, abstracts, and keywords simultaneously, which is best practice [41]. IEEE Xplore, ACM Digital Library, and Web of Science allow searching in document title, abstract, and keyword at a time. The most precise search queries possible were constructed. Table 1 shows the respective search query strings and the results returned.

We restricted our search to exclude conference proceedings, publications before 2018, theses, books or book sections, reviews, and articles in the press. Only peer-reviewed journal publications published in English from 2018 to 2025 were included in the search. Over the last decade, extensive research in social engineering, especially on phishing, has been conducted. Moreover, the accelerated advancement in Artificial Intelligence has resulted in numer-

TABLE 1. Search Strategies for each academic database.

Database	Search Query	Date	Result
Scopus	TITLE-ABS-KEY ((“artificial intelligence" OR "machine learning" OR "Deep Learning" OR ai OR ml) AND (“phishing" OR "Social Engineering”) AND "multimodal”) AND (LIMIT-TO (LANGUAGE, "English”))	09/02/2025	28
Web of Science	AB= (("phishing" or "social engineering") AND "multimodal”) AND AB= (“artificial intelligence" OR "machine learning" OR "Deep learning" OR ai OR ml OR dl)	09/02/2025	7
IEEE Xplore	ALL (("phishing" OR "social engineering”) AND "multimodal" AND (ai OR "Artificial Intelligence" OR ml OR "machine learning" OR dl OR "deep learning"))	09/02/2025	3
ACM Digital Library	"query”: Fulltext:(("phishing" or "social engineering") AND "multimodal" AND (ai OR "Artificial Intelligence" OR ml OR "machine learning" OR dl OR "deep learning")) "filter": Article Type: Research Article	09/02/2025	36

ous publications in a short time. Only journals are included to maintain the quality of our study and avoid including redundant publications.

D. SELECTION PROCESS

We aimed to find, assess, and synthesize all peer-reviewed journals related to addressing phishing scams perpetrated online through text, image, and voice. Any research on phishing dealing with text content, text image, and voice uses the following interventions: machine learning, deep learning, or multimodal analysis techniques to detect phishing are included. Techniques explored include processing of text (both web contents and URLS), image, and audio inputs. Since an email body can contain text and voices, publications on email phishing and phishing URLs are essentially included. Publications related to smishing and vishing are also included. A major chunk of the search results returned contains publications on phishing. Publications of text messages and voice-based scams are relatively limited.

We excluded publications related to blockchain, cryptocurrency, and metaverse scams that are not carried out through email, SMS, voice, or messaging platforms. Scopus returned 28 search results, Web of Science 7, IEEE Xplore 3, and ACM Digital Library 36, with the respective query strings. A total of 74 records were extracted and consolidated for preliminary screening. The searches included journal publications between 2018 and 2025.

Search results were screened for eligibility by a single author. In the first stage, a total of 74 consolidated records are extracted and downloaded into reference management software, Zotero, and exported to Rayyan.ai. With the help of Rayyan.ai [42], eight duplicates are detected and resolved, of which four duplicate records are removed before further screening. In the second stage, titles and abstracts of the remaining articles are thoroughly examined to ensure the quality and relevance of the literature for our study. Based on the predefined inclusion and exclusion criteria, records are screened into three categories on Rayyan.ai: ‘Excluded’, ‘Maybe’ and ‘Included’. 40 records were categorized into Excluded, 10 into Maybe, and 20 into Included. Those that were categorized as Excluded and Maybe are reviewed twice and finally excluded from the study. Figure 1 shows the PRISMA flow diagram for the selection process.

Fifty out of seventy records were screened out, including information about books or book sections without articles, conference proceedings, review articles, and irrelevant publications that may have been picked up during searches related to standalone keywords but are completely unrelated to our study. Similarly, retracted papers and student papers were excluded. In stage three, twenty records were sought for full text screening, of which two were initially excluded because the full text was not accessible, but were later retrieved. Consequently, this study included twenty relevant articles and twelve articles from the references of the selected records, all of which are peer-reviewed articles published in English between 2018 and 2025. A list of selected articles is given in the appendices.

TABLE 2. Standard definition of evaluation metrics used in machine learning.

SL. No.	Metrics	Definitions
1	Accuracy	Ratio of correctly classified instances to the total number of instances (Number of correct instances) / (total number of instances)
2	Precision	Ratio of correctly classified positive instances to the total predicted positive instances. (True Positives) / (True Positives + False Positives)
3	Recall	Ratio of correctly classified positive instances to the total actual positive instances. (True Positives) / (True Positives + False Negatives)
4	F1 Score	Harmonic mean of precision and recall. 2 (Precision * Recall) / (Precision + Recall)
5	ROC AUC	The area under the ROC Curve is known as AUC, which characterizes the performance of a classification model across different threshold settings.

E. DATA COLLECTION

The next stage involves collecting and extracting data from the selected articles. A uniform spreadsheet template is used to extract study characteristics, methods, results or findings, and limitations. The extracted data include the title of each article, authors, publication year, summary of the article, methodology, results, discussion, and limitations. In methodology, the authors are interested in the machine learning techniques used and in the results the efficacy of these techniques for phishing detection in terms of accuracy, precision, recall, F1 score, and AUC-ROC. The definitions of these metrics are provided in Table 2. An author reviewed the selected studies, extracted information, and summarized the articles. The summaries of the articles were later validated against the summaries generated by AI tools, such as Scispace [43]. This aims to address the limitation of manually performing these processes and improve the reliability of the extracted data.

