Suryavanshi Aditi Krushnakant, 2026, 14:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Phishing Attack Detection And Cybersecurity Defense Mechanisms: A Review Of Url, Content And Visual Machine Learning Approaches

Suryavanshi Aditi Krushnakant[1]; Prof. S. R. Borhade[2]; Prof. M. C. Kolse[3]; Prof. S. K. Said[4]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Phishing attacks use fake emails, messages and websites to trick people into revealing passwords, banking details and other personal data, causing financial losses and privacy violations. Attackers continuously create new sites and adapt their techniques, so blocklists and static rules often fail against advanced and short-lived phishing campaigns. This paper reviews phishing detection with emphasis on website and URL analysis, and places detection within a broader defence-in-depth strategy. It surveys feature-based approaches that analyse URL structure, domain information, page content and HTML elements, including the influential phishing website feature set of Mohammad et al., and machine learning classifiers such as Random Forest, Decision Tree, Support Vector Machine and Logistic Regression. Deep learning on raw URL strings and visual similarity methods that compare page screenshots and logos against legitimate brands are examined. Defence mechanisms beyond detection, including browser warnings, email authentication, phishing-resistant multi-factor authentication and user awareness training, are discussed. A layered detection and defence architecture is proposed, along with research gaps, objectives and a comparison of algorithms. The review concludes that combining automated multi-view detection with resistant authentication and informed users provides the most effective protection against phishing.

Index Terms—Phishing attacks, phishing detection, cybersecurity, machine learning, random forest, decision tree, support vector machine, malicious URLs, user awareness

I. INTRODUCTION

Phishing is one of the oldest and still most effective cyberattacks. A typical attack sends a message that appears to come from a bank, online store or government agency, urging the recipient to click a link and log in to a convincing but fake website. Reports from the Anti-Phishing Working Group [1] show that the number of unique phishing sites remains very high, and many exist for only hours before being replaced.

Traditional defences rely on blocklists of known phishing URLs. Because new sites appear constantly, there is always a window in which a site is active but not yet listed. Machine learning can examine the characteristics of a URL or page and decide whether it is likely to be phishing even if it has never been seen before.

This paper reviews phishing detection techniques, focusing on websites and URLs, and discusses complementary defence mechanisms. A layered architecture is proposed and research gaps are identified.

II. RELATED WORK

Zhang *et al.* [2] proposed CANTINA, which extracts important terms from a page using TF-IDF, searches for them, and judges the page legitimate if its domain appears among the top results; it was one of the first content-based detectors. Ma *et al.* [3] showed that lexical features of URLs together with host-based features such as registration data can classify malicious sites accurately without visiting them.

Mohammad *et al.* [4] studied features that distinguish phishing websites, such as use of an IP address in the URL, long URLs, the "@" symbol, prefixes and suffixes separated by hyphens, sub-domain count, SSL certificate status, domain age, abnormal anchors and request URLs, and pop-up windows, and built intelligent rule-based classifiers. In related work [5], they predicted phishing websites with a self-structuring neural network that adapts its structure during training. Their feature set and dataset became widely used benchmarks.

© 2026 Suryavanshi Aditi Krushnakant, This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Sahingoz *et al.* [6] extracted natural-language-processing features from URLs and compared several classifiers on a large dataset, finding that Random Forest [7] delivered the best accuracy. URLNet [8] learned URL representations with character- and word-level convolutional networks, removing the need for manual features. Sahoo *et al.* [9] surveyed malicious URL detection and categorized features and algorithms. To counter pages that copy the appearance of real sites, VisualPhishNet [10] learned a similarity metric between screenshots to detect zero-day phishing pages, and Phishpedia [11] identified brand logos on a page and checked whether the domain matched the brand. Kumaraguru *et al.* [12] showed that embedded training improves users' ability to recognize phishing.

Table I summarizes the reviewed literature.

Ref.	Method	Key Findings	Limitations
[2]	CANTINA TF-IDF + search engine	Early content-based phishing detection	Depends on search engine queries
[3]	Lexical and host-based URL features	Detected malicious URLs beyond blacklists	Host lookups add delay
[4]	Rule-based phishing website features	Influential feature set and classifiers	Some features need external lookups
[5]	Self-structuring neural network	Adaptive network for phishing prediction	Tuning complexity
[6]	NLP features + several classifiers	Random Forest achieved best accuracy	URL-only analysis
[8]	URLNet character/word CNN	No manual feature engineering	Needs large training data
[10]	VisualPhishNet screenshot similarity	Detects zero-day look-alike pages	Needs rendering; compute cost
[11]	Phishpedia logo-brand matching	Explainable brand-domain mismatch	Brand list must be maintained
[12]	Embedded anti-phishing training	Improved user recognition	Effect may fade over time

III. PRELIMINARIES

1. Anatomy Of A Phishing Attack

A phishing attack involves a lure (email, SMS, social-media message or advertisement), a deceptive URL, a fake website that imitates a trusted brand, and a mechanism to collect submitted data. Dhamija *et al.* [13] showed that users are often deceived because they rely on visual appearance and overlook security cues. Fig. 1 classifies detection approaches.

