International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 07 | July - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Literature Survey on Website Phishing Detection Using Data Science Models and Leveraging Various Datasets

Pushpa[1], Dr.Arun mozhi Selvi[2]

[1]*Computer Science Department & British University College, AJMAN,UAE*
[2]*Computer Science Department & British University College, AJMAN,UAE*

Abstract - In the modern era of computing, driven by technological advancements, data and its attributes are fueling significant shifts in technology, particularly in operations related to data analysis, management, security, and maintenance. To protect crucial data and identify and understand security incidents, it is essential to extract patterns and insights from cybersecurity data. This literature survey highlights the growing body of research in phishing detection, emphasizes the dominance of threat detection techniques, and identifies specific datasets and algorithms that have shown promising results with high accuracy. Different phishing detection approaches, including Machine Learning, Generative models, hybrid models, List-Based Models, Visual Similarity Models, Heuristic Models, and Deep Learning-based techniques, are studied and compared. Data cleaning, data balancing, feature selection, and feature extraction were carried out during the model development process to build the most sustainable model. A study and analysis were conducted on various scientific papers published in the last two years in research journals, articles, conferences, technical workshops, technology books, and high-ranking business blogs. This study enhances readers' understanding of various phishing website detection techniques, the datasets used, and the comparative performance of the algorithms employed.

Key Words: Website phishing, machine learning, deep learning, data processing, data security

1.INTRODUCTION

Security remains one of the most critical challenges in the domain of the Internet and communication. Among various cyber threats, phishing is one of the most prevalent and harmful. It involves deceptive tactics designed to steal or misuse users' personal information, such as login credentials, identity details, passwords, and financial information. Phishers often replicate legitimate websites with high accuracy, making it difficult for users to distinguish between genuine and malicious sites. This deception can lead to severe consequences, including financial loss and identity theft. The 2025 Phishing Trends Report [1] is highlighted as a groundbreaking resource for understanding real malicious clicks and phishing attacks that bypass email filters, addressing a previous lack of such data. The core question driving the analysis is "Who's clicking on what?", emphasizing the human interaction with phishing attempts. Comcast Business Cybersecurity Threat Report indicates that 80-95% of these breaches originate from phishing attacks. SlashNext reports a staggering 4,151% increase in phishing attacks since the advent of ChatGPT in 2022.

The 2025 Data Breach Investigations Report from Verizon [2] states that around 30% of breaches were linked to third-party involvement, twice as many as last year, driven in part by the exploitation of vulnerabilities and business interruptions. There was a 34% increase in attackers exploiting vulnerabilities to gain initial access and cause security breaches compared to last year's report. It is estimated that 65% of target organizations and 35% of target individuals. The forms of attack can include data breaches, Ransomware attacks that encrypt data and systems, which may lead to operational disruption and financial demands, malicious software injection into the organization's network, and business disruption damages, such as downtime, reputational damage, and economic losses.

This year, the Verizon DBIR team analyzed 22,052 real-world security incidents, of which 12,195 were confirmed data breaches that occurred within organizations of all sizes and types. This represents the highest number of breaches ever analyzed in a single report. These incidents and breaches were sourced from the case files of the Verizon Threat Research Advisory Centre (VTRAC) team, with the generous support of our global contributors, and from publicly disclosed security incidents. Together, these attacks have affected victims from 139 countries worldwide.

Although the threat landscape can vary due to organizational size, mission, and location, there are always specific overarching themes that predominate our dataset, regardless of these variables. This year is no exception. The most notable among them is the role that third-party relationships play in the occurrence and prevention of breaches.

Advances in data science and technology have enabled more sophisticated cyberattacks and data breaches while also providing new tools to defend against them. While attackers leverage automation, AI, and big data for malicious purposes (e.g., phishing at scale, deepfakes,

TABLE - 1:DATA SCIENCE & MODELS TO TACKLE DATA BREACHES

	Model	Features	Techniques	Use case
1	Anomaly Detection Models	Detect deviations from normal behavior in: Network traffic User activity Application logs	Statistical models (e.g., z-score, IQR) Clustering (e.g., DBSCAN, k-Means) Autoencoders (deep learning) Isolation Forests	Spot unusual login times, excessive data downloads, or sudden privilege escalations.
2	Behavioral Analytics with Machine Learning	Building models that learn what "normal" behavior looks like for: Individual users Devices Applications. Models should be updated regularly to adapt to evolving behavior patterns and reduce false positives.	Random Forests, Gradient Boosting LSTM networks (for time-series behavior) Bayesian networks	Detect insider threats or compromised credentials.
3.	Natural Language Processing (NLP) for Threat Intelligence	Analyze phishing emails, dark web chatter, or logs. Classify malicious vs. benign communications.	BERT or GPT-based classifiers Named Entity Recognition (NER) to extract IOCs (Indicators of Compromise)	Early detection of phishing or social engineering campaigns.
4.	Predictive Modeling for Risk Scoring	Use historical incident and vulnerability data to predict: Which systems are most likely to be attacked Which vendors pose the highest risk Which misconfigurations are most critical	Inputs:CVSS scores Network topology Past breach data External threat intelligence	Prioritize patching and vendor assessments.
5.	Automated Threat Hunting	Data science pipelines can automate: Log aggregation Event correlation Signature-less detection (based on patterns instead of known attack signatures)	ELK Stack, Splunk + ML Toolkit, custom Python pipelines	Proactively identifies advanced persistent threats (APTs), insider threats, and unknown malware within the enterprise environment.

automated vulnerability discovery), defenders can also counter these threats using data-driven models.