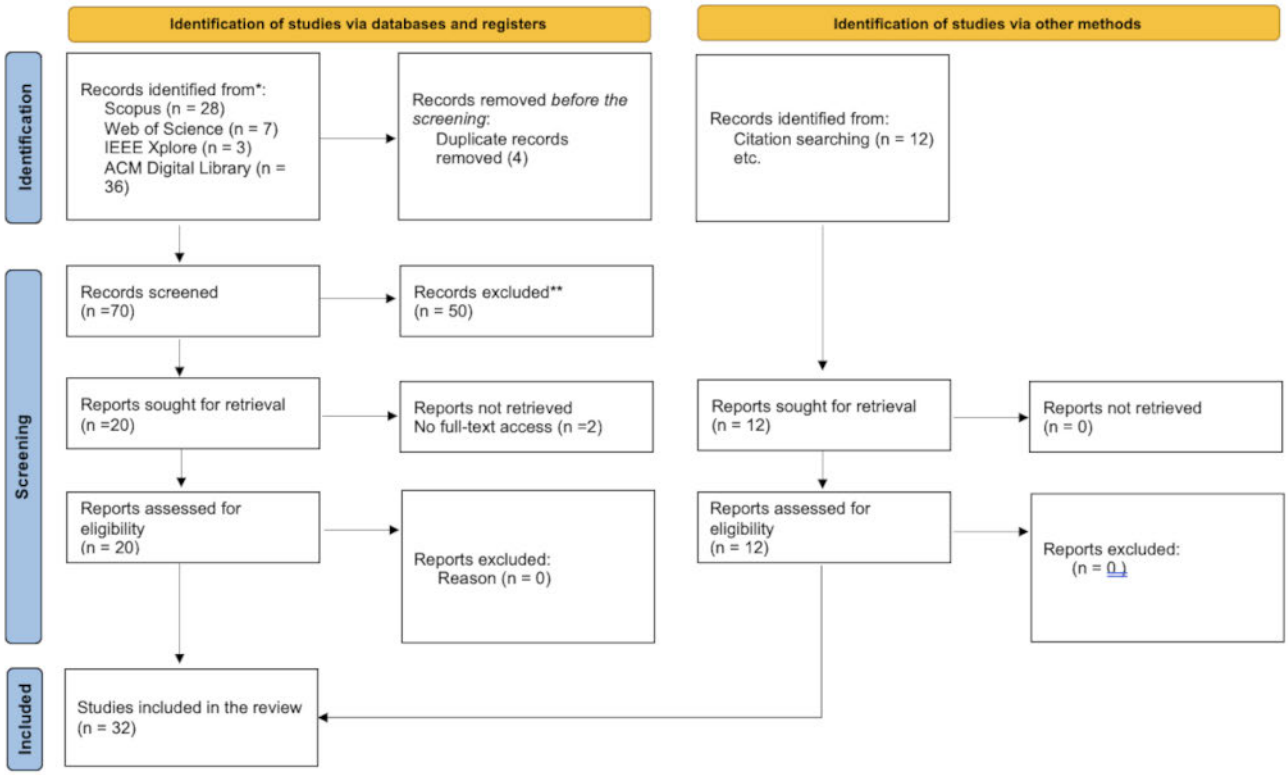

FIGURE 1. PRISMA Flow diagram showing the number of records selected at each stage.

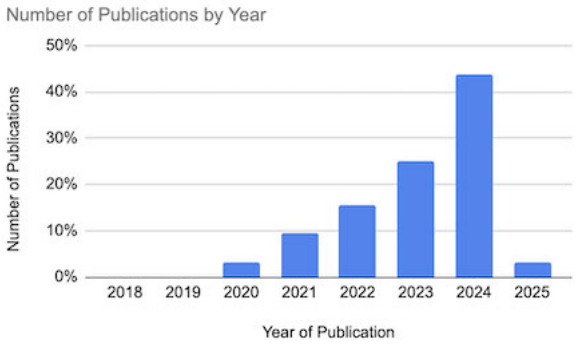

FIGURE 2. Year-wise distribution of selected publications.

IV. RESULTS

In this study, the search strategy resulted in 74 records, which were then exported to Zotero and Rayyan.ai for screening. After removing four duplicates, 70 records were screened as 'Included', 'Maybe', or 'Excluded' based on titles and abstracts, which resulted in the exclusion of 50 records and the inclusion of 20 records for retrieval. The full texts of the 20 records were retrieved to assess eligibility, and all were included in the study after retrieving the missing full text of two records. Twelve more were included after accessing the references of these 20 records. Consequently, 32 records were included in the study.

The increase in studies on phishing supported by the advent of machine learning algorithms is evident in the highest number of publications (44%) in 2024 among the total articles from 2018 to 2025 included in this study (as shown in Fig. 2). Numerous machine learning countermeasures have been studied and proposed to combat these phishing attacks [44], [45], [46], [47], [48]. Recently introduced multimodal models have garnered attention from researchers over the past few years, yet they remain underexplored in the context of phishing attack detection.

Multimodal approaches to phishing detection utilize various input types and machine learning methods to enhance accuracy and robustness. Multiple features, such as HTML, URLs, images, and visual elements [49], [50], [51], are employed to create a comprehensive detection framework. Limitations of single-modal models can effectively be addressed by multimodal approaches. These approaches strengthen resilience against sophisticated phishing attacks. With this, the authors tried to address the research questions from the synthesis of the data.

A. RQ1: WHAT ARE THE EXISTING TECHNIQUES FOR DETECTING PHISHING SCAMS IN TEXT, AUDIO, AND IMAGES?

PhishAgent, Shark-Eyes, and multimodal hierarchical attention model (MMHAM) approaches made use of HTML and URL features to detect phishing websites. PhishAgent [52] utilized multimodal large language models to integrate multimodal information retrieval to enhance brand recognition and recall. Shark-Eyes [53] combined domain

and HTML tag features for robust detection. MMHAM [49] automatically extracted deep representations of fraud cues from web pages, URLs, and visual contents using a shared dictionary learning approach to align representations with different modalities. A combination of text and visual features can reduce misclassifications and enhance early phishing detection, according to van Dooremaal et al. [54].

Many studies employed deep learning approaches like CNN, RNN, and MLP-based to handle multimodal input features. For example, a multi-feature neural network model [55] used CNN for image features, RNN for text features, and machine learning techniques for classification to enhance detection accuracy. Similarly, the Multi-Modal Deep Learning method [51] used a combination of three deep learning methods (CNN-LSTM, FastText-BiLSTM, and BERT-MLP) to identify relevant features. Chai et al. [49] used LSTM for URL string encoding at the character level and webpage text encoding at the word level to learn fraud cues. FusionNet [50] made use of Bi-LSTM to extract URL features, RNN-attention to extract HTML features, and ResNET34-attention to extract features from screenshots, which were fused and fed into a classifier to enhance phishing website detection. Nabila et al. [56] reported that the use of transfer learning and ensemble methods, such as LSTM and GRU, further enhances detection capabilities by processing both textual and image features.

B. RQ2: HOW EFFECTIVE ARE THESE TECHNIQUES IN REAL-WORLD SCENARIOS?

These approaches show promising results. Duy et al. [57] demonstrated outstanding resistance of the Shark-Eyes, multimodal-based method against adversarial examples with a detection rate of up to 99%. MMHAM achieved the best detection metrics, but also offered hierarchical interpretability of each modality, highlighting the significance of individual elements within each modality [49]. The use of visuals with textual features allowed van Dooremaal and his team [54] to achieve their approach of 99.66% accuracy for phishing classification on the dataset they had developed. Detailed comparative studies are given in Table 3.