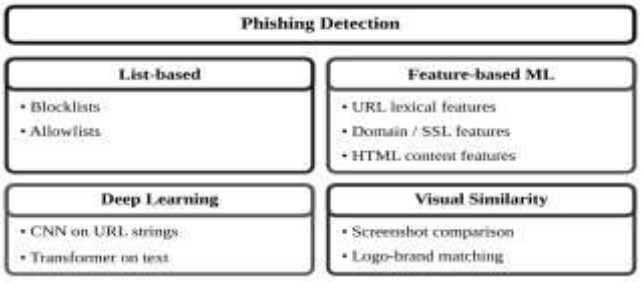


Fig. 1. Classification of phishing detection approaches.

2. Detection Approaches

List-based methods check URLs against known bad lists. Heuristic and feature-based methods extract characteristics and apply rules or ML. Visual similarity methods compare a page's appearance with protected brands. Hybrid systems combine these.

IV. PROBLEM STATEMENT AND OBJECTIVES

1. Identified Research Gaps

The following gaps are identified:

- Many models rely on features such as domain age or search rankings that are slow to obtain for real-time protection.
- Attackers increasingly use HTTPS, URL shorteners and legitimate hosting, weakening classic URL features.
- Datasets are quickly outdated, and models degrade as phishing techniques evolve.
- SMS and messaging-app phishing, common in India, receives less research attention than email and web phishing.
- Integration of detection with authentication defences and user training is seldom evaluated as a whole.

2. Objectives Of The Review

The objectives of this review are:

- To understand the working of phishing attacks and their delivery channels.
- To review URL, content and visual features used for phishing detection.
- To compare Random Forest, Decision Tree, SVM and Logistic Regression and deep learning models.
- To propose a layered phishing detection and defence architecture.
- To discuss defence mechanisms and future research directions.

3. Proposed Framework

The proposed architecture in Fig. 2 analyses links at the moment a user is about to visit them and combines detection with preventive controls.

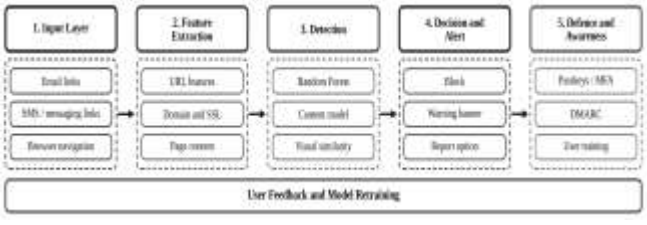


Fig. 2. Proposed layered phishing detection and defence architecture.

A. Input Layer

Links are captured from emails, SMS, messaging apps and browser navigation through a browser extension or a mail and messaging gateway.

B. Feature Extraction Layer

URL features include length, number of dots and hyphens, presence of IP addresses or "@", use of URL shorteners and suspicious keywords. Domain features include age, registrar and certificate details. Content features include forms that request passwords, external form actions, hidden elements and mismatched links.

6. Detection Layer

A Random Forest classifies URL and domain features quickly. If the score is uncertain, the page is rendered in a sandbox and checked by content and visual similarity models.

7. Decision And Alert Layer

High-risk links are blocked and the user sees a clear warning explaining the reason; medium-risk links show a caution banner. Users can report false positives.

8. Defence And Awareness Layer

Phishing-resistant authentication such as WebAuthn [14] prevents stolen passwords from being reused, email authentication with DMARC [15] reduces spoofing, and periodic simulated phishing training builds user awareness.

V. WORKING METHODOLOGY

The detection workflow is shown in Fig. 3. A Random Forest with T trees predicts the class by majority vote:

$$\hat{y} = mode\{ h_1(x), h_2(x), \ldots, h_T(x) \} \quad (1)$$

where $h_t(x)$ is the prediction of tree t for feature vector x. Each tree split is chosen to reduce Gini impurity

$$Gini = 1 - \Sigma_k \, p_k^2 \quad (2)$$

where p_k is the proportion of class k at a node. Performance is evaluated with accuracy, precision, recall, F1-score and false-positive rate, with emphasis on keeping false positives low so that legitimate sites are not blocked.

