Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.078774

REVIEW

Privacy-Preserving Phishing Detection: A Systematic Review of LLMs, Federated Learning, and Blockchain Integration

Ghadi Almaktoom, Suliman Aladhadh and Salim El Khediri*

Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

*Corresponding Author: Salim El Khediri. Email: s.elkhediri@qu.edu.sa

Received: 07 January 2026; Accepted: 01 April 2026; Published: 27 May 2026

ABSTRACT: The rapid growth of phishing attempts in the enterprise could potentially lead to bankruptcy. The primary focus of the research is on detecting phishing attacks, with no interest in how the data is processed. Attackers use fraudulent methods to obtain valuable, confidential information, resulting in billions of dollars in financial losses for enterprises. In our review, we examined the methods used in phishing-detection studies. We concluded that the two main sections, centralized and decentralized methods, were the centralized ones, which aggregate data in a central server and thus violate data protection regulations, such as GDPR. In order to properly investigate the field, we put four main questions to give the reader a proper understanding of the field: what are the major detection approaches, what are their limitations and gaps, which datasets are most commonly used and trusted across different studies, and which privacy-preserving detection approaches are used and investigated in the field of phishing detection. To address these questions, we examined 105 different papers published from 2015 to 2024. Our review covers machine learning, deep learning, hybrid methods, large language models (LLMs), federated learning, and blockchain-based detection. Our investigation led to centralized approaches that achieved more than 95% accuracy but raised privacy concerns. Keeping data local on user devices offers privacy protection, as in decentralized strategies such as federated learning, at the cost of an accuracy trade-off of 1%–3%. Other decentralized methods, such as blockchain-based systems, enhance security and transparency in the pricing of computational challenges.

KEYWORDS: Phishing detection; federated learning; blockchain; privacy-preserving; deep learning; BERT; natural language processing; smart contracts; decentralized machine learning; cybersecurity

1 Introduction

Phishing attacks seek to obtain personal information through complex techniques, strategies, and tools, including content insertion, social engineering, and more. The Anti-Phishing Working Group (APWG) offers the following description of phishing, despite other existing interpretations [1]. "Phishing employs social engineering and technical misdirection for obtaining users' identities and bank account information" [1]. These attacks can cause up to $16.6 billion in losses, as the FBI Internet Crime Complaint Center (IC3) has stated in its annual reports [2], including breaches, ransomware incidents, and substantial financial losses [3]. The International Association for Information Technology Asset Managers (IAITAM) warned that remote work could increase the risk of data breaches [4,5]. The simplicity, low cost, and reduced risk associated with phishing attacks facilitate their execution. The only conditions for performing cybercrime are an Internet connection and a computer. The anonymous nature of the Internet hinders the identification and prosecution of offenders [6,7]. Among the methods for identifying malicious and

fraudulent websites, URL analysis is the most common. In machine learning, one of the most critical domains is the classification of phishing URLs. To obtain machine-learning-based security systems and train the model on features associated with genuine and phishing website labels, a large amount of data is required. Because of their exceptional performance, machine learning algorithms can swiftly identify attacks that are hidden from users or performed for the first time and are not included on a blacklist [8]. In the last decade, a mechanism called deep learning has emerged as a robust tool for detection, particularly effective for training large-scale systems or systems lacking defined features, leading to a move towards deep learning methodologies [9]. In a typical phishing scenario, an attacker emails a phishing link; the target visits a spoofed site and unknowingly submits credentials; and the attacker then reuses those stolen credentials to access the legitimate service [10]. Fig. 1 shows the two main detection types: centralized and decentralized. The target visits a spoofed site and unknowingly submits credentials, and the attacker then reuses those stolen credentials to access the legitimate service. Misuse of individual data for central learning, exposing client data to third-party risks [11], thereby violating the regulation. That is why enterprises can reach 20 million EUR, or 4% of global revenue, if they do not comply [11,12]. The underlying conflict between efficient, centralized methods for phishing detection cannot adequately address privacy concerns. Blockchain technology offers promising potential for data protection across various domains. The authors Zhu et al. [13] proposed user data protection mechanism, which, combined with distributed hash tables and cryptography, allows users to control their data through web applications. In similar situations, Nwaiku et al. [14] considered cloud security and proposed an AI-driven anomaly detection system that leverages authentication protocols such as SAML and OAuth 2.0. They aimed to detect potential security breaches in real time using unsupervised machine learning algorithms, such as Isolation Forest.

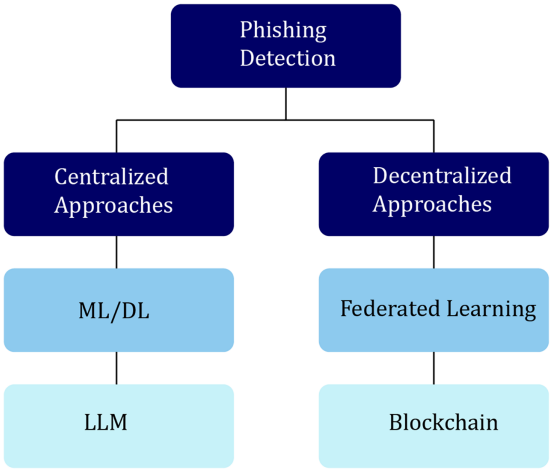

Figure 1: Comparison of centralized and decentralized phishing detection approaches.

1.1 Motivation

Usually, researchers in the field of phishing detection concentrate on achieving high accuracy. However, in the real world, maintaining the security of user data is far more important; it is about how we handle it. Different studies on this matter do not address this issue. Technologies such as federated learning and blockchain have garnered tremendous attention for enabling us to build detection systems without collecting user data in a single location. However, there is no single review that compares them with traditional methods. Highlighting what truly works is what motivates us to work in this paper.

1.2 Contribution

Our review contributes to the field of phishing detection by offering the first comprehensive analysis of the main approaches to detect traditional phishing, with privacy methods. Second, in order to distinguish between centralized and decentralized techniques, we introduce a novel classification framework. Third, our review presents the most commonly used datasets across different studies, offering their limitations and features. Fourth, we represent our view on the best method for studying privacy requirements by evaluating adversarial robustness.

1.3 Paper Structure

The paper is structured as follows: Section 2 presents related work. Section 3 describes our review methodology. Section 4 analyzes the literature on centralized approaches (machine learning, deep learning, hybrid, and LLMs) and decentralized approaches (federated learning, blockchain, and federated learning with blockchain integration). Section 5 discusses key findings, including statistical analysis of detection approaches, dataset distribution, and research gaps. Section 6 concludes the review with the main insights and recommendations. Fig. 1 illustrates the two main approaches to phishing detection examined in this review.

2 Related Work

Existing studies dedicated to detecting phishing fall short in both content and scope. While aiming for high-accuracy results and seeking the right model, they fail to address privacy and security issues. Various surveys have examined aspects of phishing detection. Here, the authors Do et al. [15] examined the ability of deep learning algorithms to detect phishing emails, focusing mainly on their strengths and limitations in this context. Similarly, Saleh and Şahin [16] review focused on detection methods, including system architectures and algorithms, and, at the end, they develop recommender systems; however, there is no mention of any security suggestions. Alkawaz et al. [17] focus specifically on developing a functional AI algorithm that outperforms other models. Also, this review did not discuss any security matters. Kytidou et al. [18] aim to investigate phishing detection approaches, such as machine learning, and the most widely used publicly accessible datasets. Kavya and Sumathi [19] (2024) survey covers traditional centralized learning, mainly machine learning and deep learning. To determine the benefits and fundamental constraints of each technique. Furthermore, Wilk-Jakubowski et al. [20] (2025) focused on identifying phishing techniques using sophisticated machine learning modelling. Alghenaim et al. [21] (2024) emphasized the use of various feature sets and classifiers to improve detection reliability, despite the ongoing challenges posed by the dynamic nature of phishing attacks and dataset imbalance.

Furthermore, Gupta et al. [22] used two main deep learning models: BERT and CNN, leveraging BERT's extraction of linguistic features and CNNs' ability to classify organizational systems. They ultimately achieve 97.5% accuracy. Synthetic Minority Over-sampling Technique (SMOTE) was used to address dataset imbalance. Bari et al. [23] provided a framework for selecting features using filters to detect phishing URLs. This framework uses several preprocessing methods, including removing constant and correlated features, using mutual information, and performing ANOVA testing. Their method used a stacking ensemble of classifiers, achieving 98.17% accuracy and a very low false-positive rate of 1.31%. That shows how important systematic feature selection is for enhancing phishing detection performance.

Table 1 compares our review with other accessible surveys. None of the previous surveys addressed federated learning, blockchain-based detection, or privacy protection. Our review aims to fill the gap by addressing these drawbacks by investigating privacy-preserving phishing detection methodologies with their classification framework to allow us differentiates between centralized and decentralized strategies.

Table 1: Comparison of related work based on various criteria.

Criteria	Do et al. [15]	Saleh and Şahin [16]	Alkawaz et al. [17]	Kytidou et al. [18]	Kavya and Sumathi [19]	Wilk-Jakubowski et al. [20]	Alghenaim et al. [21]	Our Review
Machine Learning	✓	✓	✓	✓	✓	✓	✓	✓
Deep Learning	✓	✓	✓	✓	✓	✓	✓	✓
Hybrid Approaches	×	×	×	✓	×	×	×	✓
Large Language Models (LLMs)	×	×	×	✓	×	×	×	✓
Federated Learning	×	×	×	×	×	×	×	✓
Blockchain-based Detection	×	×	×	×	×	×	×	✓
Privacy-Preserving Methods	×	×	×	×	×	×	×	✓
Centralized vs. Decentralized	×	×	×	×	×	×	×	✓
GDPR/Privacy Regulations	×	×	×	×	×	×	×	✓
Dataset Analysis	×	Partial	×	✓	×	×	×	✓
Feature Engineering	×	✓	×	✓	×	✓	✓	✓

Note: ✓ = Covered, × = Not Covered, Partial = Partially Covered.

3 Methodology

This review examines phishing detection approaches published between 2015 and 2025. We searched major databases, including IEEE Xplore, ScienceDirect, and Google Scholar, using keywords such as "phishing detection", "machine learning", "deep learning", "federated learning", and "blockchain security". We selected 105 peer-reviewed studies that focus on AI-driven phishing detection methods. To properly investigate and facilitate the search for relevant research papers, we have determined the publication years to be 2015–2025. We used datasets from IEEE Xplore, ScienceDirect, and Google Scholar, using keywords such as "phishing detection", "machine learning", "deep learning", "federated learning", and "blockchain security". We selected 105 peer-reviewed studies with a main focus on phishing detection methods. We divide the research we examined into two main groups: centralized approaches (such as machine learning, deep learning, hybrid methods, and large language models) and decentralized approaches (like federated learning and blockchain-based solutions). Fig. 2 shows the several types of phishing-detection methods examined in this evaluation. Our classification has helped us focus on traditional technologies and distributions that provide privacy.

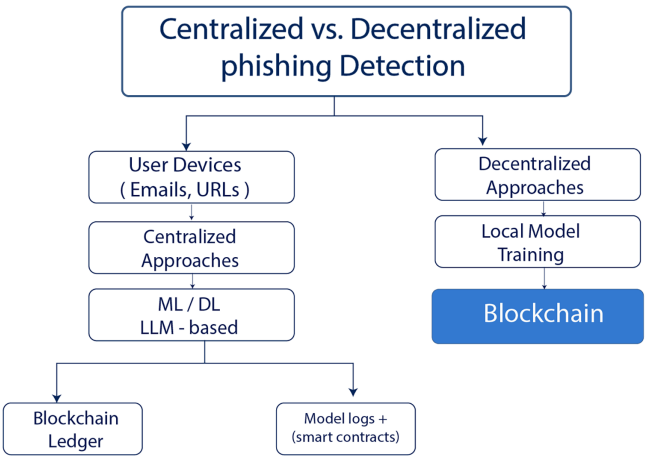

Figure 2: Phishing detection taxonomy.

3.1 Search Strategy

To cover as many relevant papers as possible, we used IEEE Xplore, ScienceDirect, and Google Scholar, concentrating on papers published between 2015 and 2025. The selected datasets were mostly known and popular among researchers. IEEE Xplore provides access to computer science journals and many conferences, while ScienceDirect covers a broader range in many computer-related fields. On the other hand, Google Scholar is a well-known search engine that primarily facilitates the search for scholarly publications. Utilizing an exclusive search engine, one can explore a wide range of disciplines and sources, including articles, theses, books, abstracts, and judicial opinions from academic publishers, professional organizations, online repositories, universities, and other sources. We searched for keywords such as phishing detection, machine learning, deep learning, federated learning, blockchain, and privacy-preserving methods. The complete search string is shown in Algorithm 1. After removing duplicates and filtering based on our inclusion and exclusion criteria, we ended up with 105 studies. Fig. 2 shows how we organized these studies into categories.

Algorithm 1: Pseudocode for defining search

```
1: String Search String =
2:    [("phishing detection" OR "phishing identification"
3:     OR "phishing classification" OR "email phishing"
4:     OR "webpage phishing" OR "URL phishing"
5:     OR "blockchain phishing" OR "cryptocurrency phishing")
6:        AND
7:    ("Machine Learning" OR "Deep Learning"
8:     OR "Artificial Intelligence" OR "Neural Network"
9:     OR "CNN" OR "RNN" OR "LSTM"
10:    OR "Support Vector Machine" OR "SVM"
11:    OR "Random Forest" OR "Decision Tree"
12:    OR "Ensemble Learning")
13:       AND
14:   ("federated learning" OR "blockchain"
15:    OR "privacy-preserving" OR "decentralized"
16:    OR "distributed learning" OR "smart contract")
17:    AND ("detection" OR "classification"
18:    OR "identification" OR "LLM"
19:    OR "transformer" OR "BERT" OR "GPT")]
```

3.2 Research Questions

This review is guided by four research questions, as summarized in Table 2.

Table 2: Research questions for the review.

#	Research Question	Aims to Answer
1	What are the major approaches for phishing detection?	To investigate the phishing detection approaches
2	What are the limitations of centralized and decentralized approaches?	To identify the commonly used techniques for identifying the phishing attempts in centralized and decentralized environments.

(Continued)

Table 2 (continued)

#	Research Question	Aims to Answer
3	What are the most used datasets in phishing detection studies?	To investigate the most used dataset in phishing detection research, also identify their type, size, and balance.
4	What are the current research gaps and future directions in privacy-preserving phishing detection?	To identify recent research gaps and outline future research directions for privacy-preserving phishing detection methods.

3.3 Inclusion and Exclusion Criteria

Table 3 presents the inclusion and exclusion criteria applied during the study selection process.

