International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.017 ISSN: 2581-4621

AI-POWERED PHISHING DETECTION : AN INTELLIGENT WEB-BASED THREAT ANALYSIS SYSTEM

Ms.D.Anusha[1], J. Sathish Reddy[2], P. Joshna[3], Sriram Mayuri[4]
[1]Assistant Professor, Department of CSE(Data Science), VidyaJyothi Institute of Technology, Hyderabad
[2,3,4,5]UG Student,Dept of CSE(Data Science) , VidyaJyothi Institute of Technology, Hyderabad

ABSTRACT
With the exponential growth of phishing attacks targeting individuals and organizations, ensuring online security has become increasingly critical. Phishing Detector is an AI-driven PHP web application designed to identify phishing URLs and detect potentially malicious web content using a combination of heuristic analysis, pattern recognition, and external threat intelligence.
The system retrieves HTML content from a user-provided URL and applies multiple detection techniques, including domain extension verification, identification of obfuscated JavaScript, analysis of suspicious form actions, and detection of brand impersonation patterns. Each identified indicator contributes to a cumulative risk scoring mechanism that classifies URLs into Low, Medium, or High Risk categories.
To enhance detection reliability, the system integrates with VirusTotal to validate URLs against global threat intelligence databases. The web interface is developed using HTML, TailwindCSS, PHP, and JavaScript, providing users with a responsive, interactive, and real-time phishing analysis platform.
This project demonstrates the practical integration of artificial intelligence, web development, and cybersecurity principles to deliver a fast, accurate, and user-friendly solution for phishing attack prevention.
Keywords: *Phishing Detection, AI, URL Analysis, Heuristic Analysis, VirusTotal, Cybersecurity, PHP, Risk Scoring, Threat Intelligence*

1. Introduction

The rapid proliferation of internet services and digital platforms has led to an alarming rise in cyber threats, with phishing being one of the most prevalent and damaging forms of online attack. Phishing attacks deceive users into revealing sensitive information such as passwords, credit card numbers, and personal data by impersonating legitimate websites and services.
Traditional approaches to phishing detection often rely on blacklisting known malicious URLs. While effective for known threats, such approaches fail to detect novel phishing domains that have not yet been catalogued. This limitation has prompted the development of more dynamic and intelligent detection mechanisms that can analyze the structural and behavioral characteristics of URLs and web content in real time.
The project titled "AI-Powered Phishing Detection" addresses this challenge by proposing a web-based application that employs heuristic analysis, pattern recognition, and external threat intelligence to evaluate the risk level of any given URL. Unlike static blacklist-based systems, the proposed solution dynamically analyzes multiple indicators of phishing behavior, enabling more comprehensive and adaptive detection.
A key motivation for this project is the increasing accessibility of cybersecurity tools. The use of PHP for backend logic and TailwindCSS for a responsive frontend interface ensures the application is lightweight, deployable, and accessible to a wide range of users. By integrating the VirusTotal API, the system gains access to a globally maintained repository of threat intelligence, further strengthening its detection capabilities.
The system is specifically designed to be user-friendly, requiring no technical expertise to operate. Users simply input a URL, and the system returns a detailed risk analysis complete with a risk score

DOI:10.46647/ijetms.2026.v10i03.017 ISSN: 2581-4621

and categorization. This makes the tool suitable for use by organizations, individuals, and security professionals alike.

In conclusion, the "AI-Powered Phishing Detection" project offers a timely and practical solution to one of the most persistent threats in modern cyberspace. By providing real-time, intelligent phishing analysis through an accessible web interface, it aims to empower users to navigate the internet more safely and confidently.

2. Literature Survey

2.1. Introduction to Phishing Threats

Phishing is a form of social engineering attack that attempts to steal sensitive user information by masquerading as a trustworthy entity. Over the decades, these attacks have evolved significantly in complexity and target specificity. Modern phishing campaigns leverage brand impersonation, lookalike domains, and obfuscated web content to bypass conventional security measures.

2.2. Blacklist-Based Detection Methods

Early phishing detection systems relied primarily on maintaining databases of known malicious URLs. Tools such as Google Safe Browsing and PhishTank maintain updated blacklists queried in real time. While effective for previously identified threats, these systems inherently fail to detect newly registered phishing domains, leaving users vulnerable during the window between domain creation and blacklist inclusion.

2.3. Heuristic and Feature-Based Approaches

To address the limitations of blacklisting, researchers developed heuristic-based methods that analyze the structural features of URLs. Indicators such as unusual domain extensions, excessive use of subdomains, presence of IP addresses in URLs, and abnormal URL length have been found to correlate strongly with phishing activity.