Data science offers powerful tools to analyze vast volumes of security-related data, uncover hidden patterns, and predict potential breach scenarios before they occur. By leveraging machine learning, anomaly detection, behavioural analytics, and threat intelligence, organizations can develop predictive models that enhance breach detection, automate threat responses, and strengthen their cybersecurity defences. These models not only help identify vulnerabilities but also reduce incident response time and minimize damage.

There is a significant need for advanced phishing protections and effective cybersecurity training models to combat and mitigate phishing attacks. The good news is that phishing risk is measurably reducible through behaviour-based training. Employees can achieve a sixfold improvement in recognizing and reporting attacks within six months. Organizations may see up to an 86% decrease in phishing incidents as a result.

The results of calculating the level of cybercrime[6] from 2016 to 2023 showed its gradual growth worldwide. Thus, during the analyzed period, the rate of growth in cybercrime across individual countries worldwide exceeded 50%. This statistic highlights a critical need for organizations to enhance their vulnerability management programs, with a focus on the rapid identification, prioritization, and complete remediation of vulnerabilities in their perimeter devices. Failing to do so leaves them significantly exposed to the escalating cyber threat landscape.

2. Background

2.1 Phishing Attacks

Phishing is a form of cyberattack in which malicious actors deceive users into revealing sensitive information such as usernames, passwords, credit card numbers, or personal identification details. Of the many phishing strategies, website phishing stands out as both highly prevalent and widely exploited. In this attack, cybercriminals create counterfeit websites that closely mimic legitimate ones, often tricking users into entering their confidential data.

These phishing websites are typically distributed through email links, social media platforms, malicious advertisements, or even compromised legitimate websites. Attackers use tactics such as:

URL spoofing (e.g., using typosquatting or punycode domains)

HTTPS abuse (many phishing sites now use SSL to appear trustworthy)

Visual cloning of popular websites like banks, e-commerce platforms, or email providers.

Once the user submits data to the fake site, the attacker captures and uses it for financial fraud, identity theft, or further attacks such as account takeover and social

engineering. The rise in automated phishing kits, AI-generated content, and phishing-as-a-service platforms has enabled even individuals with limited technical skills to carry out attacks to deploy convincing phishing websites at scale. In response, cybersecurity systems must employ advanced detection techniques, including machine learning, visual similarity analysis, URL heuristics, and user behaviour profiling.

Due to its stealth and success rate, website phishing remains a significant threat to individuals and organizations alike. Combatting it requires a combination of technical defences, user education, and continuous monitoring.

Table -2: The Evolving Threat of Phishing

	Target	Tactics	Tools
1990s–2000s	Usernames, passwords	Fake login pages, deceptive email	Basic scripting, social engineering
Now-2025	Entire networks, financial data, intellectual property	Malware-laced attachments or links (drive-by downloads) Exploits (via zero-day vulnerabilities) Ransomware delivery Business Email Compromise (BEC)	AI-generated emails, domain spoofing, QR phishing, deepfake audio/video.

The Q4 2024 phishing trends report from the Anti-Phishing Working Group (APWG)[3] highlights the rapid evolution of phishing threats, both in scale and sophistication. Explosive Growth in Phishing Attacks: There were 989,123 phishing attacks in Q4, marking the highest quarterly total yet, up from 877,536 in Q2 and 932,923 in Q3.

Website phishing involves creating a malicious website that closely resembles a legitimate one— such as a bank, login portal, or online retailer — to deceive users into providing sensitive information(credentials, card numbers, etc.).

2.2 Website Phishing techniques

Phishing attacks continue to evolve in sophistication, leveraging new techniques and exploiting user trust in digital communication platforms. The following are some of the most notable trends observed in 2024–2025:

1. Use of Homoglyph Domains

Phishers increasingly use homoglyph attacks, where visually similar characters from different character sets are substituted to mimic legitimate domains. For example:go0gle[.]com instead of google[.]com (using zero instead of the letter "o"), microsoft[.]com (using a Cyrillic "o" instead of a Latin "o")

These domains can evade casual visual inspection and are particularly effective in email or mobile interfaces, where URLs are often truncated.

2. Abuse of Obscure and Cheap Top-Level Domains (TLDs)

Malicious actors prefer low-cost or poorly regulated TLDs such as: .top, .xyz, .club, .online, .cn

These TLDs often lack strict enforcement or monitoring, making them particularly susceptible to hosting phishing sites. They are also easy to register in bulk, enabling rapid domain cycling (fast-flux) to evade detection.