Table 3: Inclusion and exclusion criteria for paper selection.

Inclusion Criteria (IC)	Exclusion Criteria (EC)
IC1: The papers are in the field of Phishing detection	EC1: Papers that are not conducted in the context of phishing detection.
IC2: The papers study phishing detection using centralized or decentralized approaches	EC2: Publications not peer-reviewed, an abstract, an editorial letter, a book review, and a scientific report.
IC3: The paper should be published in reputable journals or recognized conference proceedings	EC3: MSc and Ph.D. thesis, Posters, and Seminar.
IC4: The studies should be written in English.	EC4: Studies that are published before 2015.
IC5: Published between 2015–2025.	

3.4 Study Selection Process (PRISMA)

The study selection process followed PRISMA 2020 guidelines (Fig. 3). The PRISMA checklists are available in the supplementary files. We began with 500 entries, removed 180 duplicates, and ended up with 320 unique records. We then screened these records based on their title and abstract, leaving us with 135 full-text articles to analyze for eligibility. After disposal of 30 publications (12 that were published before 2015, 8 that weren't peer-reviewed, as well as 10 review papers), the final informal synthesis included 105 studies.

The following paper offers a comprehensive taxonomy of phishing detection approaches across seven fundamental parameters. s is provided in Fig. 4. The taxonomy gives a short overview of major papers from 2019 to 2025 on Machine Learning, Deep Learning, Hybrid methods, Federated Learning, Blockchain, and FL+Blockchain approaches. Color coding shows how private each approach is.

4 Literature Review

4.1 Centralized Approaches

In this section, we will cover the centralized approach. Usually, this type of data processing handles data in a single server, allowing the model to be trained and aggregated. The simplicity and high accuracy

have led many researchers to adopt this approach. We divided this section into four parts. The first covers machine learning, in which the model identifies the main characteristics of a phishing attempt. Deep learning derives insights from patterns stored in raw data. The hybrid approach leverages mixing different techniques, and large language models can learn in a deeper context. All of them have their own suitable context and applications, as well as the domain that best accommodates them. On the other hand, each of them has disadvantages. Access to user data is required, compromising privacy and potentially violating regulations such as the General Data Protection Regulation (GDPR).

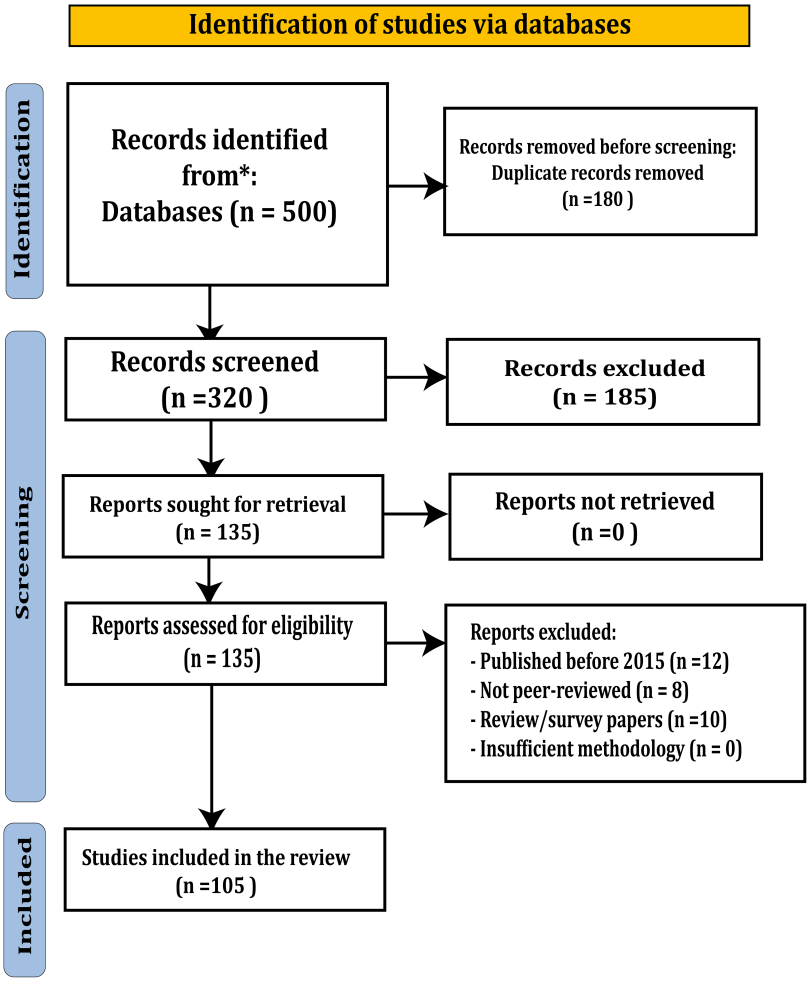

Figure 3: PRISMA 2020 flow diagram of the study selection process.

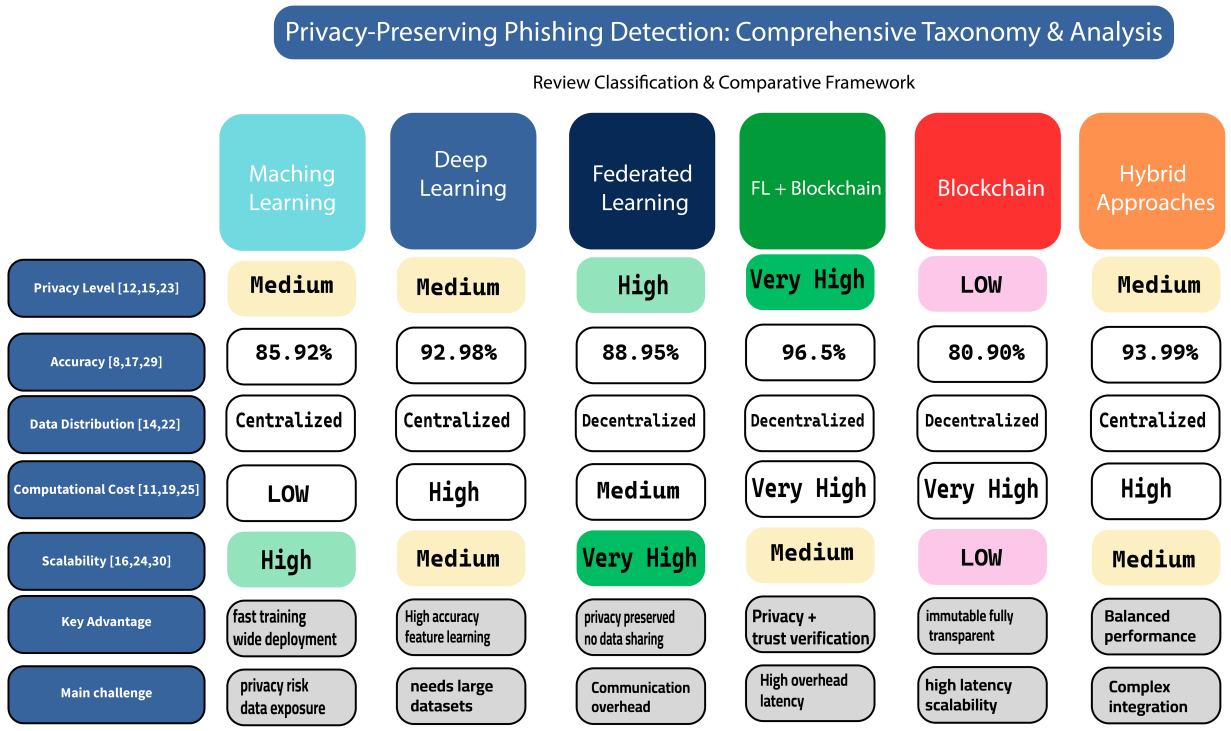

Figure 4: Privacy-preserving phishing detection: comprehensive analysis.

4.1.1 Machine Learning Techniques

Shahrivari et al. [24] examined 12 classifiers on a phishing website dataset comprising 6157 authentic websites and 4898 fraudulent websites. The classifiers investigated include Logistic Regression, Decision Tree, Support Vector Machine, AdaBoost, Random Forest, Neural Networks, K-Nearest Neighbors, Gradient Boosting, and XGBoost. The study finds that integrating multiple classification methods yields higher accuracy results. Tan et al. [25] introduced an approach, PhishWHO, for identifying phishing websites in three phases. Initially, keywords are retrieved from the websites utilizing the N-gram approach. In the second stage, these keywords are used in a web browser to identify the target domain name. Eventually, they used the same technique to verify the website's legitimacy. Chiew et al. [26] The proposed method can detect whether a web page is a phishing site. Utilize a logo image to evaluate the consistency of identification between a website's actual and represented identity. The suggested approach includes techniques for logo extraction and identity verification. The logo extraction procedure employs a machine learning technique. The identity verification technique uses Google Image Search to identify the image. Efficacy of experimental results. A graphical element, such as a logo, is more beneficial than a textual element. Harikrishnan et al. [27] used TF-IDF combined with SVD and Non-negative Matrix Factorization (NMF) representations, followed by machine learning, to categorize emails as authentic or phishing. The study concluded that decision trees and random forests achieved the highest training accuracy. We have summarized the reviewed studies in Table 4.

Studies [24–27] show that machine learning approaches can achieve high accuracy, while this high performance comes at the cost of more sophisticated techniques. Usually, these models learn threat patterns over time. The model identifies patterns in order to detect attacks. Making a model depend on a single feature to detect an attack can easily be countered. To identify fake logos in brand names, the research study [24] uses an XGBoost model with 98.3% accuracy. The cycle continues: researchers develop their tools

while attackers find ways to circumvent them. Moreover, the cycle continues indefinitely without achieving lasting protection.

Table 4: Summary of research papers on machine learning techniques.

Ref.	Method	Data	Result	Innovations	Limitations
[24]	XGBoost	PhishTank	Accuracy: 98.3%	Examined twelve classifiers	Noise has a higher impact on the result
[25]	PhishWHO	PhishTank	Accuracy: 96.10%	Identify fake web pages by searching the keywords	Image-based only
[26]	Support Vector Machine	PhishTank	Accuracy: 93.4%	Identifies the legitimate logos machine learning	Image-based only
[27]	Support Vector Machine	IWSPA	Accuracy: 99.9%	Compared multiple ML models	Overfitting due to an unbalanced dataset

4.1.2 Deep Learning Techniques

Deep learning and machine learning can learn structures and patterns from raw data and assist in detection. According to that pattern, a study by Kumar et al. [28] showed that both methods are capable of identifying intricate patterns. The main focus of the study is on using convolutional layers to detect patterns and identify links between the related content of data, enabling the model to achieve greater autonomy and efficiency on the training dataset. Sharmin et al. [29] compare CNNs to other machine learning methods. They used CNNs with a novel feature set that integrates the raw picture and Canny edges to improve earlier studies. The Support Vector Machine (SVM), Multilayer Perceptron (MLP), and CNN models outperformed conventional machine learning models with an accuracy of 99.02%. In Zavrak and Yilmaz [30], the authors' main contribution is the integration of a convolutional layer into a deep learning model to identify the most relevant components of email content for phishing. Stratification enables the model to identify patterns in the data. Their technique obtained an overall accuracy score of 0.9926. They also investigated image fraud using convolutional neural networks to compare suspicious images with the original reference image and detect potential fraud. Sharmin et al. [29] proposed a CNN model that works with SVM and MLP. Their methodology led to a high accuracy of 99.02%. Their study compares the outcomes of CNNs with those of alternative machine learning methodologies; their findings yield raw images with novel features. Ansari et al. [31] studied how well employees can spot phishing emails. They used AI-based training to test and improve employees' ability to detect phishing and strengthen security. This training utilizes AI to demonstrate how to identify phishing attempts effectively. Using AIs is an effective tool to mitigate cyberattacks. The authors Eze and Shamir [32] studied AI's ability to generate phishing emails that effectively deceive individuals. They used DeepAI to generate 865 emails containing only text. The tools, such as MALLET, Universal Data Analysis of Text (UDAT) tool, were used to recognize phrases that detect phishing patterns and to identify phishing tactics. They used the CoreNLP library with a deep neural network and Long Short-Term Memory (LSTM) to enhance the detection of phishing. Furthermore, Md et al. [33] employed a Dynamic Phishing Safeguard System (DPSS) to detect phishing emails using two main components, Anti-Phishing Neural Algorithm (APNA) and the Anti-Phishing Boosting Algorithm (APBA), to identify suspicious IP addresses, protecting user privacy through safe and timely phishing detection. The APNA results achieved 97.82% and 97.10% accuracy, respectively. In the case of Zaimi et al. [34], a two-dimensional CNN architecture is also used to detect fraudulent webpages. Their methodology consisted of

three main approaches: first, analyze the text in URLs; second, extract useful features from the text; third, use third party services to detect phishing websites. They focused on using CNN models for user protection. Their results show that 1D CNN performs better for phishing detection (96.76% accuracy), while 2D CNN is more suitable for image tasks. We have summarized the reviewed studies in Table 5.

Table 5: Summary of research papers on deep learning techniques.

Ref.	Method	Data	Result	Innovations	Limitations
[29]	CNN	Enron, SpamAssassin	Accuracy: 99.2%	Utilized CNN for word identification	High computational power
[30]	CNN	ISH	Accuracy: 99.02%	The CNN model excels in identifying phishing imagery	Image-based only
[32]	LSTM	AI-generated	Accuracy: 99.5%	Identifies AI phishing emails	Phishing tactics may not be fully covered
[33]	APBA-APNA	UCI	Accuracy: 97.82%	Able to identify phishing emails and analyze URLs	High computational power
[34]	CNN	Web Page Phishing	Accuracy: 96.76%	1D CNN excels over 2D CNN	Does not include HTML pages

These results may seem impressive and have achieved a great deal in exposing traditional phishing, except that they present a dilemma due to their potential for real-world application. If we looked closely at studies that use the same CNN model, we found that [29] reported 99.2% accuracy on the Enron dataset, which is an email dataset. In comparison, the study [30] achieved 99.02% accuracy in image-based phishing detection on the ISH dataset. Even though CNN can detect phishing easily, it yet lacks the robustness for crafted, designed phishing emails, and that is a gap in most studies [28–34] that lack testing their models against adversarial attacks.