C. RQ3: WHAT ARE THE GAPS AND CHALLENGES IN MULTIMODAL PHISHING SCAM DETECTION?

Positive and bright futures with Multimodal Models for phishing detection are not without challenges and limitations. The quality and availability of data can pose significant challenges. For example, phishing attacks are carried out not only through text-based content but also through images. Phishing websites often use images instead of text to evade detection and make it difficult for systems trained on text-based features [66]. Likewise, data collection and model training are complicated by the dynamic nature of SNS content, which is frequently updated and short-lived. Building systems with multiple modalities often leads to complex models that are difficult to scale. For instance, in [67], a system was proposed that combines Graph Convolution Networks (GCN) with Transformers, which achieved high accuracy but might struggle with scalability due to the large size of datasets and the requirement of high computational resources. Similarly, hybrid models that incorporate both textual and visual features require additional computational overhead for feature extraction and fusion [68].

The implementation of multimodal systems on social networking services requires significant computational and infrastructural resources. For example, systems like PhishAgent, which combine extensive language models with offline knowledge repositories, can be resource-intensive and difficult to deploy in environments with limited resources [52]. Similarly, advanced deep learning techniques, namely Graph Convolutional Networks (GCNs) and Transformers, are so powerful that they may not be significantly available in some settings due to the limitation of resources [67].

Multimodal systems are also not immune to adversarial attacks, as phishing attacks continually evolve to deceive detection systems. Strategies including domain rotation, logo alteration, and employment of visual representations in place of textual content may render traditional features ineffective [66]. Adversarial attacks directed at particular modalities, including logos, have the potential to compromise the efficacy of system performance [68]. Given the rapid dissemination of malicious links and posts, social networking service (SNS) platforms require real-time detection mechanisms to thwart phishing attacks. Nonetheless, numerous multimodal systems, especially those entailing intricate feature extraction and fusion processes, may not be optimized for real-time operations. For example, systems dependent on HTML DOM graph modeling can incur latency due to the requirement to parse and analyze these graphs for each webpage [67]. Addressing these challenges is crucial for ensuring the robustness, reliability, and real-time effectiveness of multimodal phishing detection systems on SNS platforms.

Integration of multimodal systems with pre-existing social networking services (SNS) is always a daunting task, and it comes with a lot of hindrances. The PhishAgent system [52], for instance, may need a major overhaul of the entire system architecture to facilitate the merger between physical knowledge repositories and digital ones. Moreover, the smoothing of the interaction among diverse elements by introducing Support Vector Machine (SVM)-based classifiers and the dimensionality reduction driven by principal component analysis (PCA) may be quite a difficult and overlong process [69]. These problems underpin the challenges of advanced multimodal systems stakes insertion into the extant SNS frameworks. In this regard, the successful deployment and operation of multimodal phishing detection systems on SNS platforms would be a matter of resolving the integration problems.

Like other systems, multimodal systems deal with the challenge of having false negatives and false positives. When a legitimate message is flagged as phishing, this is referred

TABLE 3. Comparative effectiveness of systems.

System/Model	Key features	Result
AWG [57]	Proposes Adversarial Website Generation to enhance model training and testing.	The Generator exceeds 90% domain structure generation rate. The MM Shark-Eyes model shows a detection rate of up to 99%.
FusionNet [50]	Uses specific representation learning architectures for each modality (URLs, HTML source codes, and visual features) and an attention mechanism to merge these representations.	The FusionNet model outperforms existing phishing detection methods. Multi-modal methods enhance robustness and generalization capability.
Large Multimodal Agents [58]	Proposes a two-tiered agentic approach using Gemini 1.5 Flash and GPT-4o.	Gemini 1.5 Flash model: 93% accuracy in multimodal detection. GPT-4o mini model: 94% accuracy in multimodal detection, 90% accuracy in URL-based detection, 76% accuracy with image detection.
PhishLLM [59]	Leverages LLMs to decode domain-brand relationships and analyze credential-taking intentions without needing explicit reference lists, reducing maintenance costs.	PhishLLM improves recall by over 30% without sacrificing precision It reduces runtime by half compared to DynaPhish Achieves a recall of 0.84 against public phishing feeds Enhances precision by 13% and boosts reported incidents by 83%.
PhiUSIIL [60]	Uses a Similarity Index to detect visual similarity-based attacks (e.g., zero-width characters, homographs, Punycode) and an Incremental Learning approach to continuously update its knowledge base.	The PhiUSIIL framework achieved 99.24% accuracy with incremental training. It reached 99.79% accuracy using a pre-training approach Extensive experimentation validated the effectiveness of the PhiUSIIL dataset.
CSE-ARS [61]	Introduces the Chat-based Social Engineering Attack Recognition System (CSE-ARS), which integrates five specialized deep learning models using a late fusion strategy and advanced optimization techniques to effectively detect and neutralize various CSE attack enablers.	CSE-ARS effectively detects and mitigates chat-based social engineering attacks. The system integrates predictions from multiple deep-learning models. Performance optimization is achieved through weighted linear aggregation and simulated annealing.
Hybrid Features by Combining Visual and Text Information to Improve Spam Filtering Performance [62]	Combines visual and text information to improve spam filtering. Uses three sub-models to extract topic-, word-, and image-embedding-based features, employing techniques like OCR, latent Dirichlet allocation (LDA), and word2Vec.	The model achieved an accuracy of 0.9814 and a macro-F1 score of 0.9813. Even with OCR evasion techniques, it maintained a mean macro-F1 score of 0.9607, demonstrating its effectiveness in classifying spam images.
MMHAM [49]	Jointly learn deep fraud cues from three major modalities: URLs, textual information, and visual design. Uses a shared dictionary learning approach to align representations from different modalities within the attention mechanism.	MMHAM improved phishing detection capabilities. It provided hierarchical interpretability for phishing threat intelligence. Advanced models like Bert and GPT-3 can enhance textual representations.
Multi-Modal Deep Learning for Effective Malicious Webpage Detection [51]	Proposes a method that uses three types of features (URL, JavaScript code, and webpage text), each processed by a distinct deep learning model, for a comprehensive analysis of the webpage.	The proposed method achieved an accuracy rate of 97.90%. A false negative rate of only 2% was reported.
Multiphish [63]	A novel approach that fuses multi-modal features (domain, favicon, and URL) using neural networks for phishing detection.	The model achieves 97.79% accuracy in phishing detection.
Multi-Modal Features Representation-Based CNN Model for Malicious Website Detection [64]	Proposes a multimodal representation approach that combines textual and image-based features.	The proposed model achieved 98.88% F-measure performance. The false positive rate was reduced to 3.49%. The false negative rate achieved was 0.48%.
PhishAgent [52]	A multimodal agent combining online and offline knowledge bases with Multimodal Large Language Models (MLLMs) to enhance brand recognition and recall.	PhishAgent detects 4,139 phishing webpages, outperforming KnowPhish's 681 detections.
Phishing Website Detection through Multi-Model Analysis of HTML Content [65]	Introduces a model focusing on HTML content, integrating a Multi-Layer Perceptron (MLP) for structured data and two pre-trained NLP models for textual features.	MultiText-LP achieves a 96.80 F1 score and 97.18 accuracy.

to as a false positive. On the flip side, a false negative is when a phishing message is not caught. These errors are troublesome and may cost the user in terms of money or a wasted opportunity. In the research [68], identity-based detection methods are said to be continuously disadvantaged by high false positive rates due to the challenges associated with branding the identity of a webpage. A large percentage of deception is probable on SNS because there exists a variety of content, and even authentic-looking but dubious links. In contrast, sophisticated phishing attacks that are intended to be undetected can be falsely passed by a model that does not recognize them. Attackers may use newer phishing techniques that are outside the scope of the training dataset, which causes false negative detection [52]. These problems indicate challenges in the calibration of accuracy and precision of phishing detection systems to lower both false positive and false negative cases.