3. QR Code Phishing (Quishing)

Phishing via QR codes has surged, particularly as contactless technology becomes increasingly prevalent. Attackers embed malicious URLs in QR codes and distribute them via:

Emails and flyers

Fake parking or payment notices

Public spaces (e.g., stuck to restaurant menus, posters)

Since users cannot visually verify the destination URL before scanning, this method bypasses traditional link analysis and exploits user trust.

4. HTTPS-Enabled Phishing Sites

The majority of phishing sites today use SSL/TLS certificates to display the HTTPS padlock icon in browsers. This falsely assures users that the site is legitimate, as many still equate HTTPS with trustworthiness. Free services like Let's Encrypt are often abused to obtain certificates for fraudulent domains quickly.

The detection of phishing websites has received significant attention in the research community [4] and a substantial body of literature has addressed this critical and challenging cybersecurity problem. Researchers and practitioners have proposed a variety of detection techniques and are typically divided into three overarching categories based on their underlying approach and detection targets:

1. List-Based Approaches

These methods rely on blacklists (of known phishing domains) and whitelists (of verified legitimate sites). A user-requested URL is checked against these lists to determine its legitimacy. While list-based approaches are fast and straightforward to implement, they struggle to detect zero-day phishing attacks and new, previously unseen domains.

2. Similarity-Based Approaches

These techniques compare the suspicious website to a known legitimate site by analysing various elements, such as Visual appearance (e.g., logos, layout, colour schemes), URL structure and domain similarity, and Content features like brand names or login forms.

These methods are effective against spoofing and cloning attacks, where attackers mimic popular websites to deceive users. However, they can be computationally expensive and prone to false positives, especially when legitimate sites share design templates.

3. Machine Learning-Based Approaches

Machine learning (ML)-based solutions extract a diverse array of features from the website, including:

Lexical features (e.g., domain length, presence of special characters)

Host-based features (e.g., domain age, SSL certificate info)

Content-based features (e.g., number of input fields, use of scripts)

These features are fed into classification models such as Random Forests, Support Vector Machines, Neural Networks, or Ensemble Models to determine the likelihood that a website is malicious. ML-based detection can generalise to novel phishing attacks, but it requires labelled data and careful feature engineering.

Each of these features has its strengths and limitations. In practice, hybrid systems that combine multiple approaches (e.g., list-checking with ML classifiers) often achieve the best results by balancing accuracy, speed, and resilience against evasion techniques.

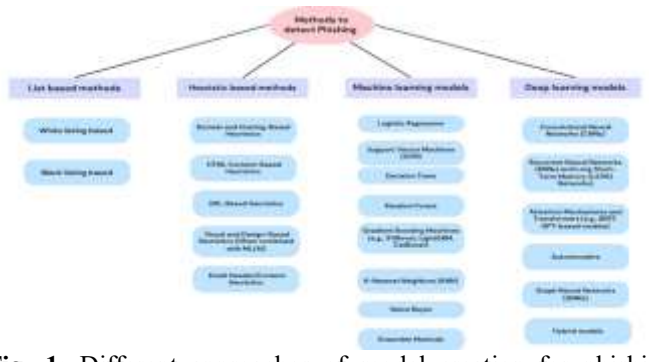

Fig -1: Different approaches of model creation for phishing detection

3. Background works

In paper [5], it is concluded that cyber-psychological manipulation tactics often precede technical attack vectors in cyberattacks. The study emphasizes that one of the leading mitigation strategies is comprehensive user training. Based on this insight, we recommend that cybersecurity stakeholders—including those in healthcare, education, government agencies, and business enterprises—invest in building both psychological awareness and technical competencies among their staff and customers. This can be achieved through structured on-the-job and off-the-job training programs. Such initiatives not only enhance the organization's resilience against cyber threats but also contribute positively to its overall reputation and trustworthiness.

In paper[7], they proposed a phishing website detector based on improving the convolutional neural network (CNN) with a self-attention mechanism. The proposed detector collects phishing Uniform Resource Locators (URLs) by treating them as strings. CNN models have proved their efficiency when dealing with text strings compared to Long Short-Term Memory (LSTM), which focuses on temporal features. Using CNN enables the learning of comprehensive features from URLs and facilitates the detection of phishing ones. Besides,the proposed detector was tested using unknown URLs and achieved excellent results. The improved CNN's detection precision of 99.7% is 2.74% higher than that of the regular CNN model. The reported results indicate that utilizing the self-attention mechanism has enhanced the detection accuracy and improved the efficiency of the CNN model in identifying phishing websites.

In the paper [8], the authors introduce a hybrid deep learning model that combines Gated Recurrent Units (GRUs) and Convolutional Neural Networks (CNNs) to enhance phishing URL detection. The fundamental aim was to integrate the strengths of both architectures:GRUs for capturing temporal dependencies in URL sequences and CNNs for extracting spatial and local features. This hybrid approach aims to develop a robust and comprehensive model that can effectively identify phishing attempts. The study evaluated three models—GRU, CNN, and the proposed GRU+CNN hybrid—using a Kaggle dataset comprising over 2.5 million labelled URL samples. The GRU model achieved an accuracy of 97.8%, whereas the CNN model slightly outperformed it with an accuracy of 98.0%. Notably,the hybrid GRU+CNN model demonstrated superior performance, attaining an accuracy of 99.0%,thereby underscoring its effectiveness in capturing both sequential and structural patterns inherent in phishing URLs. For future work,the authors plan to optimize the hybrid model for real-time phishing detection and explore its adaptability to other cybersecurity challenges, including malware classification and the detection of social engineering threats.