4.1.3 Hybrid Approaches

The authors here, Bountakas and Xenakis [35], proposed HELPED, a phishing email detection method. It analyzes linguistic characteristics of emails to enhance detection accuracy. It combines two ways: ensemble learning and hybrid attributes to improve detection. To process the hybrid features separately. They proposed two approaches for HELPED: first, an ensemble learning approach, and second, a soft voting ensemble. In each approach, they used different Machine Learning algorithms. Combining the two methods, the soft voting ensemble gives better detection results than a single-feature focus. using only content-based or text-based features. The result also shows that using the Soft Voting Ensemble method on the imbalanced email dataset achieves an F1-score of 0.9942, surpassing those of traditional machine learning and deep learning models. Additionally, Alhogail and Alsabih [36] combined deep learning, graph convolutional networks (GCNs), and natural language processing to enhance detection accuracy; they achieved 98.2% accuracy. We have summarized the reviewed studies in Table 6.

Table 6: Summary of research papers on hybrid approaches.

Ref.	Method	Data	Result	Innovations	Limitations
[35]	HELPED	Combined dataset	Accuracy: 99.43%	Novel layered technique	High computational power
[36]	GCN-NLP	CLAIR Fraud Dataset	Accuracy: 98.2%	Utilizing GCN with NLP	Text-based only

In these two studies, the main constraint is that they focus solely on combining features, even though both use ensemble learning, which requires training multiple models. In the first study [35], the combination leads to a high computational cost of hybrid methods, while the second study [36] does not justify this cost. We notice that neither study shows that the high computational cost was worth it; neither explains further practical or functional challenges.

4.1.4 Large Language Models (LLMs)

In phishing, attackers aim to exploit human weaknesses by tricking them into trusting something unauthentic. That is what Heiding et al. [37] studied in their paper. They tested the two models' ability to trick human participants. They used the two models to generate nearly 112 emails, which they presented to volunteers. The criterion was the number of clicks on the links in the emails. The GPT-generated emails had success rates of 30%–44%, while V-Triad emails achieved success rates of 69%–79%. Beyond that, they tested the model: both methods combined achieved success rates ranging from 43%–81%. Additionally, models GPT, Claude, PaLM, and LLaMA were used to detect phishing email intentions, as well as their ability to detect phishing from non-harmful emails. The results showed that LLMs outperformed human participants, especially in detecting subtle phishing attempts. Kulkarni et al. [38] developed *PhishOracle* to test how well detection systems resist attacks, which add fake and harmful web pages to legitimate ones. They used the Stack model and Phishpedia to estimate the Gemini Pro Vision performance. They put it to the test by calling for 52 participants to test whether users could recognize fake brand logos on PhishOracle-generated websites. Their results demonstrated humans can easily be deceived, while the LLM Gemini Pro showed stronger resistance to these attacks. Hua et al. [39] Phishing attackers usually aim to exploit vulnerabilities by impersonating trusted entities; this study covers both text and visual elements using ChatGPT-4 and Gemini to detect them. ChatGPT-4 shows lower recall, and also has a hard time detecting more sophisticated phishing emails with hidden malicious links. Overall, it achieved high accuracy without generating false positives. Koide et al. [40] aim to investigate how LLM contextual understanding can improve the detection of different phishing techniques. They proposed a system called ChatSpamDetector that converts email content into organized prompts. They used GPT-4 as their LLM model. Their system achieved 99.70% accuracy. Additionally, Lee et al. [41] proposed a two-stage approach to phishing detection: first, analyze the original web pages and their features, such as logos, visual themes, and brand-related elements. The second stage here is the URL classification to determine whether this web page is legitimate or malicious. Using the GPT and Claude 3 models, the results show similar precision and recall; meanwhile, Gemini performed worse, with a drop of more than 15%. The examination involved modified logos and HTML content; GPT and Claude 3 maintained strong detection performance. Furthermore, Jamal and Wimmer [42] achieved satisfactory results on both balanced and imbalanced datasets; they proposed the IPSDM, a fine-tuned transformer-based model for phishing detection, based on BERT. The IPSDM model achieved 97.50% validation accuracy and 97.10% test accuracy. Their approach improves pre-trained DistilBERT and RoBERTa models. In Mittal

et al. [43], a framework based on machine learning called DARTH, their method model handles single-phishing features, and evaluation is done separately using natural language processing and neural network techniques. Their examination shows high performance in detection with an F-score of 99.98%, trained on 150,000 emails. A summary of the reviewed LLM-based approaches is presented in Table 7.

Table 7: Summary of research papers on large language models (LLMs).

Ref.	Method	Data	Result	Innovations	Limitations
[37]	GPT-4 + V-Triad	112 volunteers	Success rate: 43%–81%	Combined GPT-4 with V-Triad psychological framework for phishing generation and detection	Relies on commercial LLMs; limited participant pool
[38]	PhishOracle/ Gemini Pro Vision	PhishOracle-generated pages	Gemini showed resilience	Generates adversarial phishing pages to test detection robustness	No standardized accuracy metric reported
[39]	ChatGPT-4 /Gemini	Brand impersonation emails	High accuracy, zero FP	Evaluated LLMs on brand impersonation with textual and visual features	Reduced recall for ChatGPT-4; struggles with concealed malicious links
[40]	ChatSpam Detector (GPT-4)	Email dataset	Accuracy: 99.70%	Transforms email content into structured prompts for LLM analysis	Dependent on GPT-4 API availability and cost
[41]	GPT/Claude3/ Gemini	Phishing webpages	GPT & Claude3: high recall/precision	Two-phase framework: brand identification + phishing classification; adversarial evaluation	Gemini performance >15% lower; requires multimodal input
[42]	IPSDM (Distil-BERT/RoBERTa)	Balanced & imbalanced datasets	Accuracy: 97.50%	Fine-tuned BERT-family transformers for phishing and spam detection	Limited to text-based features only
[43]	DARTH (NLP + NN)	150,000 emails	F-score: 99.98%	Dedicated models for individual composite phishing attributes	High computational cost; email-only

In related studies we reviewed, we noticed that cloud-based commercial models such as GPT-4, Gemini, and Claude were the fundamental concern. However, locally implemented models are more appropriate for applications that care about privacy, such as LLaMA and Mistral. Allowing models to train in a decentralized way, with fine-tuning and running on-premises. The model can adjust and operate locally, which eliminates the need to send sensitive email or URL data to external servers. The method supports a decentralized approach and could be combined with federated learning to allow collaborative training without sharing raw data. Future research should focus on integrating the local LLM deployment with federated learning for phishing detection, especially in high-value data sectors such as healthcare and finance.

4.2 Decentralized Approaches

In this section, we will conduct a review. methods that reduce the privacy risks of centralized systems, such as single-server storage, by spreading data across different nodes in a decentralized approach. This approach provides a better solution for enterprises that need to maintain high detection performance while protecting data privacy. We started this section by reviewing the general use of these technologies in security applications, and then focused on their role in phishing detection.

4.2.1 Federated Learning

Korkmaz et al. [44] found that Federated Learning (FL) is a form of Distributed machine learning, first introduced by Google, to support collaborative, decentralized, and multi-device model training. The most important difference from centralization is that a server trains the model for security, rather than each individual device training it. Each device can train its own model at home using its own data, and the model does not have to expose data it doesn't need to other devices. The benefits of FL include improved performance, increased scalability, cost savings, and faster development time. Applications of federated learning span many fields, including healthcare [45–47], industrial cyber-physical systems [48], IoT anomaly detection [49], cybersecurity [50], and privacy-preserving systems [51]. As noted by Guo et al. [52], FL generally operates in an environment with multiple users or participants; therefore, a coordinator is sometimes needed to compile the insights gathered from users. This is also one reason why FL is attractive, as it can mitigate privacy issues because users' private data is never shared when training is centralized [53]. Zeng et al. [54] reported that when using FL, the centralized global model can be created on a server and distributed to clients for local training.

As summarized in Table 8, the federated learning studies [55,56] represent a promising direction for privacy-preserving phishing detection, directly addressing the regulatory and ethical concerns raised by centralized data collection. However, only three studies focus on phishing detection, leaving the field underdeveloped. Study [57] demonstrates that FL achieves comparable accuracy to centralized while keeping email data distributed across clients. Study [55] stated that achieving protection through best practices is better than traditional centralized detection methods. Research in this field, aimed at protecting user privacy, investigated federated learning (FL). Study [56] used an SMS dataset to detect phishing attacks, with the training process kept locally on users' devices using a federated learning (FL) approach. The study describes this as a solution for protecting user privacy. Their examination achived 95.02% accuracy. Maintaining the data trained on a single local server creates serious privacy risks because it requires sharing sensitive communications for model training. Earlier studies demonstrated that federated learning is a privacy-preserving alternative that enables collaborative detection without sharing raw data, thereby preserving detection accuracy, which we observed varying from minimal reductions [56,57] to more noticeable decreases [55].

Table 8: Summary of research papers on FL for phishing detection.

Ref.	Method	Data	Result	Innovations	Limitations
[57]	FL/BERT	Collected dataset	Accuracy: 96.1%	Integrates FL with BERT to detect phishing attempts	High computational power
[56]	FL/CNN-LSTM	UCI SMS	Accuracy: 99.19%	Integrates CNN-LSTM with FL to detect phishing attempts	Lack of examination of advanced feature engineering

(Continued)

Table 8 (continued)

Ref.	Method	Data	Result	Innovations	Limitations
[55]	FL/BiLSTM	Confidential	Accuracy: 83%	Integrates BiLSTM with FL to detect phishing attempts	Low accuracy

4.2.2 Blockchain-Based

Nofer et al. [58] first introduced blockchain for digital currency, such as Bitcoin, using a distributed ledger system. They also highlighted its potential for use in other applications. Swan [59] blockchain can operate across public and private environments, making it suitable for use across fields and sectors. Additionally, Esmaili and Christensen [60] clarify that public ledgers are open to all participants without restrictions. In contrast, Azaria et al. [61] explain that a private blockchain is restricted to only certain users who meet specific conditions and are authenticated and allowed to participate. Islam et al. [62] define Blockchain as a distributed ledger containing multiple blocks that hold information, where each block is linked to the previous one to maintain a historical record. Additionally, Zheng et al. [63] noted that each block has a pointer to the preceding block, using a link that is merely a hash of that block. Fig. 5 below clarifies the application of Blockchain in security.

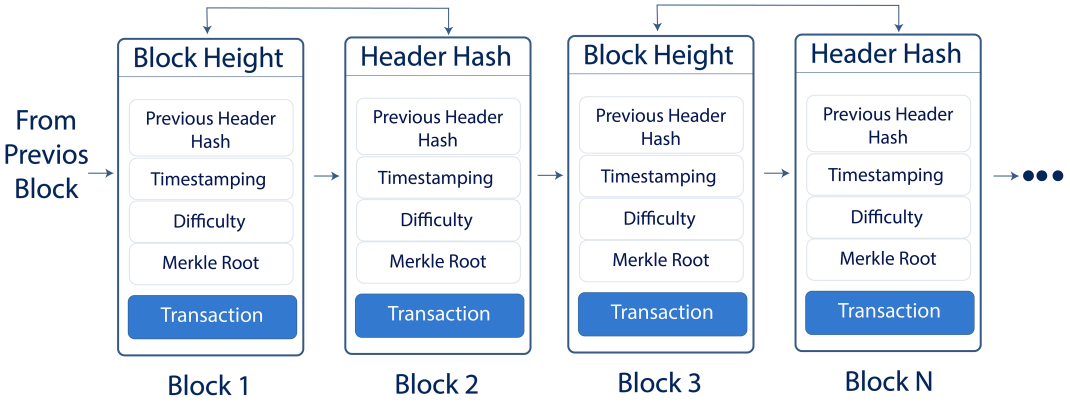

Figure 5: An example series of blocks.

4.2.3 Blockchain with Deep Learning

Scicchitano et al. [64] propose an anomaly detection system that uses an encoder-decoder deep learning model trained on aggregated data from tracking blockchain incidents. The results demonstrate the model's effectiveness in detecting publicly disclosed attacks.

Ashfaq et al. [65] blockchain technology using XGBoost and Random Forest (RF) to detect fraudulent transactions. inonther study.

Yan et al. [66] ropose ADA-Spear-an automatic phishing detection model utilizing adversarial domain adaptive learning which symbolizes the method's ability to penetrate various heterogeneous blockchains for phishing detection. Du et al. [67] Designed DeepPhishDetect, which used deep learning with blockchain to detect fraudulent nodes in blockchain networks. Their novel approach, DMFD, for node feature learning, and GAT for label dependency modeling

Nouman et al. [68] used a Histogram-based Gradient Boosting (HGB) classifier to detect the malicious node; they proposed this approach using a blockchain-based method. Their approach resulted in speed

detection. Saveetha and Maragatham [69] integrated both deep learning and blockchain to detect network intrusion. Their experiment led to high accuracy and enabled the detection of security threats. Furthermore, in a review study, Afaq and Manocha [70] highlighted that combining blockchain and deep learning improves decision-making. In other studies, both technologies were integrated.

In Hamdan et al. [71], the author proposed a solution for fraud detection that can enable risk-free, secure transactions. They propose a Deep Learning-based Blockchain Framework for Fraud Detection using Multilevel Supervision in Hierarchical Generative Hashing. Chen et al. [72] introduces a blockchain-based anti-phishing authentication protocol that enhances the security and efficiency of virtual game recharge orders. The proposed protocol integrates elliptic curve cryptography. Furthermore, Sheng et al. [73] proposed blockchain phishing detection method leveraging a dynamic feature fusion model that combines graph-based representation learning and semantic feature extraction. Varma et al. [74] proposes a novel framework, AI-MCAGCN-B-DIFCS, which integrates a Multi-Component Attention Graph Convolutional Neural Network (MCAGCN) with blockchain technology for secure and precise fraud detection. Darwish et al. [75] explores the integration of lightweight blockchain technology and deep learning for robust fraud detection in financial transactions. Lightweight blockchain ensures transaction immutability. Ref. [76] present an approach for detecting medical insurance fraud utilizing a consortium blockchain and deep learning, capable of identifying suspect medical records through an explainable model. BERT-LE is intended to assess the validity of ICD illness codes.

Zhang et al. [76] present an approach for detecting medical insurance fraud utilizing a consortium blockchain and deep learning, capable of identifying suspect medical records through an explainable model. BERT-LE is intended to assess the validity of ICD illness codes.

