Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 01, 2025

CyberFence: Intelligent Defense Against Phishing Links

[1]Prof.K.N.Agalave, [2]Anushka Bhosale, [3]Neha Chaugule, [4]Arya Nanaware, [5]Monika Patule

[1 2 3 4 5]S.B.Patil.College Of Engineering, Indapur [1,2,3,4,5]

Email: Kimaya8890@gmail.com[1] , anushkabhosale99@gmail.com[2] , chauguleneha72@gmail.com[3] , aryananaware2074@gmail.com[4] , monikapatule@gmail.com[5]

Peer Review Information

Submission: 11 Sept 2025

Revision: 10 Oct 2025

Acceptance: 22 Oct 2025

Keywords

Cybersecurity, phishing, malicious URLs, deep learning, (ResMLP), URL classification, model interpretability, web security.

Abstract

The project addresses the growing threat of phishing and malicious websites, which cause financial loss, identity theft, and distrust in online services. It proposes a real-time URL classification system that integrates lexical, structural, behavioral, and reputation-based features. By leveraging traditional ML baselines (Random Forest, Naïve Bayes) along with a Residual Multi-Layer Perceptron (ResMLP) model, the system achieves high accuracy (~95%) and low inference latency (~50ms). An interactive dashboard enhances interpretability, ensuring trust in predictions and suitability for deployment in high-throughput environments.

Introduction

The digital era has amplified cyber threats, with phishing being one of the most persistent and damaging. Attackers exploit URL manipulations to deliver malware, steal sensitive data, and compromise users and enterprises alike. Traditional methods like blacklists and rule-based detection have proven inadequate as attackers constantly evolve. To tackle this, the project introduces an AI-driven phishing URL detection system. It employs both machine learning and deep learning approaches, focusing on high detection performance, low latency, and comprehensive feature extraction. A ResMLP model, integrated with lexical, structural, and behavioral signals, provides robustness against unseen threats while maintaining interpretability for security analyst.

Building on this, phishing URL detection has increasingly shifted toward intelligent, automated systems that adapt to evolving attacker strategies. Machine learning models leverage handcrafted features such as domain length, presence of special characters, and abnormal token patterns, while deep learning approaches extract higher-level representations without manual intervention. Recent works highlight that combining lexical and structural analysis with real-time behavioral insights enhances robustness, especially against zero-day phishing attempts. The proposed ResMLP model aligns with this trend by integrating multi-source features, ensuring improved accuracy and resilience compared to traditional classifiers or shallow neural networks.

Moreover, interpretability remains a crucial requirement in cybersecurity applications. Security analysts not only need accurate detection but also clear explanations to differentiate between benign anomalies and actual threats. The ResMLP

model addresses this by maintaining transparency through feature importance visualization and explainable AI (XAI) techniques, helping analysts trace malicious indicators within URLs. Such a framework ensures the system's practical usability in enterprise environments, where rapid incident response and trust in automated decisions are critical. This blend of performance, adaptability, and interpretability positions AI-driven phishing detection as a sustainable solution to counter rapidly evolving cyber threats.