In paper [9], the authors present a comprehensive study on the application of Machine Learning techniques for phishing website detection, with an emphasis on enhancing both accuracy and computational efficiency. The proposed method combines the CfsSubsetEval attribute evaluator with the KMeans clustering algorithm to enhance detection performance. Testing of the approach was conducted on publicly accessible datasets with varying scales (2,000; 7,000 and 10,000 samples) to evaluate its robustness across different data scales. Experimental results showed that the proposed model achieved an accuracy of 89.2% on the 2,000-sample dataset, significantly outperforming the traditional kernel K-Means algorithm, which achieved an accuracy of only 51.5%. Additional evaluation using precision,recall, and F1-score metrics further validated the model's effectiveness. The study also explores the method's scalability and real-world applicability, acknowledging current limitations and suggesting avenues for future research. Overall,this work provides valuable insights into the development of efficient and adaptable phishing detection systems that can address the complex landscape of cyber threats.

In paper [11] ,the authors address the challenges of detecting phishing website by conducting comprehensive experiments using three ensemble learning classifiers:Random Forest (RF), Gradient Boosting (GB), and CatBoost (CATB). Recognizing that each classifier independently exhibits strong predictive capabilities,the study explores the effectiveness of hybrid ensemble architectures—specifically, stacking and majority voting—to enhance detection performance further. Given the computational cost commonly associated with ensemble methods, the study implements the Univariate Feature Selection (UFS) technique to reduce feature dimensionality and improve efficiency. To evaluate the scalability and consistency of the proposed models, experiments were conducted on three distinct phishing website datasets (DS-1, DS-2, and DS-3). Results show that the CatBoost (CATB) classifier consistently delivered superior accuracy across all datasets, achieving 97.9% on DS-1, 97.36% on DS2, and 98.59% on DS-3. The Random Forest (RF) classifier was the fastest in computational efficiency across all datasets, followed by CatBoost. These findings highlight that model hyperparameter tuning and the use of feature selection techniques, such as UFS, play a critical role in optimizing both accuracy and processing speed. The study concludes by identifying areas for future research,recommending the integration of deep learning algorithms, exploring mobile-based phishing scenarios, applying the approach to larger and more diverse datasets, and evaluating additional feature selection techniques to enhance model robustness and applicability.

In paper [12], the authors highlight the growing threat posed by malicious online attacks, with increasingly sophisticated techniques being used to deceive users. This study investigates explicitly the role of feature selection in enhancing the performance of phishing URL detection systems. Feature selection is a crucial preprocessing step in both machine learning (ML) and deep learning (DL), as it helps identify the most relevant attributes,thereby improving model accuracy and computational efficiency. The research evaluates multiple feature selection techniques across five diverse datasets, employing methods such as Random Forest (RF) Select-fromModel, SelectKBest using the chi-square statistic, Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE). Among these, PCA showed powerful results on the fourth dataset. Remarkably, all four classifiers —Random Forest, Decision Trees (DTs), XGBoost, and Multilayer Perceptron (MLP)—achieved 100% accuracy in phishing URL detection when combined with the selected features. These results underscore the significant impact of effective feature selection in improving phishing detection accuracy and efficiency across varied datasets. The study demonstrates how techniques like PCA can lead to optimal model performance and contribute to a deeper understanding of how feature engineering supports cybersecurity applications.

Despite some limitations, such as the risk of overfitting and the need for validation on more diverse real-world datasets,the findings strongly support the practical applicability and robustness of the proposed approach for detecting phishing threats in real-life scenarios.

In paper [13], the authors assess the effectiveness of applying machine learning (ML) techniques for phishing website classification, positioning the problem within the broader context of intrusion detection systems in cybersecurity. The study distinguishes between single classifiers and ensemble learners, emphasizing the latter as more promising due to their ability to improve detection accuracy and reduce variance. To enhance the performance of ML-based detection models,the study incorporates two key strategies:feature selection (via feature importance) and hyperparameter tuning. Two ensemble algorithms—Random Forest and Extra Trees—were employed to build phishing classification models. Both models were optimized using feature importance-based attribute selection and rigorous hyperparameter tuning. The findings suggest that the Random Forest-based model achieves a modest performance advantage over the Extra Trees model.

The study concludes that combining ensemble learners with attribute selection and hyperparameter tuning significantly improves classification performance. These findings reinforce the effectiveness of such techniques for building reliable and accurate phishing detection systems.