Ghnemat and Mosa [77] Decentralized blockchain networks are designed to make it hard to intervene in or alter records; therefore, individuals on the network can send transactions, rendering traditional fraud prevention methods ineffective. Ertam [78] This study shows that using XGBoost, LightGBM, and CatBoost can achieve 95.83%–96.46% accuracy to detect phishing attempts in Ethereum wallets. Shevchuk et al. [79] Due to increased attacks, fraud, and threat complexity, blockchain security is gaining popularity. Machine learning is replacing fundamental protection approaches as the area becomes more sophisticated. Karthika et al. [80] The proposed Phish Block on a private Ethereum blockchain has kept homographic phishing URLs. Liu et al. [81] The first effort to characterize and detect Ethereum phishing gangs was this paper. They examine phishing gang transaction habits from people's viewpoints. Bayan et al. [82] The evolution of Permissionless blockchains has become the foundation for Web3 applications, as well as decentralized finance (DeFi), resulting in privacy vulnerabilities. Vanna et al. [83] The study highlights the need for an advanced phishing detection hybrid approach to enhance the accuracy Gao et al. [84] Proposed AHGT-DFD is designed to detect phishing in blockchain-based on four categories: Feature, Encoding, Graph, and Continuous Learning there result shows 95.58% F1. Farrukh et al. [85] argued that centralized ML is more efficient than FL but might put privacy at risk. FL-blockchain hybrids reduce false positives.

Blockchain provides transparency and a decentralized ledger, as Refs. [80–84] demonstrate. However, there is no mention of how these studies address the privacy concerns. The Ethereum transaction analysis studies [86–89] focus on cryptocurrency phishing rather than email or web phishing. They rely on public blockchain data, where transaction privacy is already limited. However, these studies do not discuss privacy-preserving analysis methods or consider that fraud detection models might unintentionally reveal sensitive transaction patterns. Study [90] used a permissioned system called a "private Ethereum blockchain", which is a proposed system that can control access in the network, yet there is no evidence to support their claim. Critically, none of the blockchain studies evaluate privacy implications of their proposed systems, compare privacy properties against centralized or federated alternatives, or implement privacy-enhancing techniques

such as zero-knowledge proofs, confidential transactions, or encrypted on-chain data. Main strengths of blockchain technology are its immutability and transparency, which conflict with the privacy concept, and prior studies do not clearly show how to balance transparency and privacy without harming either. In several studies [86–89], the Ethereum transaction data used for phishing detection was collected from publicly available platforms such as Etherscan [91].

4.2.4 Federated Learning and Blockchain Integration

A combination of secure technology blockchain and federated learning can offer several solutions that are substantial to immutability and collaboratively train among different local devices. At the same time, blockchain functions as a secure, immutable ledger that documents and authenticates each model update. Together, they provide a remarkable blend of privacy and trust [92,93]. Fig. 6 demonstrates that the generalized blockchain-based federated learning paradigm incorporates blockchain as a decentralized ledger to facilitate model aggregation among distributed clients.

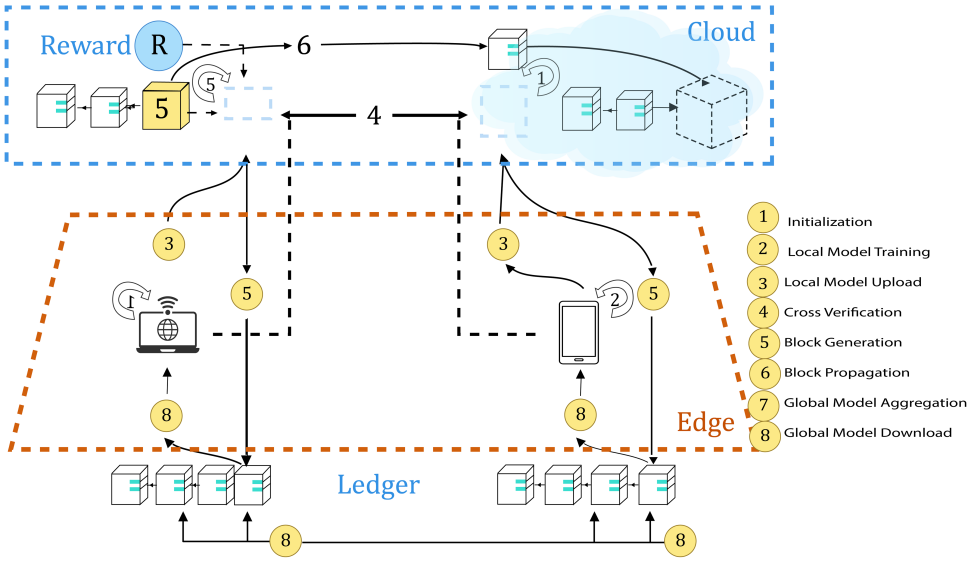

Figure 6: A generalized blockchain-based federated learning paradigm.

Many recent surveys have studied how these two technologies can function together. Research by Qammar et al. [94] investigated how blockchain can enhance the robustness of federated learning. Fig. 7 illustrates the architecture of FL members' interaction with the blockchain network via smart contracts and consensus procedures. They noted improvements in model fidelity and in protection against poisoning. Issa et al. [95] examined blockchain-based federated learning for IoT security, and Ali et al. [96] examined blockchain and FL-based intrusion detection for industrial IoT networks. Orabi et al. [97] organized current research through data partitioning to demonstrate how blockchain could enhance the security and knowledge sharing of FL systems. In blockchain-based FL setups, Sameera et al. [98] reviewed differential privacy, homomorphic encryption, and safe multiparty computation.

One of the studies, Ren et al. [99], proposed a system that uses FL with a smart contract to automate model verification. As shown in Fig. 8, the FLCoin framework uses a composed two-layer model with blocks and update blocks. Here, it records every system event, the size of the training data, contribution metrics, and validation proofs.

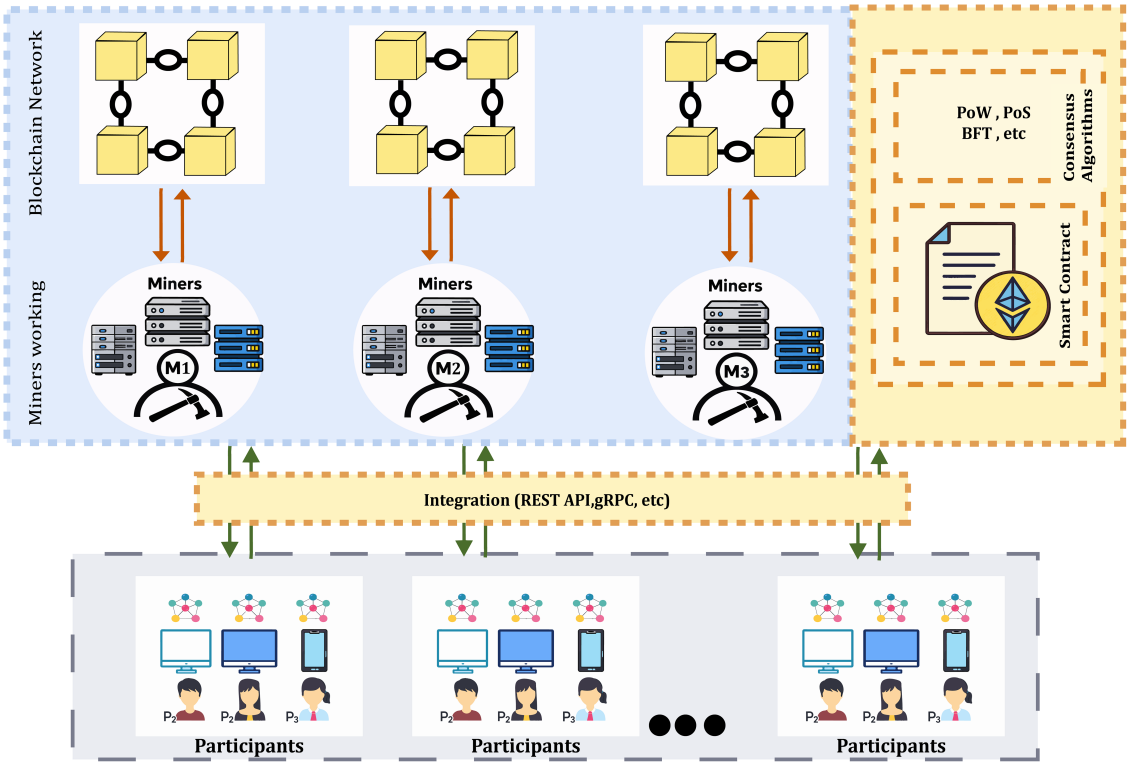

Figure 7: Architecture of a blockchain-integrated federated learning system.

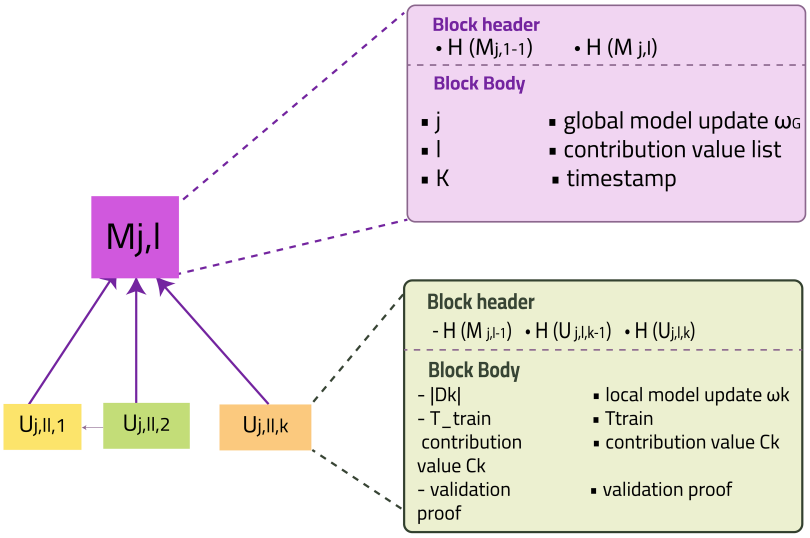

Figure 8: Structure of model and update blocks in the FLCoin framework.

Abou El Houda et al. [100] proposed using explainable AI, a blockchain, and a federated system to detect IoT intrusions. Yang and Li [101] proposed a solution to the free-rider problem, in which a system participant does not contribute to their own data. Here, the solution consists of Federated Learning (FL), Blockchain, and Incentive mechanisms to reward the participants. Li et al. [102] introduced BLADE-FL, which is in Fig. 9. Its decentralized FL blockchain framework encompasses performance evaluation metrics and

resource allocation strategies. Mainly replacing the central server with a blockchain network for aggregation and verification.

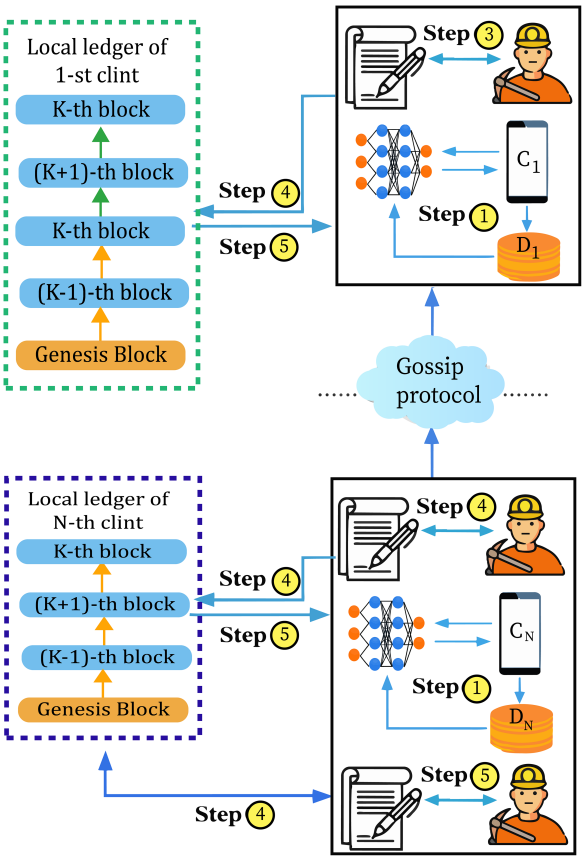

Figure 9: Architecture of the BLADE-FL framework for blockchain-assisted decentralized federated learning.

Liu et al. [103] used both blockchain and federated learning in detection. Their approach highlighted that accuracy can be improved while safeguarding the data.

The combination of decentralized technologies has driven the attention of researchers. Zhao et al. [104] developed a privacy-preserving approach specifically designed for IoT devices that have limited computing power. Lu et al. [105] designed a platform for secure data sharing in the industrial internet of things that employs the Use of a blockchain and federated learning. Fig. 10 shows the combination of blockchain technology with federated learning to enable secure data exchange among industrial IoT devices. Hallaji et al. [106] examined the security and privacy vulnerabilities of decentralized federated learning, particularly the potential of blockchain to mitigate model poisoning and data inference attacks. Han et al. [107] simultaneously provided a comprehensive examination of methodologies for ensuring both privacy and reliability in federated learning, including blockchain-based methods. Manzoor et al. [108] surveyed various defense strategies in FL, categorizing them according to what occurs before, during, and after model aggregation. Ngoupayou Limbepe et al. [109] focused on effective healthcare and examined how blockchain technology could enhance privacy in Florida-based medical systems.

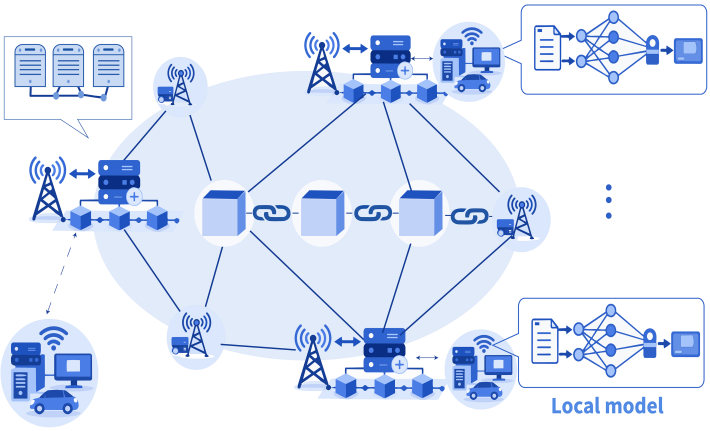

Figure 10: Blockchain-based federated learning architecture for privacy-preserving intrusion detection in Industrial IoT.

Wu et al. [110] and Ning et al. [111] have provided significant surveys that classify blockchain-based federated learning solutions according to consensus mechanisms and model aggregation methods. Cao et al. [112] have identified privacy and trust concerns given from the integration of blockchain and federated learning in intrusion detection systems. Phishing detection research is nascent yet demonstrates significant potential. Ghosh et al. [113] demonstrated that it is feasible to train phishing-detection models collaboratively across distributed nodes using FL and blockchain, while keeping transaction data private on the Ethereum network. The FedPhishLLM framework [114] took things even further by combining FL with fine-tuned multimodal large language models for phishing detection, representing one of the first attempts to merge all three technologies.