Literature Survey

1. Haq et al. (2024) proposed another CNN-based approach that concentrated on large, evolving phishing datasets. Their study revealed that traditional machine learning models underperform when faced with constantly changing URL patterns. By evaluating the CNN model on multiple datasets such as PhishTank, UNB, and Alexa, they achieved an accuracy of 99.7%. The research stressed the need for adaptability and highlighted latency as a major bottleneck for real-time applications. This work demonstrated the potential of CNNs for handling large-scale URL datasets effectively.
2. Moving towards sequential learning, Baskota (2025) explored the use of Bi-LSTM for phishing URL detection. The study aimed to capture contextual patterns and classify URLs into benign or phishing categories with higher precision. The Bi-LSTM model achieved an accuracy of 97% while offering better handling of multi-class classification scenarios. The paper suggested that incorporating additional features such as SSL details, domain age, and user behavior could further strengthen detection. This work showcased how recurrent architectures can model sequential URL characteristics more effectively than static methods.
3. Graph-based methods also attracted significant research attention, as demonstrated by Guo et al. (2025). Their study pointed out that phishing attackers can easily bypass string-based feature models, necessitating contextual and relational data for robust detection. They proposed a graph-based machine learning approach combined with Loopy Belief Propagation, constructing graphs over URLs, IPs, and name servers. With this method, they achieved 98.77% accuracy, showing promise for large-scale phishing defense systems. However, the authors acknowledged challenges in scaling to dynamic, real-time environments, indicating directions for future research.
4. In addition, Sugantham et al. (2024) proposed an improved machine learning-based approach that relied heavily on DNS data and feature engineering. By evaluating multiple machine learning models, their research showed that Random Forest outperformed others, achieving 96.38% accuracy. Their method underlined the importance of feature diversity and integration with deep learning for practical deployment. The paper also emphasized the role of user-friendly interfaces and production-ready systems, marking a step toward bridging the gap between theoretical models and real-world applicability.
5. Recent research in phishing detection has emphasized deep learning approaches to improve accuracy and explainability. Islam et al. (2024) introduced PhishGuard, a CNN-based model that focused on reducing false positives while enhancing interpretability. The model incorporated a one-dimensional convolutional neural network with extensive feature engineering, achieving an impressive accuracy of 99.85%. The study highlighted the importance of model transparency, as black-box deep learning models often struggle to gain trust from cybersecurity analysts. Their work paved the way for integrating explainable AI in phishing URL detection systems.

Research Gap

- Existing blacklists and static methods are outdated against fast-changing phishing techniques.
- Many deep learning models achieve high accuracy but lack transparency (black-box issue).
- Graph-based and CNN/LSTM models are promising but face deployment challenges in real-time, high-throughput environments.
- Very few approaches provide both accuracy and interpretability simultaneously.
- Need for an intelligent system capable of real-time inference (<50ms latency) while ensuring explainability to build analyst trust.

Problem Statement

Phishing and malware-driven websites are becoming increasingly sophisticated, exploiting deceptive URL patterns, page behaviors, and stolen domain reputations. Current defenses, such as blacklists and signature-based systems, fail against zero-day threats and rapidly evolving URL strategies. There is a pressing need for an intelligent, adaptive system that can detect malicious links in real-time with high accuracy.

Conclusion

The project demonstrates a robust defense system, **CyberFence**, that detects phishing URLs in real-time with high accuracy using ResMLP. It balances detection efficiency with interpretability, making it practical for deployment in enterprises and large-scale digital ecosystems. By addressing evolving phishing strategies and zero-day threats, the system enhances cybersecurity resilience, ultimately protecting users from financial, reputational, and data losses.

References

Kale, Archana Pritam, and Shefali Sonavane. "PF-FELM: A robust PCA feature selection for fuzzy extreme learning machine." *IEEE Journal of Selected Topics in Signal Processing* 12.6 (2018): 1303-1312.

Islam, M.R., et al., PhishGuard: A CNN-Based Model for Detecting Phishing URLs with Explainability, 2024.

Haq, Q.E., et al., Detecting Phishing URLs via 1D CNN, 2024.

Baskota, S., Phishing URL Detection using Bi-LSTM, 2025.

Guo, W., et al., Efficient Phishing URL Detection Using Graph-based ML and LBP, 2025.

Sugantham, V., Mishra, A., Agarwal, R., An Improved Method of Phishing URL Detection Using ML, 2024.

Reddy, M.S., et al., Phishing Website Detection Using ML Algorithms, 2024.

Padmini, Y., and Usha Sree, P., Phishing Website Detection Using ML (Gradient Boosting), 2024.

Raj Paul, M., Deep Learning Solutions for Phishing by URL Detection, 2024.

Smadi, A., et al., Web-based Phishing URL Detection DL Model (Springer), 2025.

Li, W., et al., PhishDebate: LLM-Based Multi-Agent Framework for Phishing Detection, 2025.