In paper [14], the authors propose a novel phishing detection framework by integrating the SMOTETomek resampling technique with the XGBoost (XGB) classifier. SMOTETomek is a hybrid data balancing approach that combines the strengths of SMOTE (Synthetic Minority Over-sampling Technique) and Tomek Links. This method simultaneously addresses class imbalance by oversampling

the minority class and applying dataset enhancement techniques by removing borderline and noisy instances,thereby improving the quality of training data. The proposed SMOTETomek-XGBoost model is evaluated against traditional classifiers and consistently outperforms them across various performance metrics, including accuracy, precision, recall, F1 Score, and ROC AUC. The study demonstrates that this hybrid approach significantly enhances phishing detection performance, offering a more effective solution for identifying online threats and improving cybersecurity readiness. The authors suggest that future work could involve integrating advanced feature engineering strategies or additional ensemble learning methods to further enhance model robustness and adaptability in real-world cybersecurity scenarios.

Paper [15] explores URL-based phishing detection using both classical machine learning and advanced deep learning (DL) approaches. The paper is essentially divided into two major investigations:

Phishing Website Detection Based on URL Features (Machine Learning Focus) This part of the study focuses on extracting specific URL characteristics— including lexical, structural, and statistical features—to train and evaluate traditional ML classifiers. Among various algorithms tested,the Random Forest (RF) classifier achieved the best performance:

Accuracy: 98.23%

False Positive Rate: Lowest among tested models

The use of feature selection techniques further optimized performance. This approach does not rely on website content or blacklists, making it fast, scalable, and suitable for real-time detection systems.

Deep Learning Approaches for Phishing Website Detection

To further enhance detection,the paper investigates deep learning models using a dataset containing both phishing and legitimate URLs. Features extracted include Structural patterns of URLs and semantic cues embedded within URL strings. A hybrid deep learning (DL) model utilizing Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) was developed to learn both spatial and sequential dependencies. Key outcomes include:

Actual Positive Rate: >90%

False Positive Rate: Minimal

Generalization Capability: High across varied phishing examples

This demonstrates that deep learning frameworks are highly effective in capturing complex URL patterns and significantly boost phishing detection accuracy.

Paper [16] addresses the persistent threat of phishing attacks by proposing a comprehensive and high-performance detection framework based on the XGBoost (Extreme Gradient Boosting) algorithm. Recognized for its robustness and scalability in classification tasks, XGBoost is employed to differentiate between phishing and legitimate websites using a rich set of extracted features.

Key Contributions and Methodology

Feature Engineering: The framework extracts a diverse set of features from both the URL and website content, including:

Lexical features (e.g., URL length, special characters)

Structural attributes (e.g., presence of subdomains, use of HTTPS)

Host-based characteristics (e.g., domain age, WHOIS info)

Content similarity to known phishing pages

Modelling with XGBoost: The algorithm is trained using labelled datasets that contain both phishing and legitimate URLs.

To optimize model performance, hyperparameter tuning is conducted.

XGBoost boosting strategy combines multiple weak learners (decision trees) to form a strong predictive model.

Evaluation: Experiments were conducted on publicly available phishing datasets containing thousands of samples.

Performance metrics include accuracy, precision,recall, and F1-score.

The model demonstrates: High classification accuracy Low false positive rate Superior performance compared to traditional classifiers (e.g., SVM, Decision Trees). Paper [17] introduces an innovative framework titled XAIAOA-WPC (Explainable Artificial Intelligence with Aquila Optimization Algorithm for Web Phishing Classification), designed to enhance the classification and explainability of phishing attacks within Cyber-Physical Systems (CPS). The proposed model addresses both the technical detection of phishing websites and the interpretability of the decision-making process.

Key Components of the XAIAOA-WPC Framework

Three-Level Preprocessing Pipeline

Data Cleaning: Noise and inconsistencies are eliminated from the input data to improve quality.

Text Preprocessing: Tokenization, stop-word removal, and normalization of textual components in URLs and web content.

Standardization: Normalizing data to ensure model compatibility and consistency. Feature Selection Using HHO (Harris' Hawks Optimization): The HHO-FS algorithm identifies optimal feature subsets to improve model performance and reduce redundancy.

MHA-LSTM Model (Multi-Head Attention-based Long Short-Term Memory): Leverages LSTM for sequential pattern learning. Integrates multi-head attention to focus on relevant parts of the input, improving phishing pattern recognition.

Aquila Optimization Algorithm (AOA): Further fine-tunes the output of the MHA-LSTM by optimizing hyperparameters and adjusting the decision boundaries.

Explainability Layer Using LIME (Local Interpretable Model-Agnostic Explanations): Provides human-interpretable explanations for model decisions. Enhances trust and transparency in AI-driven cybersecurity tools.

Performance & Evaluation Dataset: Evaluated using a benchmark phishing website dataset.

Accuracy: Achieved 99.29%, outperforming several state-of-the-art methods.