To further phishing detection, blockchain-enabled federated learning is emerging, with many recent studies presenting frameworks to address security, privacy, and scalability challenges. In an early comprehensive taxonomy of blockchain-enabled federated learning, Qu et al. [92] classified techniques by architectural frameworks and consensus mechanisms, Jiang et al. [115] conducted a comprehensive survey of blockchain-based federated learning in IoT settings, examining interactions among blockchain participants throughout the federated learning process and categorizing frameworks into three classes based on the level of integration between blockchain and federated learning. In addition, Cai et al. [116] investigated the uses of both benefits, challenges, and possible solutions; they stress model verification strategies.

Multiple frameworks have shown that combining federated learning with blockchain is effective in real-world applications. Vijay Anand et al. [117] integrated federated learning with LSTM autoencoders to safeguard blockchain network transactions, enabling diverse datasets across nodes to contribute to a global model without exchanging raw data. Their framework achieved strong anomaly detection performance while maintaining data Privacy through decentralized training. Shalan et al. [118] use knowledge distillation, transfer learning, and blockchain-enhanced FL to enable multiple IoT devices with different computing capabilities to collaborate while ensuring security via blockchain-based role-based access control.

The preservation of privacy in blockchain-enabled federated learning systems has caused increased interest among the research communities. Abuzied et al. [119] proposed FLoBC, a distributed ledger-based expandable privacy-preserving FL framework, and explored node update synchronization algorithms and associated performance trade-offs. Chen et al. [120] proposed a blockchain-based federated learning framework that establishes trust and fairness while counteracting poisoning attacks through federated computation. In study [121], the BPRFL framework was proposed to detect malicious clients and eliminate

their intrusions. Their approach uses federated learning with differential privacy. Their framework consists of tracking each client's behavior, setting two rules they must pass, and accepting updates only from participants. Their methodology achieved high accuracy.

In the study by Ali et al. FL-BCID: A Lightweight and Smart Model-Update System for Industrial IoT Systems in Decentralized Environments [122]. This paper proposes an FL-based intrusion detection system for industrial IoT environments; it uses a lightweight technique that updates the model via smart contracts and the Blockchain. Odeh and Taleb [123] proposed another combination of the two technologies, integrating federated learning with Blockchain and smart contracts, leveraging the Merkle tree to enhance integrity and eliminate unauthorized access to the network. In addition, a dual-layer hybrid blockchain–SecureChainFL–was proposed [124]. The public Blockchain ensures cryptographic auditability, and the private Blockchain (or, in some cases, a federated ledger) is used for model validation and aggregation. Moreover, for privacy protection, zero-knowledge proofs and homomorphic encryption are used.

However, several substantial obstacles persist. On public networks [94,97], transaction fees and consensus latency on blockchains can significantly delay model updates. Due to storage constraints, model parameters cannot be stored entirely on the Blockchain; as a result, systems depend on off-chain storage using IPFS [95,98]. In addition, the communication overhead inherent in federated learning, combined with blockchain verification costs, is a major stumbling block to large-scale real-time phishing detection [96,106]. Participating clients generally have non-identical data distributions, which constitute a substantial statistical challenge that blockchain technology cannot solve on its own [107,110]. Furthermore, most proposed frameworks are only evaluated in controlled lab settings, leaving many open questions about their capacity and performance in practical phishing detection [108,112]. Most importantly, the field lacks standardized benchmarks, which makes it difficult to compare various approaches fairly and consistently [109]. Table 9 summarizes the key studies on federated learning and blockchain integration.

Table 9: Summary of research papers on federated learning and blockchain integration.

Ref.	Method	Data	Result	Innovations	Limitations
[99]	FL + Blockchain + smart contracts	Edge computing data	Improved efficiency	Scalable architecture with smart contract model verification	Tested only in controlled environment
[100]	FL + Blockchain + explainable AI	IoT network traffic	Improved detection	Combined blockchain, FL, and XAI for intrusion detection	High computational overhead
[113]	FL + Blockchain phishing detection	Ethereum transactions	Accuracy: 95.8%	Decentralized defense for Ethereum phishing using FL	Limited to Ethereum transactions
[114]	FL + LLM + Blockchain	Multimodal phishing data	Improved accuracy	First framework combining FL with multimodal LLMs for phishing detection	High resource requirements for LLM fine-tuning
[102]	BLADE-FL decentralized	Distributed datasets	Improved convergence	Blockchain-assisted decentralized FL with resource allocation	Communication overhead from blockchain
[103]	FL + Blockchain vehicular	Vehicular network traffic	Accuracy: 97.5%	Collaborative intrusion detection preserving vehicle privacy	Limited to vehicular networks

(Continued)

Table 9 (continued)

Ref.	Method	Data	Result	Innovations	Limitations
[104]	FL + Blockchain for IoT	IoT device data	Accuracy: 96.2%	Privacy-preserving FL for resource-limited IoT devices	Constrained by IoT device capabilities
[105]	FL + Blockchain industrial IoT	Industrial IoT data	Improved privacy	Privacy-preserved data sharing in industrial IoT environments	Scalability not fully evaluated
[101]	FL + Blockchain fair incentives	Distributed datasets	Improved fairness	Fair incentive mechanism and secure aggregation	Free-rider issue partially addressed
[117]	FL + LSTM autoencoder	Blockchain transactions	Strong anomaly detection	FL with LSTM autoencoders for blockchain transaction security	Limited to anomaly detection
[118]	Knowledge distillation + blockchain FL	Smart home IoT data	Improved security	Knowledge distillation and transfer learning with blockchain RBAC	Limited to smart home environments
[119]	FLoBC framework	Distributed datasets	Improved privacy	Distributed ledger-based scalable privacy-preserving FL	Node synchronization trade-offs
[120]	Credible FL framework	Distributed datasets	Improved trust	Blockchain-based FL with trust and fairness against poisoning	Computational overhead of verification
[123]	BETAC-IoT	IoT network data	Improved security	Blockchain, smart contracts, FL, and Merkle tree verification	Complexity of multi-technology integration
[122]	FL-BCID	Industrial IoT traffic	Improved detection	Lightweight FL with smart contract-enabled blockchain for IDS	Limited to industrial IoT scenarios
[121]	BPRFL	Distributed datasets	Higher accuracy	Noise-separated differential privacy with reputation consensus	Performance under extreme non-IID data unclear
[124]	SecureChainFL	Distributed datasets	Enhanced privacy	Hybrid dual-layer blockchain with zero-knowledge proofs and homomorphic encryption	Scalability of ZKP verification not tested

5 Discussion

5.1 Dataset Distribution and Analysis

This paper delivered an exhaustive analysis of phishing detection techniques (and their privacy-preserving extensions). The literature was categorized into seven approaches: Machine learning (ML), deep learning (DL), hybrid (Hybrid), large language models (LLM), federated learning (FL), FL + blockchain (FL + Blockchain), and blockchain (Blockchain). Our results show a substantial increase in the number of publications on phishing detection from 2022 onwards, with 2023–2025 accounting for more than 52% of all papers we have reviewed in this area. The substantial increase in the number of papers from 2020 to 2021 is due to the COVID-19 pandemic, which led to a massive migration of the world's workforce to remote work environments and a corresponding increase in phishing attacks. The current high rate of publication on phishing detection is driven mainly by growing interest in privacy-preserving methods.

We provide a statistical evaluation of the detection performance of the respective methods using the mean accuracy and the standard deviation (see Tables 4–9). Fig. 11 illustrates that centralized methods outperform decentralized methods concerning their average accuracy. While large language models achieve an average accuracy of 99.1 99.1% (±1.1%) followed by deep learning with 98.5% (±1.0%), the average accuracy of federated learning (FL) is slightly lower at 92.8% (±7.0%). However, the combination of federated learning with blockchain improves the average accuracy to 96.5% (±0.7%) and thus exceeds the average accuracy of federated learning alone by 3.7%. Furthermore, the use of blockchain reduces the variance of average accuracy, i.e., it leads to more consistent detection. Finally, the difference in average accuracy between centralized and privacy-preserving methods is only 2.6%, indicating further improvement in decentralized methods. In Fig. 12 shows the research papers from 2015 to 2025.

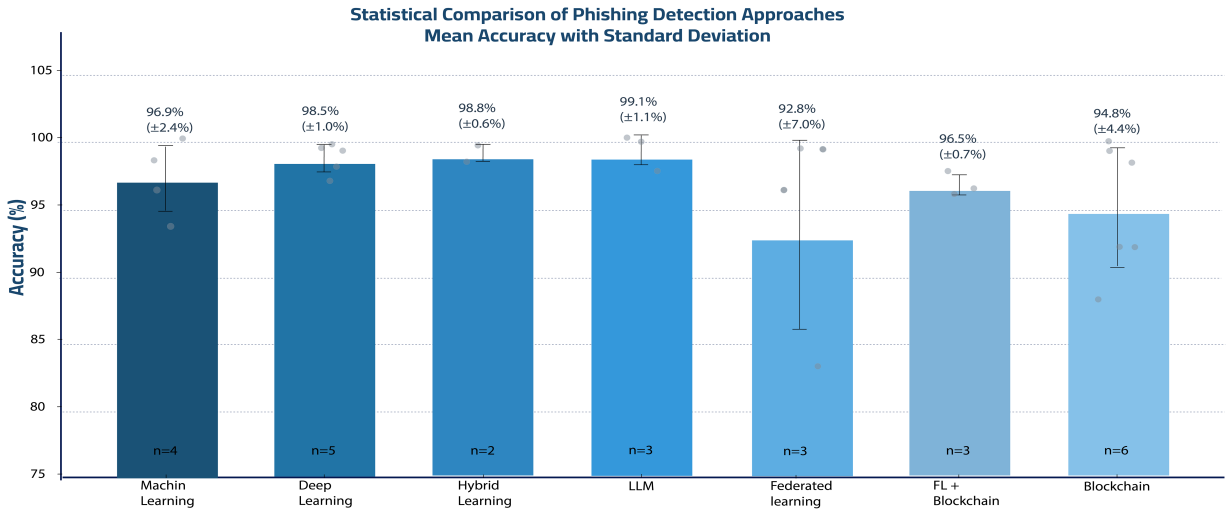

Figure 11: Statistical comparison of phishing detection approaches showing mean accuracy with standard deviation.

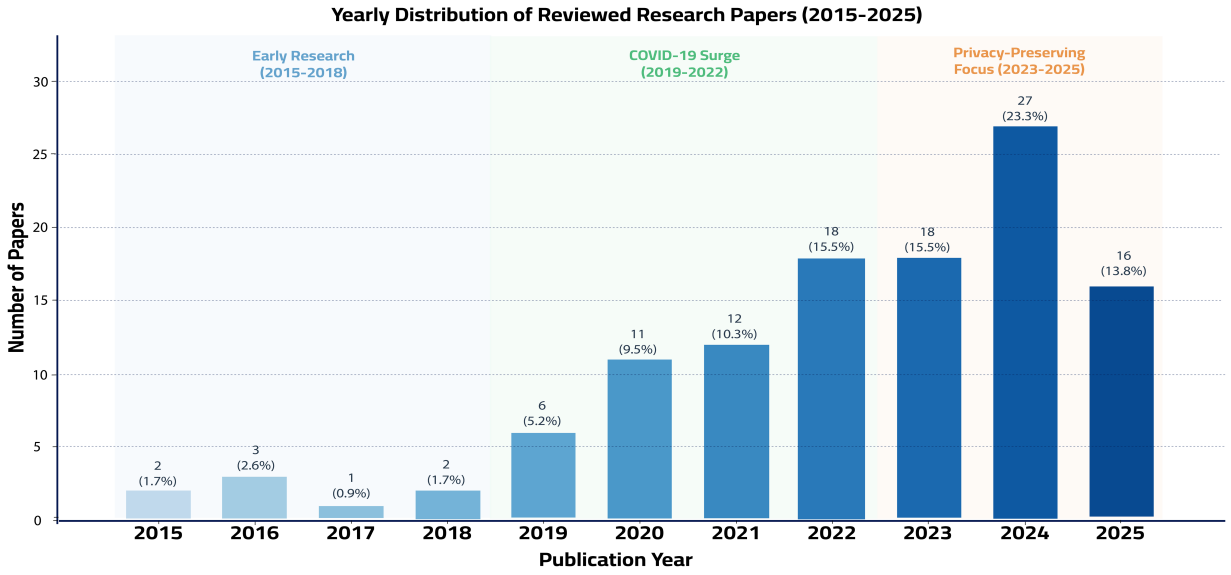

Figure 12: Yearly distribution of reviewed research papers.

Fig. 13 shows the datasets used by all of the studies we reviewed. The "Not Specified/Distributed" dataset was the most frequently used. It was used in 11 of the studies we reviewed. The second-most-frequently used dataset was the IoT dataset, which was used in 8 of the studies. The Etherscan, self-collected datasets, and other single-use datasets were each used in 7 of the studies. Six of the studies used the PhishTank dataset, indicating moderate use of this dataset for phishing detection research. Nevertheless, Sven of the studies used private or personally collected forms from others for privacy reasons; they cannot access them. In order for us to verify the result, we need the dataset to be available; not identifying them is a major limitation.

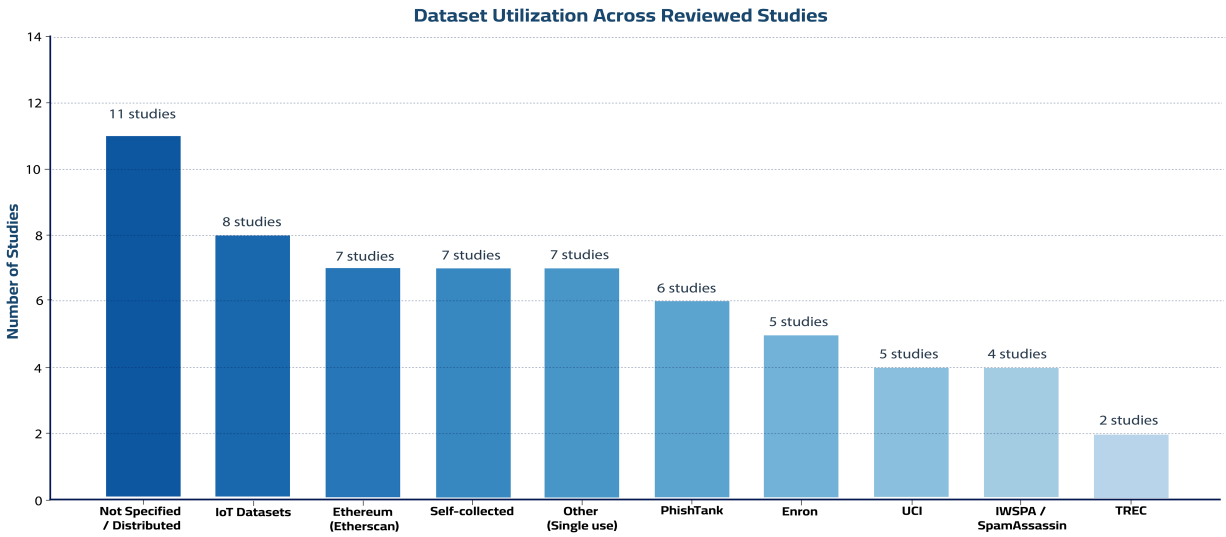

Figure 13: Dataset distribution.