V. DISCUSSIONS

The authors attempted to address the research questions based on the findings of this study. For RQ1, the authors are interested in knowing what existing techniques are used for detecting phishing scams in text, image, and audio. According to the findings, studies used HTML tags, webpage contents, URLs, and domains as text-based input and visual contents such as logos as images. None of the studies reported using audio or voice as input, but this

TABLE 4. List of selected papers.

SL. No.	Pub Year	Author(s)	Title	Pub Title	DOI
1	2024	Albishri, A.A.; Dessouky, M.M.	A Comparative Analysis of Machine Learning Techniques for URL Phishing Detection	Engineering, Technology and Applied Science Research	10.48084/etasr.8920
2	2024	Duy, P.T.; Minh, V.Q.; Dang, B.T.H.; Son, N.D.H.; Quyen, N.H.; Pham, V.-H.	A Study on Adversarial Sample Resistance and Defense Mechanism for Multimodal Learning-Based Phishing Website Detection	IEEE Access	10.1109/ACCESS.2024.3436812
3	2022	Chai, Y.; Zhou, Y.; Li, W.; Jiang, Y.	An Explainable Multi-Modal Hierarchical Attention Model for Developing Phishing Threat Intelligence	IEEE Transactions on Dependable and Secure Computing	10.1109/TDSC.2021.3119323
4	2022	Rao, Routhu Srinivasa; Umarekar, Amey; Pais, Alwyn Roshan	Application of word embedding and machine learning in detecting phishing websites	Telecommunication Systems	10.1007/s11235-021-00850-6
5	2024	Tsinganos, N.; Fouliras, P.; Mavridis, I.; Gritzalis, D.	CSE-ARS: Deep Learning-Based Late Fusion of Multimodal Information for Chat-Based Social Engineering Attack Recognition	IEEE Access	10.1109/ACCESS.2024.3359030
6	2024	Avci, Cigdem; Tekinerdogan, Bedir; Cat	Design tactics for tailoring transformer architectures to cybersecurity challenges	Cluster Computing	10.1007/s10586-024-04355-0
7	2023	Rustam, Furqan; Saher, Najia; Mehmood, Arif; Lee, Ernesto; Washington, Sandrilla; Ashraf, Imran	Detecting ham and spam emails using feature union and supervised machine learning models	Multimedia Tools Appl.	10.1007/s11042-023-14814-2
8	2024	Dharini, N.; Sudarsan, E.P.V.; Praveen, B.; Thirugnanam, D.; Sibi, K.	Enhanced Phishing Detection: Integrating Random Forest Classifier and Domain Analysis for Proactive Cybersecurity	8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings	10.1109/I-SMAC61858.2024.10714859
9	2023	Sun, Y.; Liu, G.; Han, X.; Zuo, W.; Liu, W.	FusionNet: An Effective Network Phishing Website Detection Framework Based on Multi-Modal Fusion	2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application	10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00071
10	2022	Nam, S.-G.; Jang, Y.; Lee, D.-G.; Seo, Y.-S.	Hybrid Features by Combining Visual and Text Information to Improve Spam Filtering Performance	Electronics (Switzerland)	10.3390/electronics11132053
11	2020	Azeez, Nureni Ayofe; Salaudeen, Balikis Bolanle; Misra, Sanjay; Damaševičius, Robertas; Maskeliūnas, Rytis	Identifying phishing attacks in communication networks using URL consistency features	Int. J. Electron. Secur. Digit. Forensic	10.1504/ijesdf.2020.106318
12	2023	Muralidharan, Trivikram; Nissim, Nir	Improving malicious email detection through novel designated deep-learning architectures utilizing entire email	Neural Network	10.1016/j.neunet.2022.09.002
13	2024	Trad, Fouad; Chehab, Ali	Large Multimodal Agents for Accurate Phishing Detection with Enhanced Token Optimization and Cost Reduction	Arxiv	arXiv:2412.02301
14	2024	Liu, Ruofan; Lin, Yun; Teoh, Xiwen; Liu, Gongshen; Huang, Zhiyong; Dong, Jin Song	Less defined knowledge and more true alarms: reference-based phishing detection without a pre-defined reference list - Proceedings of the 33rd USENIX Conference on Security Symposium	Proceedings of the 33rd USENIX Conference on Security Symposium	Not available
15	2024	Gallagher, S.; Gelman, B.; Taoufiq, S.; Vörös, T.; Lee, Y.; Kyadige, A.; Bergeron, S.	Phishing and Social Engineering in the Age of LLMs	Large Language Models in Cybersecurity: Threats, Exposure and Mitigation	10.1007/978-3-031-54827-7_8
16	2021	Anupam, Sagnik; Kar, Arpan Kumar	Phishing website detection using support vector machines and nature-inspired optimization algorithms	Telecommun. System	10.1007/s11235-020-00739-w