In paper [18], the authors address the challenge of enhancing human technology integration in the detection of phishing emails, highlighting the limitations of approaches that focus solely on technology or human factors. They propose and evaluate a prototype visual risk indicator that conveys differentiated risk ratings in an accessible format, empowering users to identify phishing attempts. Drawing on the Protection Motivation Theory (PMT) and the Heuristic-Systematic Model (HSM),the study develops a research model tested through a preliminary online survey and a main eye-tracking experiment. The results demonstrate that both implicit and explicit visual cues significantly affect user information processing,that visual risk indicators effectively guide decision-making and risk discrimination, and that decisionmaking anomalies can arise in situations of conflicting signals. The findings support the practical feasibility of integrating visual risk indicators into email interfaces to mitigate phishing risks, providing both theoretical insights for cybersecurity research and practical implications for designing adequate security warnings in organizational contexts.

Paper [19] investigates the effectiveness of a Bidirectional Gated Recurrent Unit (BiGRU) model combined with feature selection techniques for detecting phishing websites. Using the Phishing Websites dataset sourced from the UCI Machine Learning Repository, the study involves data cleaning, preprocessing, feature normalization, and selection of the most relevant attributes for classification. The BiGRU model, which excels at capturing temporal dependencies in sequential data, is applied to the task. The evaluation employs five-fold crossvalidation to ensure robustness. Experimental results demonstrate outstanding performance, with mean accuracy, precision,recall, F1 score, and AUC all reaching 1.0, indicating the model's exceptional capability in distinguishing phishing from legitimate websites. The study underscores the potential of integrating BiGRU architectures with feature selection and rigorous validation methods to build precise phishing detection systems. The authors suggest that future work could focus on optimizing model parameters, exploring alternative deep learning architectures, and combining these methods with quantum computing approaches. Additionally, external validation and further evaluation under varied real-world conditions are recommended to confirm the generalizability of the findings.

Paper [20] proposes a phishing detection system that integrates behavioral analysis, OCR, and NLP under an Explainable Machine Learning (XML) framework. The system employs supervised learning models, including ensemble methods,for accurate classification alongside anomaly detection techniques to enable adaptive learning. To enhance model transparency and interpretability, explainability tools such as SHAP and LIME are utilized, facilitating cyber security experts' understanding of model decisions. Experimental results demonstrate high detection accuracy, adaptability, and improved reliability compared to traditional methods. This approach provides a robust solution for cybersecurity resilience by enabling real-time phishing detection, alerting, and continuous adaptive learning,thereby enhancing organizational phishing defences.

Paper [21] proposes three deep learning-based techniques for phishing website detection: Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and a hybrid LSTM–CNN model. The experimental findings indicate that the CNN-based approach achieves the highest accuracy at 99.2%, outperforming the LSTM–CNN (97.6%) and LSTM (96.8%) models. The findings demonstrate the effectiveness of CNN for phishing detection on the evaluated dataset. The study notes variability in performance across the models and highlights the superiority of CNN in accuracy. Future work aims to optimize the training process by reducing training time and enhancing feature engineering to improve overall detection accuracy. Additionally,the authors plan to develop methods that incorporate both webpage content and URL context to enhance phishing detection capabilities further.

4. EVALUATION METRICS

The survey on website phishing detection has compared several detection techniques. Hence, it is helpful to introduce the evaluation metrics used in the phishing literature. In the case of a binary classification problem, where we detect websites as phishing or legitimate instances, only four classification possibilities exist—usually represented using the confusion matrix.

Let's denote the four outcomes from the confusion matrix as :

TP (True Positives) = NP→P (Correctly classified phishing instances)

FP (False Positives) = NL→P (Legitimate instances incorrectly classified as phishing)

FN (False Negatives) = NP→L (Phishing instances incorrectly classified as legitimate)

TN (True Negatives) = NL→L (Correctly classified legitimate instances)

TABLE -3: CLASSIFICATION CONFUSION MATRIX

	Classified as phishing	Classified as legitimate
Is phishing	$N_{P\to P}$	$N_{P\to L}$
Is legitimate	$N_{L\to P}$	$N_{L\to L}$

Based on our review of the literature,the following are the most commonly used evaluation metrics:

Accuracy (ACC)

Definition: The proportion of total predictions that were correct (both true positives and true negatives). It measures the overall correctness of the model.

Formula: Accuracy= $\frac{TP+TN}{TP+TN+FP+FN}$

Phishing Context: While seemingly straightforward, accuracy can be misleading in phishing detection due to dataset imbalance. If a dataset consists of 99% legitimate sites and only 1% phishing sites, a model that classifies everything as legitimate would achieve 99% accuracy; however, it would fail to detect any phishing attacks (resulting in 100% false negatives for phishing).Hence, relying solely on accuracy is often insufficient.

Precision (PR)

Definition: The proportion of predicted positive instances that were correct positives. It answers the question: "Of all the websites the model flagged as phishing, how many were phishing?"

Formula: Precision= $\frac{TP}{TP+FP}$

Phishing Context: High precision is vital in phishing detection. A low precision means a high number of false positives (NL→P), which can lead to legitimate websites being blocked, user frustration, and distrust in the security system. It prioritizes minimizing false alarms.