As shown Table 10 previously in our review, the traditional centralized model in LLM and deep learning section can achieve an accuracy rate (up to 99.70%), while the decentralized model with the aggregated models can achieve (95.80%) at a cost, providing privacy. Our report shows the accuracy of phishing detection, privacy, and the trade-off between the two.

Table 10: Comparative analysis of phishing detection approaches across centralized and decentralized categories.

Ref.	Method	Data	Result	Innovations	Limitations
Machine Learning					
[24]	XGBoost	PhishTank	Accuracy: 98.3%	Examined 12 classifiers	Noise impact on results
[27]	SVM	IWSPA	Accuracy: 99.9%	Compared multiple ML models	Overfitting due to unbalanced dataset
Deep Learning					
[29]	CNN	Enron, SpamAssassin	Accuracy: 99.2%	CNN for word identification	High computational power

(Continued)

Table 10 (continued)

Ref.	Method	Data	Result	Innovations	Limitations
[32]	LSTM	AI-generated	Accuracy: 99.5%	Identifies AI phishing emails	Phishing tactics may not be fully covered
Hybrid Approaches					
[35]	HELPED	Combined dataset	Accuracy: 99.43%	Novel layered technique	High computational power
[36]	GCN-NLP	CLAIR Fraud	Accuracy: 98.2%	Utilizing GCN with NLP	Text-based only
Large Language Models (LLMs)					
[40]	ChatSpam Detector (GPT-4)	Email dataset	Accuracy: 99.70%	Structured prompts for LLM analysis	Dependent on GPT-4 API cost
[42]	IPSDM (Distil-BERT/RoBERTa)	Balanced & imbalanced datasets	Accuracy: 97.50%	Fine-tuned BERT-family transformers	Text-based features only
Federated Learning					
[56]	FL/CNN-LSTM	UCI SMS	Accuracy: 99.19%	CNN-LSTM integrated with FL	Lack of advanced feature engineering
[57]	FL/BERT	Collected dataset	Accuracy: 96.1%	FL integrated with BERT	High computational power
FL + Blockchain					
[113]	FL + Blockchain	Ethereum transactions	Accuracy: 95.8%	Decentralized defense for Ethereum phishing	Limited to Ethereum transactions
[114]	FL + LLM + Blockchain	Multimodal phishing data	Improved accuracy	First FL + multimodal LLM framework	High resource requirements for LLM fine-tuning
Blockchain					
[125]	Blockchain/CNN-LSTM, Bi-LSTM	Ethereum-lists	Accuracy: 99.72%	Different DL methods with blockchain	Imbalanced dataset

(Continued)

Table 10 (continued)

Ref.	Method	Data	Result	Innovations	Limitations
[89]	Blockchain/GCN	Etherscan, XBlock	Accuracy: 98.11%	Double-layer graph convolutional network	Limited dataset

One of the most notable conclusions derived from this analysis is that, while data-set imbalance is prevalent across the studies we analyzed, it was rarely fully addressed. In many studies reporting proportions of each sample type, the number of legitimate samples greatly outnumbered the number of phishing samples, typically 2:1 or more. The major problem is the overrated accuracy of term detection of legitimate samples, rather than the ability to detect phishing attacks. If we used the term accuracy, it reflects the system's ability to detect and classify. In the Study [35], the problem was investigated using a learning technique to handle class imbalance. while study [34] applied data sanitization and class balancing through randomization. Although the majority of reviewed studies did not use methods for reducing class imbalance, including oversampling (SMOTE), undersampling, cost-sensitive learning, and/or data augmentation; it seems this is an area that needs attention as accuracy, the most frequently used metric in the field to assess the performance of the detection system, is not always the best measure of performance when dealing with unbalanced datasets. Therefore, future studies should evaluate detection systems using additional measures (Table 11), e.g., precision, recall, F1 score, and false positive rate (FPR), in addition to accuracy, to better reflect how well detection systems perform when operating with large differences in sample sizes between classes.

Table 11: Summary of main datasets used in phishing detection studies.

Dataset	Type	Features	Phishing	Benign	Ratio (P:B)	Avail.	Used in Study	Ref.
PhishTank	Webpage	Logo extraction	N/S	N/S	Unk.	Public	[8,10,25,26,35]	[21]
Alexa	Webpage	Logo extraction	N/A	N/S	N/A	Public	[8,10,26]	[21]
IWSPA Email Dataset	Email	Headers, content	1113	9170	1:8.2	Public	[9,54]	[126]
Phishing Dataset	Webpage	URL extraction	4898	6157	1:1.3	Public	[9]	[127]
OpenPhish	Webpage	URL extraction	N/S	N/S	Unk.	Public	[8]	[128]
AI-generated emails	Email	Email content	N/S	N/S	Unk.	N/A	[24]	–
ISH Dataset	Image	Image extraction	920	810	1:0.9	Public	[17]	[128]
Challenge dataset 1,2	Image	Image extraction	N/D	N/D	Unk.	N/D	[17]	–
UCI	Webpage	URL extraction	N/S	N/S	Unk.	Public	[25,34,35,50]	[129]
MillerSmiles	Webpage	URL extraction	N/S	N/S	Unk.	Public	[25]	[130]
TREC	Email	Email content	50071	25217	2:1	Public	[16,32]	[131]
GenSpam	Email	Email content	30761	9186	3.3:1	Public	[16]	–
SpamAssassin	Email	Email content	1892	4144	1:2.2	Public	[16,32]	[132,133]
Enron	Email	Email content	17110	16544	1:0.97	Public	[16,30,32,50,54]	[134]
Ling spam	Email	Email content	481	2412	1:5.0	Public	[16]	[135]
spam.csv	Email	Email content	N/S	N/S	Unk.	N/S	[15]	–
emails.csv	Email	Email content	N/S	N/S	Unk.	N/S	[15]	–
Fraud dataset	Email	Email content	3685	4894	1:1.3	Public	[28]	–
NapierOne	Email, URL	Email content	N/S	N/S	Unk.	N/S	[35]	–
PhishingEmailData	Email	Email content	N/S	N/S	Unk.	Public	[34]	–
Open Phish Webpages	Webpage	Webpage content	1500	3000	1:2.0	Public	[33]	[128]

(Continued)

Table 11 (continued)

Dataset	Type	Features	Phishing	Benign	Ratio (P:B)	Avail.	Used in Study	Ref.
Ethereum Dataset	Transaction	TX details	N/S	N/S	Unk.	Public	[88]	[91]
PhishBlock	URL	Phishing URL	N/D	N/D	Unk.	Private	[80,86]	–
Microsoft 365	Email	Email content	N/D	N/D	Unk.	Private	[50]	–
CLAIR Fraud Dataset	Email	Email content	3685	4894	1:1.3	Restr.	[36]	–
Ethereum Wallet (LGBM)	Transaction	Behavioral & TX	2179	7662	1:3.5	Public	[78]	[136]
Ethereum Ponzi & Phishing (Xblock)	Transaction	Graph-based	2708	1397	1.9:1	Public	[84]	[137]
Etherscan Phishing Gangs	Transaction	On-chain TX	5363	330000+	1:62	Public	[81]	[138]

As demonstrated by the numbers in Table 11, an overwhelming class imbalance issue is apparent through all of the studies contained within this review. All but two of the datasets examined contain a significant level of imbalance. The worst example of imbalance is found with the Etherscan Phishing Gangs dataset where only one percent (1.6%) of all entries (5363 phishing out of 335,000+) are labeled as phishing. On the other hand, balanced datasets have been identified in studies on the Enron Corpus (ratio 1:0.97), and in studies using the UCI Phishing Dataset (ratio 1:1.3). The class imbalance present in these highly unbalanced datasets will significantly affect how classifiers can be evaluated. It's possible for a classifier to predict all benign and still reach an accuracy greater than ninety-five percent while not finding even a single phishing entry. Therefore, it would be beneficial for researchers to include precision, recall, F1-Score, and false positive rate (FPR) along side accuracy when evaluating their models. Also, several datasets–particularly those from federated learning studies–do not provide information about the count of either phishing or benign examples, nor do they provide conditions under which they can be accessed, preventing independent verification. At a minimum, researchers conducting federated learning studies should provide the aggregate count of phishing vs. benign examples as well as the number of client participants so that others can reproduce the results without having to compromise data security.

5.2 Answering Research Questions

This study identified four key research inquiries regarding methods for phishing detection, the need to identify additional data resources to improve the current dataset, and the need to provide a guide for future investigations.

5.2.1 What Are the Major Approaches for Phishing Detection?

In this review, we looked at studies that applied traditional machine learning methods (Naive Bayes, Random Forests, SVMs) to identify and extract attributes of phishing content [24–27]. In addition, we evaluated the application of a deep learning method [29,30,32–34] using CNNs and LSTMs, as well as a hybrid method combining machine learning and deep learning. In [35,36], we identified two hybrid approaches that combine both methods for improved detection performance. Federated learning and blockchain were the focus of studies [55–57,80,86–89,125,139,140], whereas studies [24–27,29,30,32–34, 37–43] were focused on improving detection results. The large language models (BERT and GPT) described in [37–43] demonstrate exceptional capabilities for understanding semantic content, exceeding expectations for BERT's ability to understand context and GPT's ability to generate a comprehensive representation of the content.

5.2.2 What Are the Limitations of Centralized and Decentralized Approaches?

The European Union EU designed the GDPR to protect individuals from unauthorized use of their personal data. Although it was developed to provide legal protections for individuals' rights with regard to data privacy, centralized approaches [24–27,29,30,32–34,37–43] rely on model training on a central server, which creates legal challenges when deploying those models. Decentralized methodologies [55–57,80,86–89,125,139,140] have exhibited potential advantages over centralized methods, as they can incur 1%–3% accuracy loss due to trade-offs made to preserve user data at the device level. However, the greatest challenge for methodologies that can preserve accuracy is the lack of production-ready implementations. In addition, although blockchain-based studies [80,86–89,125,139–141] demonstrate high levels of security, they do not address computational inefficiencies.

5.2.3 What Are the Most Used Datasets in Phishing Detection Studies?

We researched many of the major datasets that have been studied, namely those related to URLs (such as PhishTank, Alexa, and OpenPhish) and email data (such as IWSPA, Enron, and Nazario). Despite their importance, we found that phishing techniques have changed over time, yet these established datasets do not reflect that change. The fact that so many studies use this same data set, which is well-balanced between actual and phishing examples, will provide the best possible accuracy and effectiveness for our study. Existing datasets differ in format; some include URLs, email content, and IP addresses.

5.2.4 What Are the Current Research Gaps and Future Directions in Privacy-Preserving Phishing Detection?

The key findings from our review include a major gap in combining blockchain and federated learning in production-ready systems, along with other gaps. We noticed that most of the papers we reviewed focused mainly on accuracy; they neglected other metrics such as precision, recall, F1-score, and false positives. Furthermore, large language models with federated learning were not used to detect phishing in a privacy-preserving manner. Additionally, there is a need for standardized multilingual benchmarks to evaluate phishing detection and to broaden its scope to include new and emerging types of phishing (e.g., mobile, voice, and IoT-based phishing). Further, the issue of class imbalance needs to be systematically addressed across all phishing detection studies using techniques such as SMOTE, cost-sensitive learning, and generative data augmentation to guarantee dependable, generalizable results.

6 Conclusions

The goal of this study was to analyze the detection of Phishing and the privacy concerns of users of the above categories, i.e., (1) Centralized Detection Methods (Machine Learning, Deep Learning, Hybrid Systems, Large Language Models), and (2) Decentralized Detection Methods (Federated Learning and Blockchain Technologies). The results of our study show that central training achieves very high accuracy; however, it also enables the exploitation of user data, raising serious privacy concerns and legal obligations. We investigated other available solutions, such as Federated Learning, in which the training process is executed locally while the weights are stored on the server. At the same time, the model parameters are stored on the server. However, we found that there is currently a lack of understanding of how these systems function under intentional attacks. Therefore, we suggest that future studies create an expanded dataset of the most recent phishing attempts and incorporate multiple privacy-preserving techniques.

Acknowledgement: The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2026).

Funding Statement: The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2026).

Author Contributions: Ghadi Almaktoom: Conceptualization, methodology, data collection, writing—original draft. Suliman Aladhadh: Supervision, validation, writing—review & editing. Salim El Khediri: Supervision, validation, writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This is a review article. All data analyzed in this study are from previously published studies, which are cited in the reference list. No new datasets were generated.

Ethics Approval: Not applicable. This study is a literature review and did not involve human participants, animal subjects, or personal data collection.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at https://www.techscience.com/doi/10.32604/cmes.2026.078774/s1. The PRISMA checklists are available in the supplementary files.