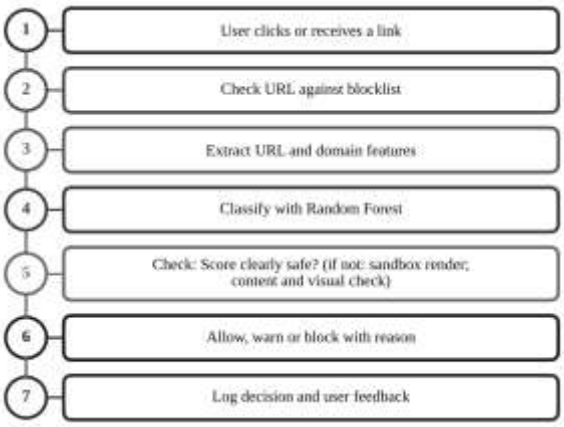


Fig. 3. Phishing URL detection workflow.

VI. DISCUSSION

1. Comparison Of Approaches

Table II compares classifiers commonly used on phishing feature datasets. Random Forest generally achieves the best balance of accuracy and robustness; Decision Trees provide readable rules; SVM performs well on smaller datasets; Logistic Regression is a simple, interpretable baseline; and deep models on raw URLs avoid manual feature design.

Classifier	Accuracy	Speed	Interpretability
Logistic Regression	Good	Very fast	High
Decision Tree	Good	Very fast	High
SVM	High	Medium	Low
Random Forest	Very high	Fast	Medium
CNN on URLs	Very high	Medium	Low

TABLE II Comparison of Classifiers for Phishing Detection

2. Open Challenges And Future Work

Table III summarizes key challenges. Future work should address phishing generated with AI tools that produces fluent, personalized messages, QR-code and SMS phishing, on-device detection that protects privacy, and continuous learning from user reports. Wider adoption of passkeys will reduce the value of stolen passwords, shifting attackers toward other forms of social engineering.

Challenge	Mitigation
Short-lived sites	Real-time ML instead of blocklists only
HTTPS phishing	Content and visual features, not padlock alone
Evolving tactics	Continuous retraining with fresh data
False positives	Warnings rather than blocks for medium risk
Credential theft	Passkeys and phishing-resistant MFA

TABLE III Challenges and Mitigation

VII. CONCLUSION AND FUTURE WORK

Phishing remains a major cybersecurity threat because it exploits human trust rather than technical weaknesses. The reviewed research, including the influential feature-based work of Mohammad *et al.*, shows that machine learning classifiers such as Random Forest can detect phishing websites with high accuracy from URL, domain and content features, while deep and visual models extend detection to new and well-disguised pages. Detection alone is not enough: phishing-resistant authentication, email authentication and user awareness complete the defence. A layered approach combining these mechanisms can significantly reduce phishing risk and protect sensitive information.

1. Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

REFERENCES

1. Anti-Phishing Working Group, "Phishing activity trends report," APWG, quarterly reports. [Online]. Available: https://apwg.org/trendsreports/
2. Y. Zhang, J. I. Hong, and L. F. Cranor, "CANTINA: A content-based approach to detecting phishing web

sites," in Proc. 16th Int. Conf. World Wide Web (WWW), 2007, pp. 639–648.

3. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs," in Proc. 15th ACM SIGKDD, 2009, pp. 1245–1254.
4. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Intelligent rule-based phishing websites classification," IET Inf. Secur., vol. 8, no. 3, pp. 153–160, 2014.
5. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Neural Comput. Appl., vol. 25, no. 2, pp. 443–458, 2014.
6. O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Syst. Appl., vol. 117, pp. 345–357, 2019.
7. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
8. H. Le, Q. Pham, D. Sahoo, and S. C. H. Hoi, "URLNet: Learning a URL representation with deep learning for malicious URL detection," arXiv:1802.03162, 2018.
9. D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL detection using machine learning: A survey," arXiv:1701.07179, 2017.
10. S. Abdelnabi, K. Krombholz, and M. Fritz, "VisualPhishNet: Zero-day phishing website detection by visual similarity," in Proc. ACM CCS, 2020, pp. 1681–1698.
11. Y. Lin et al., "Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages," in Proc. 30th USENIX Security Symp., 2021, pp. 3793–3810.
12. P. Kumaraguru et al., "Protecting people from phishing: The design and evaluation of an embedded training email system," in Proc. CHI, 2007, pp. 905–914.
13. R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in Proc. SIGCHI Conf. Human Factors Comput. Syst. (CHI), 2006, pp. 581–590.
14. W3C, "Web Authentication: An API for accessing public key credentials, Level 2," W3C Recommendation, Apr. 2021.
15. M. Kucherawy and E. Zwicky, "Domain-based Message Authentication, Reporting, and Conformance (DMARC)," IETF RFC 7489, 2015.