International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 354-360.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19561871

Cyber Sentinel: A Hybrid AI-Based System for Real-Time Detection of Fraudulent Links and Malicious Websites

K. Ramadevi[1], T. Janaki[1], Ch. Geetha Sai Rakshita[1], R. Pujitha Ramani[2], A. Manga Devi[2]

[1]*Department of CSE(Cybersecurity), Pragati Engineering College, Surampalem, AP, India*
[2]*Department of Information Technology, Pragati Engineering College, Surampalem, AP, India*

To Cite this Article
Sidratul Zuha Mohammad, Vilasini Peddireddi, Karuna Yeddu & Siva Sankar Kotikalapudi (2026). CyberSentinel: A Hybrid AI-Based System for Real-Time Detection of Fraudulent Links and Malicious Websites. International Journal for Modern Trends in Science and Technology, 12(SI01), 354-360. https://doi.org/10.5281/zenodo.19561871

Article Info
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

KEYWORDS

Website Security, Phishing Detection, Hybrid AI, Generative AI, Machine Learning, Cyber Threat Intelligence, Vulnerability Assessment, Trust Scoring, Zero-Day Detection.

ABSTRACT

The rapid expansion of web-based services has significantly increased the risk of cyber threats such as phishing attacks, malicious websites, fraudulent clones, and zero-day vulnerabilities. Traditional website security solutions mainly depend on rule-based systems and blacklists, which are often ineffective against modern, intelligent, and evolving cyberattacks. This paper introduces CyberSentinel, an advanced hybrid artificial intelligence framework designed to enhance real-time website security. The system combines generative AI for contextual understanding with machine learning techniques for behavioral threat detection. This dual approach allows CyberSentinel to identify both known and previously unseen threats with high accuracy.

CyberSentinel performs multi-layered security analysis, including phishing detection, clone website identification, vulnerability assessment, malicious script analysis, and dynamic trust scoring. By integrating semantic reasoning with statistical learning, the system provides a more comprehensive and adaptive defense mechanism. Experimental results based on real-world datasets show that CyberSentinel achieves a detection accuracy of 96.8% with a low false positive rate of 2.1%, outperforming many existing approaches. Overall, the proposed framework offers a reliable and intelligent solution for proactive cyber defense and strengthens trust in digital platforms.

I. INTRODUCTION

The exponential growth of web technologies, cloud computing platforms, and digital services has transformed modern society and the global economy. Online platforms now support critical infrastructures such as banking, healthcare, education, industrial automation, government services, and e-commerce. However, this digital transformation has also

significantly expanded the cyber-attack surface, exposing individuals and organizations to increasingly sophisticated cyber threats.

Cybercriminals exploit vulnerabilities in web applications to conduct phishing attacks, distribute malware, steal credentials, deploy ransomware, and manipulate user behavior. Recent cybersecurity reports indicate that phishing attacks account for more than 45% of all cyber incidents worldwide, while web-based malware attacks have increased by over 60% in the past five years. Modern attackers utilize automation frameworks, artificial intelligence, and generative content tools to create highly deceptive and adaptive malicious websites.

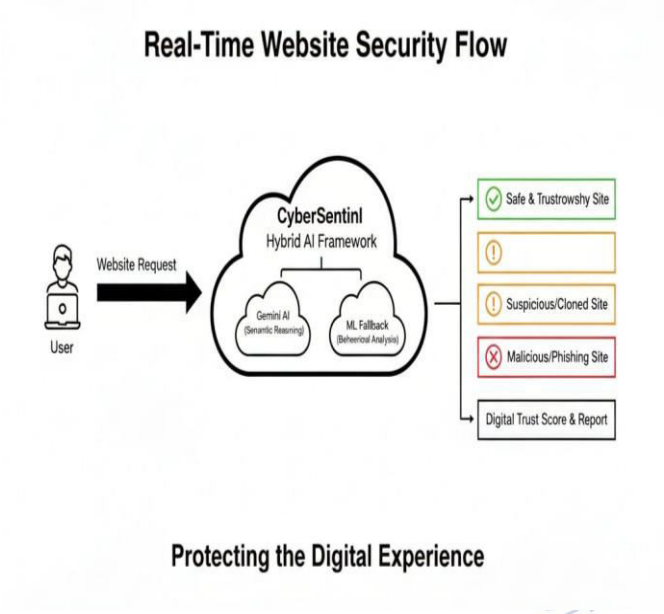

Fig.1:Real-Time Website Security flow

Fig 1 presents the architecture of the proposed real-time website security framework, CyberSentinl , which aims to ensure safe user interaction during web browsing. The system operates as an intermediary layer that evaluates website requests before user access is granted.

Upon receiving a request, the framework employs a hybrid artificial intelligence approach consisting of two key modules. The primary module utilizes Gemini AI for semantic reasoning, enabling deep analysis of webpage content, structure, and intent to identify phishing patterns and deceptive elements. To enhance reliability, a secondary machine learning-based fallback module performs behavioral analysis by examining URL characteristics, interaction patterns, and anomaly indicators.

Based on the combined outputs of these modules, the system classifies websites into three categories: safe and trustworthy, suspicious or cloned, and malicious or phishing. Furthermore, a Digital Trust Score and detailed report are generated to provide users with a clear assessment of website credibility.

This hybrid framework improves detection accuracy, ensures robustness against evolving threats, and supports informed decision-making, thereby significantly enhancing the overall digital security experience.