References

1. APWG. APWG phishing activity trends reports. 2020 [cited 2026 Jan 1]. Available from: https://apwg.org/trendsreports/.
2. Federal Bureau of Investigation. 2024 Internet crime report. Internet crime complaint center (IC3); 2024 [cited 2026 Jan 1]. Available from: https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf.
3. Hijji M, Alam G. A multivocal literature review on growing social engineering based cyber-attacks/threats during the COVID-19 pandemic: challenges and prospective solutions. IEEE Access. 2021;9:7152–69. doi:10.1109/access.2020.3048839.
4. Di Pietro R, Raponi S, Caprolu M, Cresci S. New dimensions of information warfare. Berlin/Heidelberg, Germany: Springer; 2020. p. 1–4.
5. Al-Qahtani AF, Cresci S. The COVID-19 scamdemic: a survey of phishing attacks and their countermeasures during COVID-19. IET Inf Secur. 2022;16(5):324–45.
6. Atlam HF, Oluwatimilehin O. Business E-mail compromise phishing detection based on machine learning: a systematic literature review. Electronics. 2022;12(1):42. doi:10.3390/electronics12010042.
7. Nisha T, Bakari D, Shukla C. Business E-mail compromise—techniques and countermeasures. In: Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE); 2021 Mar 4–5; Greater Noida, India. p. 217–22.
8. Karim A, Shahroz M, Mustofa K, Belhaouari SB, Joga SRK. Phishing detection system through hybrid machine learning based on URL. IEEE Access. 2023;11(3):36805–22. doi:10.1109/access.2023.3252366.
9. Sahingoz OK, Bube E, Kugu E. Dephides: deep learning based phishing detection system. IEEE Access. 2024;12:8052–70.
10. Drainakis G, Katsaros KV, Pantazopoulos P, Sourlas V, Amditis A. Federated vs. centralized machine learning under privacy-elastic users: a comparative analysis. In: Proceedings of the 2020 IEEE 19th International Symposium on Network Computing and Applications (NCA); 2020 Nov 24–27; Online. p. 1–8.
11. Li Z, Lee G, Raghu T, Shi Z. Impact of the general data protection regulation on the global mobile app market: digital trade implications of data protection and privacy regulations. Inf Syst Res. 2025;36(2):669–89.
12. Frey CB, Presidente G. Privacy regulation and firm performance: estimating the GDPR effect globally. Econ Inq. 2024;62(3):1074–89.
13. Zhu R, Wang M, Zhang X, Peng X. Investigation of personal data protection mechanism based on blockchain technology. Sci Rep. 2023;13(1):21918. doi:10.1038/s41598-023-48661-w.
14. Nwaiku M, Diyan M, Almakdi S, Asghar I, Olugbenga A. Enhancing cloud security through anomaly detection: an artificial intelligence driven approach to secure authentication and authorization in SAML and OAuth 2.0

protocols. In: Proceedings of the International Conference on Smart Systems and Emerging Technologies. Berlin/Heidelberg, Germany: Springer; 2024. p. 432–43.
15. Do NQ, Selamat A, Krejcar O, Herrera-Viedma E, Fujita H. Deep learning for phishing detection: taxonomy, current challenges and future directions. IEEE Access. 2022;10:36429–63.
16. Saleh M, Şahin S. Phishing detection using machine learning and deep learning techniques: a review. J Comput Anal Appl. 2024;33(8):894–902. doi:10.52783/pst.1643.
17. Alkawaz MH, Steven SJ, Hajamydeen AI, Ramli R. A comprehensive survey on identification and analysis of phishing website based on machine learning methods. In: Proceedings of the 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE); 2021 Apr 3–4; Penang, Malaysia. p. 82–7.
18. Kytidou E, Tsikriki T, Drosatos G, Rantos K. Machine learning techniques for phishing detection: a review of methods, challenges, and future directions. Intell Decis Technol. 2025;19(6):4356–79.
19. Kavya S, Sumathi D. Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection. Artif Intell Rev. 2024;58(2):50.
20. Wilk-Jakubowski JL, Pawlik L, Wilk-Jakubowski G, Sikora A. Machine learning and neural networks for phishing detection: a systematic review (2017–2024). Electronics. 2025;14(18):3744. doi:10.3390/electronics14183744.
21. Alghenaim M, Alkawsi G, Barnhart CR. the state of the art in AI-based phishing detection: a systematic literature review. Curr Future Trends AI Appl. 2025;1178:431–58. doi:10.1007/978-3-031-75091-5_23.
22. Gupta BB, Gaurav A, Arya V, Attar RW, Bansal S, Alhomoud A, et al. Advanced BERT and CNN-based computational model for phishing detection in enterprise systems. Comput Model Eng Sci. 2024;141(3):2165–83. doi:10.32604/cmes.2024.056473.
23. Bari N, Saleem T, Shah M, Algarni A, Patel A, Ullah I. A filter-based feature selection framework to detect phishing URLs using stacking ensemble machine learning. Comput Model Eng Sci. 2025;145(1):1167–87. doi:10.32604/cmes.2025.070311.
24. Shahrivari V, Darabi MM, Izadi M. Phishing detection using machine learning techniques. arXiv:2009.11116. 2020.
25. Tan CL, Chiew KL, Wong K, Sze SN. PhishWHO: phishing webpage detection via identity keywords extraction and target domain name finder. Decis Support Syst. 2016;88:18–27.
26. Chiew KL, Chang EH, Sze SN, Tiong WK. Utilisation of website logo for phishing detection. Comput Secur. 2015;54(1):16–26. doi:10.1016/j.cose.2015.07.006.
27. Harikrishnan NB, Vinayakumar R, Soman KP. A machine learning approach towards phishing email detection: CEN-Security@IWSPA 2018. In: Proceedings of the 1st Anti-Phishing Shared Task Pilot at 4th ACM IWSPA Co-Located with 8th ACM Conference on Data and Application Security and Privacy (CODASPY 2018); 2018 Mar 21; Tempe, AZ, USA. p. 21–8.
28. Kumar N, Sonowal S, Nishant. Email spam detection using machine learning algorithms. In: Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA); 2020 Jul 15–17; Coimbatore, India. p. 108–13.
29. Sharmin T, Di Troia F, Potika K, Stamp M. Convolutional neural networks for image spam detection. Inf Secur J A Glob Perspect. 2020;29(3):103–17. doi:10.1080/19393555.2020.1722867.
30. Zavrak S, Yilmaz S. Email spam detection using hierarchical attention hybrid deep learning method. Expert Syst Appl. 2023;233(2):120977. doi:10.1016/j.eswa.2023.120977.
31. Ansari MF, Sharma PK, Dash B. Prevention of phishing attacks using AI-based cybersecurity awareness training. Prevention. 2022;3(6):61–72. doi:10.47893/ijssan.2022.1221.
32. Eze CS, Shamir L. Analysis and prevention of AI-based phishing email attacks. Electronics. 2024;13(10):1839. doi:10.3390/electronics13101839.
33. Md AQ, Jaiswal D, Daftari J, Haneef S, Iwendi C, Jain SK. Efficient dynamic phishing safeguard system using neural boost phishing protection. Electronics. 2022;11(19):3133. doi:10.3390/electronics11193133.
34. Zaimi R, Hafidi M, Lamia M. A deep learning approach to detect phishing websites using CNN for privacy protection. Intell Decis Technol. 2023;17(3):713–28. doi:10.3233/idt-220307.
35. Bountakas P, Xenakis C. Helphed: hybrid ensemble learning phishing email detection. J Netw Comput Appl. 2023;210:103545.

36. Alhogail A, Alsabih A. Applying machine learning and natural language processing to detect phishing email. Comput Secur. 2021;110(8):102414. doi:10.1016/j.cose.2021.102414.
37. Heiding F, Schneier B, Vishwanath A, Bernstein J, Park PS. Devising and detecting phishing emails using large language models. IEEE Access. 2024;12:42131–46. doi:10.1109/access.2024.3375882.
38. Kulkarni A, Balachandran V, Divakaran DM, Das T. From ML to LLM: evaluating the robustness of phishing web page detection models against adversarial attacks. Digit Threat Res Pract. 2025;6(2):1–25. doi:10.1145/3737295.
39. Hua J, Wang P, Lutchkus P. How effective are large language models in detecting phishing emails? Issues Inf Syst. 2024;25(3):327–41. doi:10.48009/3_iis_2024_125.
40. Koide T, Fukushi N, Nakano H, Chiba D. Chatspamdetector: leveraging large language models for effective phishing email detection. Berlin/Heidelberg, Germany: Springer; 2024. p. 297–319.
41. Lee J, Lim P, Hooi B, Divakaran DM. Multimodal large language models for phishing webpage detection and identification. In: Proceedings of the 2024 APWG Symposium on Electronic Crime Research (eCrime); 2024 Sep 24–26; Boston, WA, USA. p. 1–13.
42. Jamal S, Wimmer H. An improved transformer-based model for detecting phishing, spam, and ham: a large language model approach. arXiv:2311.04913. 2023.
43. Mittal A, Engels D, Kommanapalli H, Sivaraman R, Chowdhury T. Phishing detection using natural language processing and machine learning. SMU Data Sci Rev. 2022;6(2):14.
44. Korkmaz A, Alhonainy A, Rao P. An evaluation of federated learning techniques for secure and privacy-preserving machine learning on medical datasets. In: Proceedings of the 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR); 2022 Oct 11–13; Washington, DC, USA. p. 1–7.
45. Amin MS, Ahmad S, Loh WK. Federated learning for Healthcare 5.0: a comprehensive survey, taxonomy, challenges, and solutions. Soft Comput. 2025;29(2):673–700. doi:10.1007/s00500-025-10508-z.
46. Nasajpour M, Pouriyeh S, Parizi RM, Han M, Mosaiyebzadeh F, Liu L, et al. Federated learning in smart healthcare: a survey of applications, challenges, and future directions. Electronics. 2025;14(9):1750.
47. Pfitzner B, Steckhan N, Arnrich B. Federated learning in a medical context: a systematic literature review. ACM Trans Internet Technol. 2021;21(2):1–31. doi:10.1145/3412357.
48. Li B, Wu Y, Song J, Lu R, Li T, Zhao L. DeepFed: federated deep learning for intrusion detection in industrial cyber-physical systems. IEEE Trans Ind Inform. 2020;17(8):5615–24. doi:10.1109/tii.2020.3023430.
49. Wang X, Wang Y, Javaheri Z, Almutairi L, Moghadamnejad N, Younes OS. Federated deep learning for anomaly detection in the Internet of Things. Comput Electr Eng. 2023;108(7):108651. doi:10.1016/j.compeleceng.2023.108651.
50. Ferrag MA, Friha O, Maglaras L, Janicke H, Shu L. Federated deep learning for cyber security in the internet of things: concepts, applications, and experimental analysis. IEEE Access. 2021;9:138509–42. doi:10.1109/access.2021.3118642.
51. Jimenez-Gutierrez DM, Falkouskaya Y, Hernandez-Ramos JL, Anagnostopoulos A, Chatzigiannakis I, Vitaletti A. On the security and privacy of federated learning: a survey with attacks, defenses, frameworks, applications, and future directions. Inf Fusion. 2026;131:104155.
52. Guo W, Zhuang F, Zhang X, Tong Y, Dong J. A comprehensive survey of federated transfer learning: challenges, methods and applications. Front Comput Sci. 2024;18(6):186356. doi:10.1007/s11704-024-40065-x.
53. Kairouz P, McMahan HB. Advances and open problems in federated learning. Found Trends Mach Learn. 2021;14(1–2):1–210. doi:10.1561/2200000083.
54. Zeng R, Mi B, Huang D. A federated learning framework based on CSP homomorphic encryption. In: Proceedings of the 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS); 2023 May 12–14; Xiangtan, China. p. 196–201.
55. Sun Y, Chong N, Ochiai H. Privacy-preserving phishing email detection based on federated learning and LSTM. arXiv:2110.06025. 2021.
56. Dafni Rose J, JN M, JP S. Next-gen phishing detection system based on federated learning integrated CNN-LSTM for SMS communication. In: Proceedings of the 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV); 2024 Mar 11–12; Online. p. 367–72.

57. Thapa C, Tang JW, Abuadbba A, Gao Y, Camtepe S, Nepal S, et al. Evaluation of federated learning in phishing email detection. Sensors. 2023;23(9):4346. doi:10.3390/s23094346.
58. Nofer M, Gomber P, Hinz O, Schiereck D. Blockchain. Bus Inf Syst Eng. 2017;59(3):183–7.
59. Swan M. Blockchain: blueprint for a new economy. Sebastopol, CA, USA: O'Reilly Media, Inc.; 2015.
60. Esmaili M, Christensen K. Performance modeling of public permissionless blockchains: a survey. ACM Comput Surv. 2025;57(7):1–35.
61. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: using blockchain for medical data access and permission management. In: 2016 2nd International Conference on Open and Big Data (OBD); 2016 Aug 22–24; Vienna, Austria. p. 25–30.
62. Islam I, Munim KM, Oishwee SJ, Islam AN, Islam MN. A critical review of concepts, benefits, and pitfalls of blockchain technology using concept map. IEEE Access. 2020;8:68333–41. doi:10.1109/access.2020.2985647.
63. Zheng Z, Xie S, Dai HN, Chen X, Wang H. Blockchain challenges and opportunities: a survey. Int J Web Grid Serv. 2018;14(4):352–75. doi:10.1504/ijwgs.2018.095647.
64. Scicchitano F, Liguori A, Guarascio M, Ritacco E, Manco G. A deep learning approach for detecting security attacks on blockchain. CEUR Workshop Proc. 2020;2597:212–22.
65. Ashfaq T, Khalid R, Yahaya AS, Aslam S, Azar AT, Alsafari S, et al. A machine learning and blockchain based efficient fraud detection mechanism. Sensors. 2022;22(19):7162.
66. Yan C, Han X, Zhu Y, Du D, Lu Z, Liu Y. Phishing behavior detection on different blockchains via adversarial domain adaptation. Cybersecurity. 2024;7(1):45. doi:10.1186/s42400-024-00237-5.
67. Du W, Huang Q, Xu RR. Follow the vine to get the melon: a deep framework for blockchain phishing fraud detection. Decis Support Syst. 2025;199:114555.
68. Nouman M, Qasim U, Nasir H, Almasoud A, Imran M, Javaid N. Malicious node detection using machine learning and distributed data storage using blockchain in WSNs. IEEE Access. 2023;11(5):6106–21. doi:10.1109/access.2023.3236983.
69. Saveetha D, Maragatham G. Design of Blockchain enabled intrusion detection model for detecting security attacks using deep learning. Pattern Recognit Lett. 2022;153:24–8.
70. Afaq Y, Manocha A. Blockchain and deep learning integration for various application: a review. J Comput Inf Syst. 2024;64(1):92–105.
71. Hamdan IK, Aziguli W, Zhang D, Tiwari A. Deep learning-based blockchain framework for fraud detection using multilevel supervision in hierarchical generative hashing. Hum Centric Comput Inf Sci. 2026;17(1):15. doi:10.1007/s13042-025-02916-2.
72. Chen CM, Xiong Z, Wu TY, Kumari S, Alenazi MJ. Protecting virtual economies: a blockchain-based anti-phishing authentication protocol for metaverse applications. IEEE Internet Things J. 2025;12(13):24244–58.
73. Sheng Z, Song L, Wang Y. Dynamic feature fusion: combining global graph structures and local semantics for blockchain phishing detection. IEEE Trans Netw Serv Manag. 2025;22(5):4706–18.
74. Varma MS, Varma GP, Hemalatha I. Enhanced AI methods for cheque book scam prevention using block chain and multi-component attention graph convolutional networks. J Comput Virol Hacking Tech. 2026;22(1):33. doi:10.1007/s11416-026-00607-2.
75. Darwish SM, EL-Naggar S, Elkaffas SM. Securing financial transactions: exploring the role of lightweight blockchain-enabled deep learning for fraud detection in FinTech systems. Cybersecurity. 2026;9(1):8. doi:10.1186/s42400-025-00436-8.
76. Zhang G, Zhang X, Bilal M, Dou W, Xu X, Rodrigues JJ. Identifying fraud in medical insurance based on blockchain and deep learning. Future Gener Comput Syst. 2022;130(1):140–54. doi:10.1016/j.future.2021.12.006.
77. Ghnemat R, Mosa H. Blockchain-based fraud detection: a systematic review of Ethereum network applications. Clust Comput. 2025;28(16):1080.
78. Ertam F. Near real-time Ethereum fraud detection using explainable AI in blockchain networks. Appl Sci. 2025;15(19):10841. doi:10.3390/app151910841.
79. Shevchuk R, Martsenyuk V, Adamyk B, Benson V, Melnyk A. Anomaly detection in blockchain: a systematic review of trends, challenges, and future directions. Appl Sci. 2025;15(15):8330. doi:10.3390/app15158330.