TABLE 4. *(Continued.)* **List of selected papers.**

SL. No.	Pub Year	Author(s)	Title	Pub Title	DOI
17	2024	Prasad, Arvind; Chandra, Shalini	PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning	Computer Security	10.1016/j.cose.2023.103545
18	2023	Vo Quang, M.; Bui Tan Hai, D.; Tran Kim Ngoc, N.; Ngo Duc Hoang, S.; Nguyen Huu, Q.; Phan The, D.; Pham, V.-H.	Shark-Eyes: A multimodal fusion framework for multi-view-based phishing website detection	ACM International Conference Proceeding Series	10.1145/3628797.3629003
19	2022	Tanha, Jafar; Zarei, Zahra	The Bombus-terrestris bee optimization algorithm for feature selection	Applied Intelligence	10.1007/s10489-022-03478-4
20	2024	Singh, Tejveer; Kumar, Manoj; Kumar, Santosh	Walkthrough phishing detection techniques	Computer Electrical Engineering	10.1016/j.compeleceng.2024.109374
21	2024	Lee, Jehyun; Lim, Peiyuan; Hooi, Bryan; Divakaran, Dinil Mon	Multimodal Large Language Models for Phishing Webpage Detection and Identification	2024 APWG Symposium on Electronic Crime Research (eCrime)	10.1109/eCrime66200.2024.00007
22	2021	Van Dooremaal, Bram; Burda, Pavlo; Allodi, Luca; Zannone, Nicola	Combining Text and Visual Features to Improve the Identification of Cloned Webpages for Early Phishing Detection	Proceedings of the 16th International Conference on Availability, Reliability and Security	10.1145/3465481.3470112
23	2023	Mishra, S.; Soni, D.	DSmishSMS-A System to Detect Smishing SMS	Neural Computing and Applications	10.1007/s00521-021-06305-y
24	2023	Tan, Colin Choon Lin; Chiew, Kang Leng; Yong, Kelvin S.C.; Sebastian, Yakub; Than, Joel Chia Ming; Tiong, Wei King	Hybrid phishing detection using joint visual and textual identity	Expert Systems with Applications	10.1016/j.eswa.2023.119723
25	2023	Belfedhal, Alaa Eddine	Multi-Modal Deep Learning for Effective Malicious Webpage Detection	Revue d'Intelligence Artificielle	10.18280/ria.370422
26	2024	Alsaedi, Mohammed; Ghaleb, Fuad A.; Saeed, Faisal; Ahmad, Jawad; Alasli, Mohammed	Multi-Modal Features Representation-Based Convolutional Neural Network Model for Malicious Website Detection	IEEE Access	10.1109/ACCESS.2023.3348071
27	2023	Al-Kabbi, H.A.; Feizi-Derakhshi, M.-R.; Pashazadeh, S.	Multi-Type Feature Extraction and Early Fusion Framework for SMS Spam Detection	IEEE Access	10.1109/ACCESS.2023.3327897
28	2021	Zhang, Lei; Zhang, Peng; Liu, Luchen; Tan, Jianlong	Multiphish: Multi-Modal Features Fusion Networks for Phishing Detection	ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	10.1109/ICASSP39728.2021.9415016
29	2024	Cao, Tri; Huang, Chengyu; Li, Yuexin; Wang, Huilin; He, Amy; Oo, Nay; Hooi, Bryan	PhishAgent: A Robust Multimodal Agent for Phishing Webpage Detection	Proceedings of the AAAI Conference on Artificial Intelligence	10.1609/aaai.v39i27.35003
30	2022	Zhang, Sirui; Yan, Zhiwei; Dong, Kejun; Li, Hongtao; Yuchi, Xuebiao	Phishing Domain Name Detection Based on Hierarchical Fusion of Multimodal Features	2022 IEEE 16th International Conference on Big Data Science and Engineering (BigDataSE)	10.1109/BigDataSE56411.2022.00010
31	2024	Yoon, Jun-Ho; Buu, Seok-Jun; Kim, Hae-Jung	Phishing Webpage Detection via Multi-Modal Integration of HTML DOM Graphs and URL Features Based on Graph Convolutional and Transformer Networks	Electronics	10.3390/electronics13163344
32	2025	Çolhak, Furkan; Ecevit, Mert İlhan; Uçar, Bilal Emir; Creutzburg, Reiner; Dağ, Hasan	Phishing Website Detection Through Multi-model Analysis of HTML Content	Proceedings of International Conference on Theoretical and Applied Computing	10.1007/978-981-97-6957-5_15

could mean that phishing based on voice (Vishing) was studied separately and not as part of the multimodal mode. In terms of machine learning techniques, many studies employed deep learning approaches like CNN, RNN, and MLP-based techniques to handle multimodal input features. Dooremal et al. [54] asserted that a combination of text and visual features can reduce misclassifications and enhance early phishing detection. This highlights the importance of leveraging multiple features and sophisticated models to improve the accuracy and robustness of phishing detection systems. Nabila et al. [56] reported that using transfer learning and ensemble methods, such as LSTM and GRU, further enhances detection capabilities by processing both textual and image features. These approaches demonstrate the

effectiveness of combining various DL techniques to improve the accuracy and robustness of phishing detection.

For RQ2, the authors are interested in knowing how effective these techniques are in a real-world scenario. Combining multiple features enhances effectiveness and improves models boost phishing detection. Studies report very high accuracy for their models. The high detection rates of these models underscore their potential for improving the robustness and accuracy of phishing detection systems. However, reports of phishing scams are increasing yearly, as reported in [1] and [6]. This highlights the ongoing challenge of the dynamic nature of phishing scams and the gap between phishing research and detection in the real world, despite advancements in detection techniques, emphasizing the need for continuous improvement and adaptation in phishing detection methods.

With RQ3, the authors are interested in further uncovering the gaps and challenges in phishing scams specific to multimodal phishing. The study uncovered several gaps and challenges. The first significant challenge is the quality and availability of data. Models trained on text-based input may not work for image-based input. Moreover, this is exacerbated by the dynamic nature of SNS content, which is frequently updated and short-lived. Building a system with multimodalities often leads to complex models that are difficult to scale and require high computational resources. Addressing these challenges in developing and maintaining effective multimodal phishing detection systems is crucial for the continued advancement and practical implementation of multimodal models in phishing detection. Other challenges include being prone to adversarial attacks such as domain rotations, logo alterations, and replacement of text contents with visual representations. False positives and false negatives add to the limitations, as well as integration with existing SNS adds another level of challenges.

VI. LIMITATIONS

This study was mainly carried out by a single author, who is a student and mentored by a supervisor. Due to this, the requirement of a team to conduct a systematic review may not be satisfied. The potential of inducing biases in the study cannot be ruled out, though due diligence has been considered. Similarly, composing query strings for each academic database had to be tailored, which might seem to be biased, but filters were applied to make the search as uniform as possible. This highlights the limitations of the study.

VII. CONCLUSION

This paper systematically reviewed recent advances and persistent challenges in multimodal input processing to address phishing attacks perpetrated through - and image-based attack vectors. The analysis found that current research studies predominantly focus on multimodal features such as HTML tags, webpage contents, URLs, domains, and logos for the detection of phishing URLs and websites. For handling these multimodal input features, deep learning approaches like CNN, RNN, and MLP-based techniques are employed. These approaches have shown promising results, suggesting that the introduction of multimodal phishing detection systems in SNS can maximize accuracy and robustness. However, several challenges remain, including scarcity and heterogeneity of high-quality labeled datasets, the computational overhead associated with training and inference in deep multimodal architectures, susceptibility to adversarial perturbations, and the difficulty of integrating detection systems into existing social networking service (SNS) infrastructures. Moreover, the prevalence of false positives and negatives remains a significant barrier to user trust and system reliability.