Recall (RC) / Sensitivity / True Positive Rate (TPR)

Definition: The proportion of actual positive instances that the model precisely identified. It answers the question: "Of all the actual phishing websites, how many did the model correctly detect?" Formula:Recall= $\frac{TP}{TP+FN}$

Phishing Context: High recall is highly critical in phishing detection. A low recall corresponds a high number of false negatives (NP→L), indicating that many real phishing attacks are missed by the system, directly exposing users to harm. It prioritizes minimizing missed threats.

F1-Score

Definition: The harmonic mean of Precision and Recall. It provides a single metric that balances both precision and recall, which is especially useful when there is an uneven class distribution (imbalanced dataset). It penalizes models that have perfect precision but poor recall, or vice versa.

Formula: F1-Score=2× $\frac{Precision*Recall}{Precision+Recall}$

Phishing Context: The F1-Score is often considered a more robust metric than accuracy for phishing detection because it directly accounts for both false positives and false negatives. A high F1 score indicates a good balance between accurately detecting most phishing attacks and minimizing false alarms.

Specificity / True Negative Rate (TNR)

Definition: The proportion of actual negative instances (legitimate websites) that were correctly identified as negative. It answers: "Of all the actual legitimate websites, how many did the model correctly identify as legitimate?

Formula: Specificity= $\frac{TN}{TN+FP}$

Phishing Context: High specificity is desirable as it indicates the model is good at not flagging legitimate sites as malicious. It is the complement of the False Positive Rate (FPR=1−Specificity).

False Positive Rate (FPR) / Fall-out

Definition: The proportion of actual negative instances that were incorrectly classified as positive. Formula: FPR=$\frac{FP}{TN+FP}$

Phishing Context: This metric directly reflects the rate of false alarms. Minimizing false positives (FPR) is crucial for both user experience and system efficiency.

5. Research issues and future directions

Our literature review highlights several critical challenges and open research issues in the domain of phishing website detection and its prevention. These challenges span the entire pipeline from data collection to output interpretation—and are summarized as follows: Datasets: The availability and quality of datasets remain a significant bottleneck. Many existing phishing datasets are outdated and may not reflect current phishing tactics and behavioural patterns. This limits the effectiveness and generalizability of detection models. There is a pressing need for up-todate, diverse, and representative datasets that capture the evolving nature of phishing attacks. Feature Engineering: Identifying and extracting meaningful features from website data is both crucial and challenging. Effective feature engineering requires deep domain knowledge and adaptability to evolving phishing techniques. Ongoing research is being conducted into automated and context aware feature extraction methods to enhance detection performance. Data Balancing: Phishing datasets are typically imbalanced, with a significantly higher number of legitimate samples than phishing samples. This imbalance can bias models and reduce their effectiveness. Techniques such as undersampling, over-sampling (e.g., SMOTE), and synthetic data generation are

commonly used, but each introduces its limitations and trade-offs. Hyperparameter Tuning: The optimal configuration of model hyperparameters is essential for maximizing performance. However, hyperparameter tuning is often computationally expensive and time-consuming. Automated and efficient optimization strategies (e.g., grid search,random search, Bayesian optimization) remain an active area of exploration. Model Selection: Choosing the most appropriate modelling approach— whether traditional machine learning, hybrid models, deep learning, ensemble techniques, or extensible architectures—is a key challenge. Each phishing detection approach has distinct strengths and limitations in terms of interpretability, scalability, and adaptability to novel phishing strategies. Output Analysis and Parameter Selection: Analyzing model outputs and selecting performance metrics (e.g., accuracy, recall, F1 score, precision, AUC) are crucial for evaluating and comparing detection systems. More attention is needed on context-specific metrics and interpretability tools that provide actionable insights, especially for real-time and high-risk environments.

6. Conclusion

The comprehensive analysis of various data science methodologies, particularly machine learning, has significantly enhanced phishing website detection by enabling the development of automated systems that can analyze website content, URLs, and other features to identify potential threats. These systems are capable of adapting to the continuously evolving tactics used in phishing attacks, detect zero-hour attacks, and reduce false positive rates. The study also highlights the crucial role of selecting a feasible dataset for training and testing, as well as employing feature engineering and data balancing techniques to overcome undersampling and oversampling. From traditional phishing detection methodologies to cutting-edge models, researchers have explored various systems to prevent phishing attacks effectively.

The growing sophistication and frequency of phishing attacks emphasize the importance of advanced, adaptive, and scalable detection systems. This review examines a range of detection methodologies, including list-based, heuristic, machine-learning, ensemble and deep-learning approaches, highlighting their respective strengths and limitations. Empirical evidence suggests that machine learning models, such as XGBoost and random forest, achieve high performance when combined with data balancing techniques across diverse datasets. Deep learning architectures, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), demonstrate notable effectiveness in capturing complex phishing patterns due to their hierarchical learning capabilities. However, deploying such models in real-world environments remains a significant challenge. Issues related to computational overhead, scalability, and adaptability to continuously evolving phishing techniques must be addressed to ensure practical and sustainable implementation.