80. Karthika R, Valliyammai C, Naveena M. Phish block: a blockchain framework for phish detection in cloud. Comput Syst Sci Eng. 2023;44(1):777–94.
81. Liu J, Chen J, Wu J, Wu Z, Fang J, Zheng Z. Fishing for fraudsters: uncovering Ethereum phishing gangs with blockchain data. IEEE Trans Inf Forensics Secur. 2024;19:3038–50.
82. Bayan T, Yazici A, Banach R. Permissionless blockchain recent trends, privacy concerns, potential solutions and secure development lifecycle. Future Internet. 2025;17(12):547. doi:10.3390/fi17120547.
83. Vanna K, Rahaman M, Gaurav A, Arya V, Hsu CH, Gupta BB, et al. Critical analysis of advanced hybrid models for mobile phishing detection through data mining and machine learning. Int J Data Warehous Min. 2025;21(1):1–32. doi:10.4018/ijdwm.394800.
84. Gao J, Richard BS, Xia H, Victor K, Fabien EB, Xia Q. AHGT-DFD: adaptive hierarchical graph transformer for dynamic fraud detection in blockchain networks. IEEE Trans Dependable Secur Comput. 2025;23(2):2229–41.
85. Farrukh H, Zafar S, Rehman ZU, Shah AA, Alshammry N. Blockchain-based fraud detection: a comparative systematic literature review of federated learning and machine learning approaches. Electronics. 2025;14(24):4952. doi:10.3390/electronics14244952.
86. Chen W, Guo X, Chen Z, Zheng Z, Lu Y. Phishing scam detection on Ethereum: towards financial security for blockchain ecosystem. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20); 2020 Jul 11–17; Yokohama, Japan. p. 4456–62.
87. Wu J, Yuan Q, Lin D, You W, Chen W, Chen C, et al. Who are the phishers? Phishing scam detection on Ethereum via network embedding. IEEE Trans Syst Man Cybern Syst. 2020;52(2):1156–66.
88. Fu B, Yu X, Feng T. CT-GCN: a phishing identification model for blockchain cryptocurrency transactions. Int J Inf Secur. 2022;21(6):1223–32.
89. Kabla AHH, Anbar M, Manickam S, Karupayah S. Eth-PSD: a machine learning-based phishing scam detection approach in Ethereum. IEEE Access. 2022;10:118043–57.
90. Kumar N, Goel V, Ranjan R, Altuwairiqi M, Alyami H, Asakipaam SA. A blockchain-oriented framework for cloud-assisted system to countermeasure phishing for establishing secure smart city. Secur Commun Netw. 2023;2023(1):8168075. doi:10.1155/2023/8168075.
91. Etherscan. Ethereum blockchain data. 2023 [cited 2026 Jan 1]. Available from: https://etherscan.io/.
92. Qu Y, Uddin MP, Gan C, Xiang Y, Gao L, Yearwood J. Blockchain-enabled federated learning: a survey. ACM Comput Surv. 2022;55(4):1–35. doi:10.1145/3524104.
93. Zhang H, Jiang S, Xuan S. Decentralized federated learning based on blockchain: concepts, framework, and challenges. Comput Commun. 2024;216(1):140–50. doi:10.1016/j.comcom.2023.12.042.
94. Qammar A, Karim A, Ning H, Ding J. Securing federated learning with blockchain: a systematic literature review. Artif Intell Rev. 2023;56(5):3951–85.
95. Issa W, Moustafa N, Turnbull B, Sohrabi N, Tari Z. Blockchain-based federated learning for securing internet of things: a comprehensive survey. ACM Comput Surv. 2023;55(9):1–43. doi:10.1145/3560816.
96. Ali S, Li Q, Yousafzai A. Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: a survey. Ad Hoc Netw. 2024;152(6):103320. doi:10.1016/j.adhoc.2023.103320.
97. Orabi MM, Emam O, Fahmy H. Adapting security and decentralized knowledge enhancement in federated learning using blockchain technology: literature review. J Big Data. 2025;12(1):55. doi:10.1186/s40537-025-01099-5.
98. Sameera K, Nicolazzo S, Arazzi M, Nocera A, KA. RR, Vinod P, et al. Privacy-preserving in blockchain-based federated learning systems. Comput Commun. 2024;222(4):38–67. doi:10.1016/j.comcom.2024.04.024.
99. Ren S, Kim E, Lee C. A scalable blockchain-enabled federated learning architecture for edge computing. PLoS One. 2024;19(8):e0308991. doi:10.1371/journal.pone.0308991.
100. Abou El Houda Z, Moudoud H, Brik B, Khoukhi L. Securing federated learning through blockchain and explainable AI for robust intrusion detection in IoT networks. In: Proceedings of the IEEE INFOCOM 2023-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 2023 May 17–20; Hoboken, NJ, USA; p. 1–6.
101. Yang X, Li T. A Blockchain-based federated learning framework for secure aggregation and fair incentives. Connect Sci. 2024;36(1):2316018. doi:10.1080/09540091.2024.2316018.

102. Li J, Shao Y, Wei K, Ding M, Ma C, Shi L, et al. Blockchain assisted decentralized federated learning (BLADE-FL): performance analysis and resource allocation. IEEE Trans Parallel Distrib Syst. 2021;33(10):2401–15.
103. Liu H, Zhang S, Zhang P, Zhou X, Shao X, Pu G, et al. Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing. IEEE Internet Things J. 2021;70(6):6073–84. doi:10.1109/tvt.2021.3076780.
104. Zhao Y, Zhao J, Jiang L, Tan R, Niyato D, Li Z, et al. Privacy-preserving blockchain-based federated learning for IoT devices. IEEE Internet Things J. 2020;8(3):1817–29.
105. Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. IEEE Trans Ind Inform. 2019;16(6):4177–86. doi:10.1109/tii.2019.2942190.
106. Hallaji E, Razavi-Far R, Saif M, Wang B, Yang Q. Decentralized federated learning: a survey on security and privacy. IEEE Trans Big Data. 2024;10(2):194–213.
107. Han Q, Lu S, Wang W, Qu H, Li J, Gao Y. Privacy preserving and secure robust federated learning: a survey. Concurr Comput Pract Exp. 2024;36(13):e8084. doi:10.1002/cpe.8084.
108. Manzoor HU, Shabbir A, Chen A, Flynn D, Zoha A. A survey of security strategies in federated learning: defending models, data, and privacy. Future Internet. 2024;16(10):374. doi:10.3390/fi16100374.
109. Ngoupayou Limbepe Z, Gai K, Yu J. Blockchain-based privacy-enhancing federated learning in smart healthcare: a survey. Blockchains. 2025;3(1):1. doi:10.3390/blockchains3010001.
110. Wu L, Ruan W, Hu J, He Y. A survey on blockchain-based federated learning. Future Internet. 2023;15(12):400. doi:10.3390/fi15120400.
111. Ning W, Zhu Y, Song C, Li H, Zhu L, Xie J, et al. Blockchain-based federated learning: a survey and new perspectives. Appl Sci. 2024;14(20):9459.
112. Cao Y, Ku CS, Kumar R, Khan A. Privacy and trust in blockchain-federated intrusion detection systems: taxonomy, challenges and perspectives. J Reliab Secur Comput. 2025;1(1):4–24. doi:10.62762/jrsc.2025.399812.
113. Ghosh PK, Bhushan A, Kumar D, Singh AK. Decentralized defences from federated learning for Ethereum phishing detection. In: International Conference on Advanced Network Technologies and Intelligent Computing. Berlin/Heidelberg, Germany: Springer; 2024. p. 243–57.
114. Li W, Manickam S, Chong YW. FedPhishLLM: a privacy-preserving and explainable phishing detection mechanism using federated learning and LLMs. J King Saud Univ Comput Inf Sci. 2025;37(8):252.
115. Jiang Y, Ma B, Wang X, Yu G, Yu P, Wang Z, et al. Blockchained federated learning for Internet of Things: a comprehensive survey. ACM Comput Surv. 2024;56(10):1–37. doi:10.1145/3659099.
116. Cai Z, Chen J, Fan Y, Zheng Z, Li K. Blockchain-empowered federated learning: benefits, challenges, and solutions. IEEE Trans Big Data. 2025;11(5):2244–63.
117. Vijay Anand R, Magesh G, Alagiri I, Brahmam MG, Balusamy B, Selvan CP, et al. Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions. Sci Rep. 2025;15(1):1615. doi:10.1038/s41598-024-83564-4.
118. Shalan M, Hasan MR, Bai Y, Li J. Enhancing smart home security: blockchain-enabled federated learning with knowledge distillation for intrusion detection. Smart Cities. 2025;8(1):35.
119. Abuzied Y, Ghanem M, Dawoud F, Gamal H, Soliman E, Sharara H, et al. A privacy-preserving federated learning framework for blockchain networks. Clust Comput. 2024;27(4):3997–4014. doi:10.1007/s10586-024-04273-1.
120. Chen L, Zhao D, Tao L, Wang K, Qiao S, Zeng X, et al. A credible and fair federated learning framework based on blockchain. IEEE Trans Artif Intell. 2024;6(2):301–16. doi:10.1109/tai.2024.3355362.
121. Guo J, Liu R, Xing J. A blockchain-based privacy-preserving reputation consensus federated learning. Alex Eng J. 2025;133(8):444–60. doi:10.1016/j.aej.2025.11.016.
122. Ali A, Husain M, Hans P. Federated learning-enhanced blockchain framework for privacy-preserving intrusion detection in industrial IoT. arXiv:2505.15376. 2025.
123. Odeh A, Taleb A. Federated learning and blockchain framework for scalable and secure IoT access control. Comput Mater Contin. 2025;84(1):447. doi:10.32604/cmc.2025.065426.
124. Usharani S, Manju Bala P, Balachandar A, Glorindal G. Enhanced privacy preserving deep learning using blockchain and federated learning. In: Blockchain and federated learning synergy for privacy-focused deepfex solutions. Berlin/Heidelberg, Germany: Springer; 2025. p. 145–64.

125. Ogundokun RO, Arowolo MO, Damaševičius R, Misra S. Phishing detection in blockchain transaction networks using ensemble learning. Telecom. 2023;4(2):279–97. doi:10.3390/telecom4020017.
126. IWSPA. IWSPA-AP email dataset. 2018 [cited 2026 Jan 1]. Available from: https://dasavisha.github.io/IWSPA-sharedtask/.
127. Mohammad R, McCluskey L. Phishing websites dataset. UCI Mach Learn Repos. 2015. doi:10.24432/C51W2X.
128. OpenPhish. OpenPhish phishing intelligence. 2023 [cited 2026 Jan 1]. Available from: https://openphish.com/.
129. Dredze M, Gevaryahu R, Elias-Bachrach A. Learning fast classifiers for image spam. Brussels, Belgium: Council of European Aerospace Societies; 2007. p. 487–93.
130. Mohammad RM, Thabtah F, McCluskey L. An assessment of features related to phishing websites using an automated technique. In: Proceedings of the 2012 International Conference for Internet Technology and Secured Transactions (ICITST); 2012 Dec 10–12; London, UK. p. 492–7.
131. MillerSmiles. Phishing scams archive. 2023 [cited 2026 Jan 1]. Available from: http://www.millersmiles.co.uk/.
132. Cormack GV. TREC 2007 spam track overview. In: Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007); 2007 Nov 6–9; Gaithersburg, MD, USA. p. 1–9.
133. Apache SpamAssassin Project. SpamAssassin public corpus. 2005 [cited 2026 Jan 1]. Available from: https://spamassassin.apache.org/old/publiccorpus/.
134. Cohen WW. Enron email dataset. 2015 [cited 2026 Jan 1]. Available from: http://www.cs.cmu.edu/.
135. Androutsopoulos I, Koutsias J, Chandrinos KV, Paliouras G, Spyropoulos CD. Ling-spam dataset. 2000 [cited 2026 Jan 1]. Available from: https://www.kaggle.com/datasets/mandygu/lingspam-dataset.
136. Aziz RM, Baluch MF, Patel S, Ganie AH. LGBM: a machine learning approach for Ethereum fraud detection. Int J Inf Technol. 2022;14:3321–31.
137. Jin C, Zhou J, Xie C, Yu S, Xuan Q, Yang X. Enhancing Ethereum fraud detection via generative and contrastive self-supervision. IEEE Trans Inf Forensics Secur. 2025;20:839–53.
138. Chen T, Li Z, Zhang Y, Luo X, Chen A, Yang K, et al. DataEther: data exploration framework for Ethereum. In: Proceedings of the IEEE 39th International Conference on Distributed Computing Systems (ICDCS); 2021 Mar 4–5; Greater Noida, India; 2021. p. 1369–80.
139. Khalifa O, Nor THSBT, Ahmed MZ, El-Khazmi E, Esgiar AN. Blockchain based email security to mitigate phishing attack. Asian J Electr Electron Eng. 2024;4(2):77–86. doi:10.69955/ajoeee.2024.v4i2.73.
140. Pitre V, Joshi A, Das S. Blockchain and machine learning based approach to prevent phishing attacks. In: Proceedings of the 2023 3rd Asian Conference on Innovation in Technology (ASIANCON); 2023 Aug 25–27; Pune, India. p. 1–6.
141. Rathore S, Pan Y, Park JH. BlockDeepNet: a blockchain-based secure deep learning for IoT network. Sustainability. 2019;11(14):3974.