Addressing these challenges requires a holistic approach that focuses on improving data and computational efficiency while anticipating adversarial attacks, mitigating privacy breaches, as well as reducing false positives and negatives. Future research should prioritize the development of scalable, real-time multimodal models capable of processing diverse inputs, including text, images, and audio. Emphasis should also be placed on improving robustness against adversarial attacks, optimizing resource usage, reducing false detections, and ensuring seamless integration with existing SNS platforms. By leveraging the strengths of multimodal systems and continuously adapting to evolving threats, it is possible to build more effective and reliable phishing detection solutions.

APPENDIX A
LIST OF SELECTED PAPERS

See Table 4.

APPENDIX B
DATA AVAILABILITY

Data for this study is available from [here.]

ACKNOWLEDGMENT

The Tad Gonsalves AI Laboratory at Sophia University supported this study by providing an excellent research environment, and the Project for Human Resource Development Scholarship (JDS) funded the Ph.D. study.

REFERENCES

[1] *APWG | Phishing Activity Trends Reports*. Accessed: Mar. 10, 2025. [Online]. Available: https://apwg.org/trendsreports/

[2] I. Del Pozo, M. Iturralde, and F. Restrepo, "Social engineering: Application of psychology to information security," in *Proc. 6th Int. Conf. Future Internet Things Cloud Workshops (FiCloudW)*, Aug. 2018, pp. 108–114. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8488183

[3] (May 2024). *What is Phishing? | IBM*. [Online]. Available: https://www.ibm.com/think/topics/phishing

[4] S. Mishra and D. Soni, "DSmishSMS—A system to detect smishing SMS," *Neural Comput. Appl.*, vol. 35, no. 7, pp. 4975–4992, Mar. 2023.

[5] J. Figueiredo, A. Carvalho, D. Castro, D. Gonçalves, and N. Santos, "On the feasibility of fully AI-automated vishing attacks," 2024, *arXiv:2409.13793*.

[6] (2024). *State of Cybersecurity 2024*. [Online]. Available: https://www.isaca.org/resources/reports/state-of-cybersecurity-2024

[7] E. D. Frauenstein and S. Flowerday, "Susceptibility to phishing on social network sites: A personality information processing model," *Comput. Secur.*, vol. 94, May 2020, Art. no. 101862. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404820301346

[8] T. Ige, C. Kiekintveld, A. Piplai, A. Waggler, O. Kolade, and B. H. Matti, "An investigation into the performances of the current state-of-the-art naive Bayes, non-Bayesian and deep learning based classifier for phishing detection: A survey," 2024, *arXiv:2411.16751*.

[9] P. H. Kyaw, J. Gutierrez, and A. Ghobakhlou, "A systematic review of deep learning techniques for phishing email detection," *Electronics*, vol. 13, no. 19, p. 3823, Sep. 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/19/3823

[10] S. Kavya and D. Sumathi, "Staying ahead of phishers: A review of recent advances and emerging methodologies in phishing detection," *Artif. Intell. Rev.*, vol. 58, no. 2, p. 50, Dec. 2024.

[11] S. Ahmad, M. Zaman, A. S. Al-Shamayleh, R. Ahmad, S. M. Abdulhamid, I. Ergen, and A. Akhunzada, "Across the spectrum in-depth review AI-based models for phishing detection," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 2065–2089, 2025.

[12] L. Yang, J. Zhang, X. Wang, Z. Li, Z. Li, and Y. He, "An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113863.

[13] R. B. Basnet, A. H. Sung, and Q. Liu, "Rule-based phishing attack detection," in *Proc. Int. Conf. Secur. Manage. (SAM)*, vol. 1, no. 1, Jul. 2011, pp. 624–630.

[14] A. Safi and S. Singh, "A systematic literature review on phishing website detection techniques," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 2, pp. 590–611, Feb. 2023.

[15] Y. A. Alsariera, M. H. Alanazi, Y. Said, and F. Allan, "An investigation of AI-based ensemble methods for the detection of phishing attacks," *Eng., Technol. Appl. Sci. Res.*, vol. 14, no. 3, pp. 14266–14274, Jun. 2024.

[16] H. Rashid, H. B. Liaqat, M. U. Sana, T. Kiren, H. Karamti, and I. Ashraf, "Framework for detecting phishing crimes on Twitter using selective features and machine learning," *Comput. Electr. Eng.*, vol. 124, May 2025, Art. no. 110363.

[17] S. Asiri, Y. Xiao, S. Alzahrani, S. Li, and T. Li, "A survey of intelligent detection designs of HTML URL phishing attacks," *IEEE Access*, vol. 11, pp. 6421–6443, 2023.

[18] R. Alazaidah, A. Al-Shaikh, M. R. Al-Mousa, H. Khafajah, G. Samara, M. Alzyoud, N. Al-Shanableh, and S. Almatarneh, "Website phishing detection using machine learning techniques," *J. Statist. Appl. Probab.*, vol. 13, no. 1, pp. 119–129, Jul. 2023.

[19] L. Mat Rani, C. Mohd Foozy, and S. Mustafa, "Feature selection to enhance phishing website detection based on URL using machine learning techniques," *J. Soft Comput. Data Mining*, vol. 4, no. 1, pp. 30–41, Jan. 2023.

[20] A. A. Albishri and M. M. Dessouky, "A comparative analysis of machine learning techniques for URL phishing detection," *Eng., Technol. Appl. Sci. Res.*, vol. 14, no. 6, pp. 18495–18501, Dec. 2024.

[21] M. Almseidin, A. M. Abu Zuraiq, M. Al-Kasassbeh, and N. Alnidami, "Phishing detection based on machine learning and feature selection methods," *Int. J. Interact. Mobile Technol.*, vol. 13, no. 12, pp. 171–183, Dec. 2019.

[22] M. Ramaiah, V. Chandrasekaran, V. Chand, A. Vasudevan, S. Mohamma, E. Soon, Q. Shambour, and M. Alshurideh, "Enhanced phishing detection: An ensemble stacking model with DT-RFECV and SMOTE," *Appl. Math. Inf. Sci.*, vol. 18, no. 6, pp. 1481–1493, Nov. 2024.

[23] K. M. Sudar, M. Rohan, and K. Vignesh, "Detection of adversarial phishing attack using machine learning techniques," *Sādhanā*, vol. 49, no. 3, p. 232, Aug. 2024.

[24] M. A. Tamal, M. K. Islam, T. Bhuiyan, A. Sattar, and N. U. Prince, "Unveiling suspicious phishing attacks: Enhancing detection with an optimal feature vectorization algorithm and supervised machine learning," *Frontiers Comput. Sci.*, vol. 6, p. 1428013, Jul. 2024.

[25] P. M. Paithane, "URLGuard: A holistic hybrid machine learning approach for phishing detection," *Int. J. Inf. Eng. Electron. Bus.*, vol. 17, no. 2, pp. 95–110, Apr. 2025.