REFERENCES

1. https://hoxhunt.com/guide/phishing-trends-report
2. https://www.verizon.com/business/resources/reports/dbir/
3. https://apwg.org/trendsreports/
4. Zieni, Rasha & Massari, Luisa & Calzarossa, Maria Carla. (2023). Phishing or Not Phishing? A Survey on the Detection of Phishing Websites. IEEE Access. 11. 18499-18519. 10.1109/ACCESS.2023.3247135.
5. Bundala, Ntogwa. (2025). Detecting and Mitigating Cyber-Psychological Tricks and Cyber-Technical Tricks in Cyberattacks.
6. Kuzior, Aleksandra & Tiutiunyk, Inna & Zielińska, Anetta & Kelemen, Roland. (2024). Cybersecurity and cybercrime: Current trends and threats. JOURNAL OF INTERNATIONAL STUDIES. 17. 220-239. 10.14254/2071-8330.2024/17-2/12.
7. Said, Yahia & Alsheikhy, Ahmed & Lahza, Husam & Shawly, Tawfeeq. (2024). Detecting phishing websites through improving convolutional neural networks with Self-Attention mechanism. Ain Shams Engineering Journal. 102643. 10.1016/j.asej.2024.102643.
8. M, Sangeetha & K, Navaz & Ravva, Santosh & Rudra, Roopa & Balaji, Penubaka & T, Ravi. (2025). Enhanced Phishing URL Detection Using a Novel GRU-CNN Hybrid Approach. Journal of Machine and Computing. 089-101. 10.53759/7669/jmc202505007.
9. al Sabbagh, Abdallah & Hamze, Khalil & Khan, Samiya & Elkhodr, Mahmoud. (2024). An Enhanced K-Means Clustering Algorithm for Phishing Attack Detections. Electronics. 13. 10.3390/electronics13183677.
10. Shafin, Sakib. (2024). An Explainable Feature Selection Framework for Web Phishing Detection with Machine Learning. Data Science and Management. 10.1016/j.dsm.2024.08.004.
11. Adane, Kibreab & Beyene, Berhanu & Yimer, Mohammed. (2023). Single and Hybrid-Ensemble Learning-Based Phishing Website Detection: Examining Impacts of Varied Nature Datasets and Informative Feature Selection Technique. Digital Threats: Research and Practice. 4. 10.1145/3611392.
12. Preeti, Preeti & Sharma, Priti. (2024). Enhancing phishing URL detection through comprehensive feature selection: a comparative analysis across diverse datasets. Indonesian Journal of Electrical Engineering and Computer Science. 36. 1182. 10.11591/ijeecs.v36.i2.pp1182-1188.
13. Gbolagade, Morufat & Oyelakin, Akinyemi & Ogundele, Temitope & Auwal, Jibrin & Akanni, Oluseye. (2023). Efficient Ensemble-based Phishing Website Classification Models using Feature Importance Attribute Selection and Hyper parameter Tuning Approaches. Journal of Information Technology and Computing. 4. 1-10. 10.48185/jitc.v4i2.891.
14. Omari, Kamal & Oukhatar, Ayoub. (2025). Advanced Phishing Website Detection with SMOTETomek-XGB: Addressing Class Imbalance for Optimal Results. Procedia Computer Science. 252. 289-295. 10.1016/j.procs.2024.12.031.
15. Kumar, A. & Prathiba, A. & Ashritha, A. & Reddy, N. & Shiny, Dr. (2025). Phishing Website Detection Based on URL Features. International Journal Of Scientific Research In Engineering & Technology. 73-78. 10.59256/ijsreat.20250502011.
16. Kumavat, Aditya. (2025). Phishing URL and Website Detection using MI. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 09. 1-9. 10.55041/IJSREM41473.
17. Alotaibi, Refa & Alkahtani, Hend & Aljebreen, Mohammed & Alshuhail, Asma & Saeed, Muhammad & Ebad, Shouki & Almukadi, Wafa & Alotaibi, Moneerah. (2024). Explainable artificial intelligence in web phishing classification on secure IoT with cloud-based cyber-physical systems. Alexandria Engineering Journal. 110. 16. 10.1016/j.aej.2024.09.115.
18. Baltuttis, Dennik & Teubner, Timm. (2024). Effects of Visual Risk Indicators on Phishing Detection Behavior: An Eye-Tracking

Experiment. Computers & Security. 144. 103940. 10.1016/j.cose.2024.103940.

19. Setiadi, De Rosal Ignatius Moses & Widiono, Suyud & Safriandono, Achmad & Budi, Setyo. (2024). Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection. Journal of Future Artificial Intelligence and Technologies. 1. 10.62411/faith.2024-15.
20. Himani, Pandya. (2025). Detection of Phishing Websites and Emails Using Explainable Machine Learning Models. Journal of Information Systems Engineering and Management. 10. 978-983. 10.52783/jisem.v10i40s.7545.
21. Alshingiti, Zainab & Alaqel, Rabeah & Al-Muhtadi, Jalal & Haq, Qazi & Saleem, Kashif & Faheem, Muhammad. (2023). A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN. Electronics. 12. 232. 10.3390/electronics12010232.