2.4. Machine Learning in Phishing Detection

Machine learning has been widely applied to phishing detection. Algorithms such as Random Forest, Support Vector Machines (SVM), and Naive Bayes have been trained on labeled datasets of phishing and legitimate URLs. These models extract lexical and host-based features and classify URLs with high accuracy, though they require curated training datasets and periodic retraining.

2.5. Deep Learning Approaches

Recent advances in deep learning have introduced neural network-based solutions for phishing detection. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have demonstrated improved classification accuracy by learning character-level and sequence-level patterns in URLs, capable of identifying subtle obfuscation techniques.

2.6. Threat Intelligence Integration

External threat intelligence platforms such as VirusTotal aggregate data from multiple antivirus engines and security vendors, providing a comprehensive view of a URL's risk profile. Integration of such APIs into detection systems enables real-time cross-validation against a global database of known threats, significantly enhancing detection reliability.

2.7. Content-Based Analysis Techniques

Beyond URL analysis, researchers have explored content-based techniques that examine the HTML and JavaScript structure of web pages. Detection of obfuscated scripts, hidden iframes, suspicious form actions directing to third-party servers, and unauthorized use of brand logos are among the indicators used to assess page-level phishing risk.

2.8. Challenges in Existing Systems

Despite significant progress, existing phishing detection systems face several persistent challenges including high false positive rates that block legitimate websites, latency issues in real-time analysis, evasion techniques that allow sophisticated phishing pages to bypass detection, and difficulty in keeping threat databases current.

3. Existing System

DOI:10.46647/ijetms.2026.v10i03.017 ISSN: 2581-4621

Current phishing detection solutions predominantly operate at the browser or network level. Browser extensions and built-in security features such as Google Safe Browsing alert users to known malicious pages. Email security gateways scan incoming links for phishing indicators before messages reach end users. These systems are primarily reactive, responding to known threats rather than proactively analyzing unfamiliar URLs.

Standalone URL scanning services such as VirusTotal and URLScan.io provide detailed analysis of submitted URLs, drawing on data from numerous security vendors. While powerful, these services are primarily designed for security professionals and may be inaccessible or complex for general users. They also do not always perform deep content analysis of retrieved web pages.

Some organizations employ machine learning models embedded within their security infrastructure to classify URLs in real time. While effective, these models require substantial computational resources and ongoing maintenance, including dataset curation and retraining cycles. They are typically not available as lightweight, standalone web applications.

PHP-based web applications for URL analysis exist but tend to offer limited detection depth, relying primarily on domain checks or simple keyword matching. They lack integration with comprehensive threat intelligence databases and do not perform multi-indicator scoring that captures the cumulative risk of multiple phishing signals.

Overall, existing systems either lack depth of analysis, require technical expertise to operate, or fail to combine multiple detection methodologies into a unified, user-friendly platform. These limitations highlight the need for an integrated, accessible, and intelligent phishing detection application.

4. Proposed System

The proposed "AI-Powered Phishing Detection" system is a web-based application built on PHP that enables users to analyze the phishing risk of any URL through a simple and intuitive interface. The system combines multiple layers of detection into a single, cohesive platform, overcoming the fragmentation observed in existing solutions.

Upon submission of a URL, the system first retrieves the HTML content of the target page using PHP's cURL library. This retrieved content forms the basis of a multi-indicator analysis pipeline. The system evaluates several distinct phishing signals including the domain extension, the presence of obfuscated or suspicious JavaScript, the action attributes of embedded HTML forms, and brand impersonation patterns.

Each detected indicator is assigned a weighted score. These scores are aggregated into a cumulative risk value mapped to one of three classification levels: Low Risk, Medium Risk, or High Risk. This scoring mechanism allows the system to provide nuanced assessments rather than binary safe/unsafe verdicts.

To further strengthen detection accuracy, the system integrates with the VirusTotal API. After performing local heuristic analysis, the URL is submitted to VirusTotal for cross-validation against its global database of threat intelligence. The combined output of local analysis and external validation produces a robust and comprehensive risk assessment.

The user interface is developed using HTML, TailwindCSS, and JavaScript, ensuring a responsive and visually clear presentation of results. Users receive a detailed breakdown of identified risk indicators alongside the overall risk score and classification, enabling transparency and understanding of the analysis.

The system is designed for accessibility, requiring no installation or technical background. It is deployable on any standard PHP-capable web server, making it practical for use by individuals, small organizations, and educational institutions.