[26] A. Karim, M. Shahroz, K. Mustofa, S. Belhaouari, and S. Joga, "Phishing detection system through hybrid machine learning based on URL," *IEEE Access*, vol. 11, pp. 36805–36822, 2023.

[27] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. A. Alzahrani, "A deep learning-based innovative technique for phishing detection in modern security with uniform resource locators," *Sensors*, vol. 23, no. 9, p. 4403, Apr. 2023.

[28] S. Asiri, Y. Xiao, and T. Li, "PhishTransformer: A novel approach to detect phishing attacks using URL collection and transformer," *Electronics*, vol. 13, no. 1, p. 30, Dec. 2023.

[29] U. Zara, K. Ayyub, H. U. Khan, A. Daud, T. Alsahfi, and S. G. Ahmad, "Phishing website detection using deep learning models," *IEEE Access*, vol. 12, pp. 167072–167087, 2024.

[30] S. Kavya and D. Sumathi, "Multimodal and temporal graph fusion framework for advanced phishing website detection," *IEEE Access*, vol. 13, pp. 74128–74146, 2025.

[31] S. Remya, M. Pillai, B. Aparna, S. Rama Subbareddy, and Y. Cho, "BGL-PhishNet: Phishing website detection using hybrid model-BERT, GNN, and LightGBM," *IEEE Access*, vol. 13, pp. 47552–47569, 2025.

[32] D. M. Divakaran and A. Oest, "Phishing detection leveraging machine learning and deep learning: A review," *IEEE Secur. Privacy*, vol. 20, no. 5, pp. 86–95, Sep. 2022.

[33] A. Alhuzali, A. Alloqmani, M. Aljabri, and F. Alharbi, "In-depth analysis of phishing email detection: Evaluating the performance of machine learning and deep learning models across multiple datasets," *Appl. Sci.*, vol. 15, no. 6, p. 3396, Mar. 2025.

[34] J. Setu, N. Halder, A. Islam, and M. Amin, "RSTHFS: A rough set theory-based hybrid feature selection method for phishing website classification," *IEEE Access*, vol. 13, pp. 68820–68830, 2025.

[35] S. S. Patil, N. M. Shekokar, and S. C. Iyer, "Design of intelligent feature selection technique for phishing detection," *IIUM Eng. J.*, vol. 26, no. 1, pp. 254–277, Jan. 2025.

[36] Y. Tashtoush, M. Alajlouni, F. Albalas, and O. Darwish, "Exploring low-level statistical features of n-grams in phishing URLs: A comparative analysis with high-level features," *Cluster Comput.*, vol. 27, no. 10, pp. 13717–13736, Dec. 2024.

[37] M. Diviya, M. Subramanian, and D. Gopala Krishnan, "An optimized phishing detection model using hybrid feature selection and a fine-tuned narrow neural network with dynamic Jaya optimization to overcome cyberthreats," *Eng. Res. Exp.*, vol. 7, no. 1, Mar. 2025, Art. no. 015202.

[38] R. Meléndez, M. Ptaszynski, and F. Masui, "Comparative investigation of traditional machine-learning models and transformer models for phishing email detection," *Electronics*, vol. 13, no. 24, p. 4877, Dec. 2024.

[39] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Systematic Rev.*, vol. 10, no. 1, p. 71, Mar. 2021. [Online]. Available: https://www.bmj.com/lookup/doi/10.1136/bmj.n71

[40] B. Kitchenham, L. Madeyski, and D. Budgen, "SEGRESS: Software engineering guidelines for REporting secondary studies," *IEEE Trans. Softw. Eng.*, vol. 49, no. 3, pp. 1273–1298, Mar. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9772383

[41] *Title, Abstract and Keywords | Springer—International Publisher*. Accessed: May 6, 2025. [Online]. Available: https://www.springer.com/gp/authors-editors/authorandreviewertutorials/writing-a-journal-manuscript/title-abstract-and-keywords/10285522

[42] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan—A web and mobile app for systematic reviews," *Systematic Rev.*, vol. 5, no. 1, p. 210, Dec. 2016. [Online]. Available: http://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-016-0384-4

[43] *The AI for Academic Research | SciSpace*. Accessed: Feb. 11, 2025. [Online]. Available: https://scispace.com

[44] T. Muralidharan and N. Nissim, "Improving malicious email detection through novel designated deep-learning architectures utilizing entire email," *Neural Netw.*, vol. 157, pp. 257–279, Jan. 2023, doi: 10.1016/j.neunet.2022.09.002.

[45] R. S. Rao, A. Umarekar, and A. R. Pais, "Application of word embedding and machine learning in detecting phishing websites," *Telecommun. Syst.*, vol. 79, no. 1, pp. 33–45, Jan. 2022. [Online]. Available: https://link.springer.com/10.1007/s11235-021-00850-6

[46] N. Dharini, E. P. V. Sudarsan, B. Praveen, D. Thirugnanam, and K. Sibi, "Enhanced phishing detection: Integrating random forest classifier and domain analysis for proactive cybersecurity," in *Proc. 8th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Oct. 2024, pp. 633–643.

[47] N. A. Azeez, B. B. Salaudeen, S. Misra, R. Damaševičius, and R. Maskeliūnas, "Identifying phishing attacks in communication networks using URL consistency features," *Int. J. Electron. Secur. Digit. Forensics*, vol. 12, no. 2, pp. 200–213, Jan. 2020, doi: 10.1504/ijesdf.2020.106318.

[48] S. Anupam and A. K. Kar, "Phishing website detection using support vector machines and nature-inspired optimization algorithms," *Telecommun. Syst.*, vol. 76, no. 1, pp. 17–32, Jan. 2021, doi: 10.1007/s11235-020-00739-w.

[49] Y. Chai, Y. Zhou, W. Li, and Y. Jiang, "An explainable multi-modal hierarchical attention model for developing phishing threat intelligence," *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 2, pp. 790–803, Mar. 2022.

[50] Y. Sun, G. Liu, X. Han, W. Zuo, and W. Liu, "FusionNet: An effective network phishing website detection framework based on multi-modal fusion," in *Proc. IEEE Int. Conf. High Perform. Comput. Commun., Data Sci. Syst., Smart City Dependability Sensor, Cloud Big Data Syst. Appl. (HPCC/DSS/SmartCity/DependSys)*, Dec. 2023, pp. 474–481.

[51] A. E. Belfedhal, "Multi-modal deep learning for effective malicious webpage detection," *Revue d'Intell. Artificielle*, vol. 37, no. 4, pp. 1005–1013, Aug. 2023. [Online]. Available: https://www.iieta.org/journals/ria/paper/10.18280/ria.370422

[52] T. Cao, C. Huang, Y. Li, H. Wang, A. He, N. Oo, and B. Hooi, "PhishAgent: A robust multimodal agent for phishing webpage detection," 2024, *arXiv:2408.10738*.

[53] M. V. Quang, B. T. H. Dang, N. T. K. Ngoc, N. D. H. Son, N. H. Quyen, P. T. Duy, and V.-H. Pham, "Shark-eyes: A multimodal fusion framework for multi-view-based phishing website detection," in *ACM Int. Conf. Proc. Ser.*, Dec. 2023, pp. 793–800.

[54] B. van Dooremaal, P. Burda, L. Allodi, and N. Zannone, "Combining text and visual features to improve the identification of cloned webpages for early phishing detection," in *Proc. 16th Int. Conf. Availability, Rel. Secur.*, Vienna Austria, Aug. 2021, pp. 1–10. [Online]. Available: https://dl.acm.org/doi/10.1145/3465481.3470112

[55] S. Yu, C. An, T. Yu, Z. Zhao, T. Li, and J. Wang, "Phishing detection based on multi-feature neural network," in *Proc. IEEE Int. Perform., Comput., Commun. Conf. (IPCCC)*, Austin, TX, USA, Nov. 2022, pp. 73–79. [Online]. Available: https://ieeexplore.ieee.org/document/9894337/

[56] O. G. Nabila, H. R. Wicaksono, Girinoto, R. N. Yasa, and H. Setiawan, "Benchmarking model URL features and image based for phishing URL detection," in *Proc. Int. Conf. Informat., Multimedia, Cyber Informations Syst. (ICIMCIS)*, Jakarta Selatan, Indonesia, Nov. 2023, pp. 177–182. [Online]. Available: https://ieeexplore.ieee.org/document/10349059/

[57] P. T. Duy, V. Q. Minh, B. T. H. Dang, N. D. H. Son, N. H. Quyen, and V.-H. Pham, "A study on adversarial sample resistance and defense mechanism for multimodal learning-based phishing website detection," *IEEE Access*, vol. 12, pp. 137805–137824, 2024.

[58] F. Trad and A. Chehab, "Large multimodal agents for accurate phishing detection with enhanced token optimization and cost reduction," 2024, *arXiv:2412.02301*.

[59] R. Liu, Y. Lin, X. Teoh, G. Liu, Z. Huang, and J. S. Dong, "Less defined knowledge and more true alarms: Reference-based phishing detection without a pre-defined reference list," in *Proc. 33rd USENIX Secur. Symp.*, Aug. 2024, pp. 7067–7084.

[60] A. Prasad and S. Chandra, "PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning," *Comput. Secur.*, vol. 136, Jan. 2024, Art. no. 103545, doi: 10.1016/j.cose.2023.103545.

[61] N. Tsinganos, P. Fouliras, I. Mavridis, and D. Gritzalis, "CSE-ARS: Deep learning-based late fusion of multimodal information for chat-based social engineering attack recognition," *IEEE Access*, vol. 12, pp. 16072–16088, 2024.

[62] S.-G. Nam, Y. Jang, D.-G. Lee, and Y.-S. Seo, "Hybrid features by combining visual and text information to improve spam filtering performance," *Electronics*, vol. 11, no. 13, p. 2053, Jun. 2022.

[63] L. Zhang, P. Zhang, L. Liu, and J. Tan, "Multiphish: Multi-modal features fusion networks for phishing detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*. Toronto, ON, Canada: IEEE, Jun. 2021, pp. 3520–3524. [Online]. Available: https://ieeexplore.ieee.org/document/9415016/

[64] M. Alsaedi, F. A. Ghaleb, F. Saeed, J. Ahmad, and M. Alasli, "Multi-modal features representation-based convolutional neural network model for malicious website detection," *IEEE Access*, vol. 12, pp. 7271–7284, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10375501/

[65] F. Çolhak, M. İ. Ecevit, B. E. Uçar, R. Creutzburg, and H. Dağ, "Phishing website detection through multi-model analysis of HTML content," in *Proc. Int. Conf. Theor. Appl. Comput.*, L. Mathew, K. G. Subramanian, and A. K. Nagar, Eds., Singapore: Springer, Jan. 2024, pp. 171–184.

[66] S. Zhang, Z. Yan, K. Dong, H. Li, and X. Yuchi, "Phishing domain name detection based on hierarchical fusion of multimodal features," in *Proc. IEEE 16th Int. Conf. Big Data Sci. Eng. (Big-DataSE)*, Wuhan, China, Dec. 2022, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10062762/

[67] J.-H. Yoon, S.-J. Buu, and H.-J. Kim, "Phishing webpage detection via multi-modal integration of HTML DOM graphs and URL features based on graph convolutional and transformer networks," *Electronics*, vol. 13, no. 16, p. 3344, Aug. 2024.

[68] C. C. L. Tan, K. L. Chiew, K. S. C. Yong, Y. Sebastian, J. C. M. Than, and W. K. Tiong, "Hybrid phishing detection using joint visual and textual identity," *Expert Syst. Appl.*, vol. 220, Jun. 2023, Art. no. 119723. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417423002245

[69] H. B. Saadallah and O. N. Uçan, "Cyber security method for phishing and malicious link detection in social meda using data mining techniques," in *Proc. 7th Int. Symp. Innov. Approaches Smart Technol. (ISAS)*, Istanbul, Turkiye, Nov. 2023, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/10391779/

TANDIN WANGCHUK received the bachelor's degree in computer science from the University of New Brunswick, NB, Canada, in 2007, and the Master of Information Technology degree in network computing from the University of Canberra, Australia, in 2012. He is currently pursuing the Ph.D. degree with Sophia University, Tokyo, Japan.

Since 2008, he has been a Lecturer with the Information Technology Department, College of Science and Technology, one of the constituent colleges of the Royal University of Bhutan. He has taught numerous undergraduate courses, including programming languages (C, C++, and Java), data structures, algorithms, professional practices in IT, and cryptography. He is currently on study leave until 2027.

TAD GONSALVES (Member, IEEE) received the B.S. degree in theoretical physics and the M.S. degree in astrophysics, and the Ph.D. degree in information systems from Sophia University, Tokyo, Japan.

He is currently a Full Professor with the Department of Information and Communication Sciences, Faculty of Science and Technology, Sophia University. His research interests include bio-inspired optimization techniques and the application of deep learning to diverse problems, such as autonomous driving, drones, digital art and music, and computational linguistics. More recently, he has also been working on affective computing models. His research laboratory Tad Gonsalves AI Laboratory, Tokyo, specializes in applications of deep learning and multi-GPU computing. He has published over 150 papers in international conferences and journals. He is the author of *Artificial Intelligence: A Non-Technical Introduction* (Sophia University Press), Tokyo, and the co-author of *Artificial Intelligence for Business Optimization: Research and Applications* (2021) (CRC Press), London.