6. Results

6.1. Successful URL Retrieval and Content Analysis

DOI:10.46647/ijetms.2026.v10i03.017 ISSN: 2581-4621

The system successfully retrieves HTML content from submitted URLs in real time using PHP cURL. Content from a diverse range of websites, including legitimate and simulated phishing pages, was correctly fetched and passed through the detection pipeline without errors.

6.2. Accurate Risk Indicator Detection

The multi-indicator analysis module consistently identified phishing signals such as suspicious domain extensions, obfuscated JavaScript, malicious form actions, and brand impersonation patterns. Detection accuracy remained high across test cases involving both clear-cut and ambiguous phishing URLs.

6.3. Effective Risk Scoring and Classification

The cumulative scoring mechanism correctly classified URLs into Low, Medium, and High Risk categories. Simulated phishing pages registered high risk scores, while legitimate URLs scored in the low risk range, demonstrating the effectiveness of the weighted scoring approach.

6.4. VirusTotal Integration

Integration with the VirusTotal API operated reliably, returning detailed threat intelligence reports for submitted URLs. Cross-validation with VirusTotal data improved overall detection accuracy, particularly for URLs that exhibited few local heuristic signals but were listed in global threat databases.

6.5. Real-Time Performance

The system delivered analysis results within an acceptable timeframe under typical network conditions. Local heuristic analysis completed in near real time; the primary latency factor was the external API call to VirusTotal.

6.6. Limitations Observed

- Dependence on VirusTotal API rate limits may introduce delays under heavy usage.
- Highly sophisticated phishing pages using dynamic content loading may partially evade static HTML analysis.
- URLs requiring login sessions cannot be fully analyzed due to access restrictions.

7. Conclusion

7.1. Summary of the Project

The "AI-Powered Phishing Detection" project successfully delivers an intelligent, web-based platform for identifying phishing URLs through a combination of heuristic analysis, risk scoring, and external threat intelligence validation. The system addresses key limitations of existing solutions by providing a unified, accessible, and multi-layered approach.

7.2. Achievement of Objectives

All primary objectives were achieved. The system retrieves and analyzes web content, identifies multiple phishing indicators, assigns cumulative risk scores, integrates with VirusTotal, and presents results through a responsive user interface capable of classifying URLs in real time with meaningful accuracy.

7.3. Practical Applications

The system is applicable across multiple domains including individual user security, organizational cybersecurity training, educational institutions teaching internet safety, and security professionals requiring a quick-analysis tool for suspicious URLs.

7.4. Scope for Enhancement

Future improvements may include the incorporation of deep learning models for more advanced URL and content classification, support for dynamic content analysis using headless browsers, expansion of the brand impersonation detection database, and implementation of user feedback loops to continuously improve detection accuracy.

7.5. Overall Conclusion

This project demonstrates that a practical, cost-effective, and user-friendly phishing detection system can be developed by combining AI principles with web technologies and external threat intelligence. It represents a meaningful contribution to applied cybersecurity and a foundation for more advanced phishing prevention tools.

DOI:10.46647/ijetms.2026.v10i03.017 ISSN: 2581-4621

References

1. A. K. Jain and B. B. Gupta, "Phishing Detection: Analysis of Visual Similarity Based Approaches," Security and Communication Networks, 2017.
2. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Phishing Websites Features," University of Huddersfield, 2015.
3. S. Garera et al., "A Framework for Detection and Measurement of Phishing Attacks," ACM Workshop on Recurring Malcode, 2007.
4. I. Fette, N. Sadeh, and A. Tomasic, "Learning to Detect Phishing Emails," WWW Conference, 2007.
5. M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, 2013.
6. VirusTotal API Documentation. Available: https://developers.virustotal.com/
7. OWASP Phishing Prevention Cheat Sheet. Available: https://cheatsheetseries.owasp.org/
8. Google Safe Browsing API Documentation. Available: https://developers.google.com/safe-browsing
9. PhishTank Developer Information. Available: https://phishtank.org/developer_info.php
10. TailwindCSS Documentation. Available: https://tailwindcss.com/docs
11. PHP Manual - cURL Functions. Available: https://www.php.net/manual/en/book.curl.php
12. A. Oest et al., "PhishFarm: A Scalable Framework for Measuring the Effectiveness of Evasion Techniques Against Browser Phishing Blacklists," IEEE S&P, 2019.
13. N. Chou et al., "Client-side Defense Against Web-based Identity Theft," NDSS, 2004.
14. L. Wenyin et al., "Detection of Phishing Webpages Based on Visual Similarity," WWW Conference, 2005.
15. Y. Zhang, J. Hong, and L. Cranor, "Cantina: A Content-Based Approach to Detecting Phishing Web Sites," WWW, 2007.