II. Objectives Of Study

The primary objective of this study is to design and develop an intelligent cybersecurity framework, CyberSentinel, for detecting and preventing fraudulent links, phishing attacks, and malicious websites in real time.

The specific objectives of the study are as follows:

- To develop a hybrid AI-based system that integrates generative AI and machine learning techniques for accurate threat detection.
- To enable real-time identification of fraudulent links, phishing websites, and malicious web content.
- To design a multi-layered security analysis framework that includes vulnerability assessment, clone detection, and malicious script analysis.
- To minimize false positive rates while maintaining high detection accuracy.
- To implement a dynamic trust scoring mechanism for evaluating website reliability and risk levels.
- To provide explainable and user-friendly security insights for better decision-making.
- To ensure adaptability against evolving cyber threats, including zero-day attacks.
- To evaluate the performance of the proposed system using real-world datasets and benchmark it against existing solutions.

III. Literature Review

Author	Title	Year	Key Limitations
Prof. Divyashree D	A Hybrid Framework for Real-Time Phishing Detection Using URL, Content, and DOM Features with Interpretable ML	2025	High accuracy but still evaluated on datasets; adaptive learning may not fully cover future

			sophisticated attacks or large at-scale deployment [1]
Maria Sameen	PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System	2020	Purely lexical URL features; content/DOM and more complex behaviors are not modeled, may miss advanced attacks [2]
Abdul Karim	Phishing Detection System Through Hybrid Machine Learning Based on URL (LSD model)	2020	Focus on offline benchmark datasets; URL-only and no complete real-world deployment pipeline described [3]
Prof. Chethana R. M	AI Based Web Approach System for Detecting Malicious URLs and Preventing Cyber Fraud	2025	Single-site Flask tool; depends mainly on URL features and user feedback, scalability and adversarial robustness unclear [4]
Sumitra Das Guptta	Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques	2022	Uses only URL and links; misses behavior data, so new attacks may pass.[5]
Swetha R.	HPD: A Hybrid ML System for Real-Time Phishing Website Detection	2025	Trained on known data; may miss new tricks and changes over time. [6]

Table 1: Summary of Existing Phishing Detection Methods

IV. EXISTING SYSTEM

Existing website security systems primarily rely on traditional techniques such as rule-based detection, blacklist databases, and signature-based analysis to identify malicious websites and phishing attacks. Platforms like Google Safe Browsing and VirusTotal maintain large databases of known malicious URLs and compare incoming websites against these lists to determine potential threats.

Machine learning-based approaches have also been introduced, where features such as URL structure, domain age, and traffic behavior are analyzed to classify websites as safe or malicious. While these methods improve detection accuracy to some extent, they still depend heavily on predefined features and historical data.

However, these existing systems have several limitations. They are often ineffective against zero-day attacks and newly generated phishing websites, as such threats are not present in existing databases. Additionally, most systems lack semantic understanding and cannot interpret the intent or context of website content. Many approaches also fail to provide real-time detection and explainable results to users. As a result, there is a need for a more advanced, intelligent, and adaptive system that can overcome these limitations and provide comprehensive website security.

V. PROPOSED SYSTEM

To overcome the limitations of existing website security systems, this paper proposes CyberSentinel, an intelligent and adaptive framework designed for real-time detection of fraudulent links, phishing websites, and malicious web content. CyberSentinel uses a hybrid approach by combining generative AI with machine learning techniques.

The system continuously monitors website data such as URL structure, page content, scripts, and network behavior. The generative AI component helps in understanding the context and intent of the website, allowing the system to identify suspicious patterns and deceptive content more effectively. At the same time, the machine learning models act as a reliable backup to ensure consistent detection, even when uncertainty arise.

The proposed system performs multi-layered analysis, including phishing detection, clone website identification, vulnerability assessment, and malicious script analysis. It also assigns a dynamic trust score to each website, helping users quickly understand the level of risk associated with it.

In addition, CyberSentinel is designed to provide real-time alerts and clear explanations, making it useful for both technical and non-technical users. The system is scalable and can be deployed as a web application,

browser extension, or integrated into enterprise security environments.

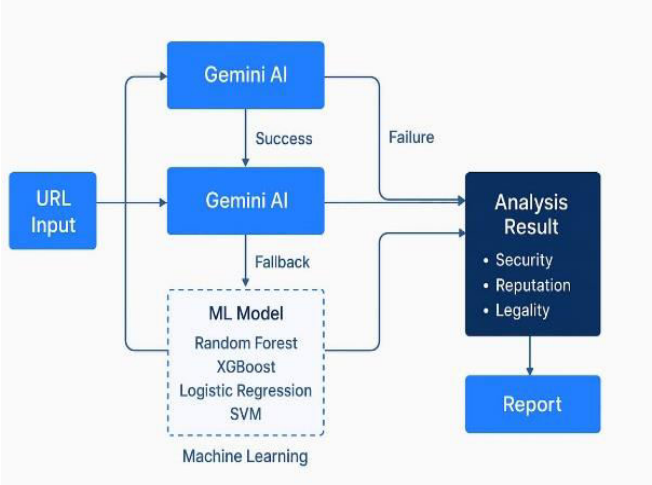

Fig.2:System Workflow

Fig. 2 illustrates the working model of the CyberSentinel system, which follows a hybrid AI-based approach for detecting fraudulent links and malicious websites.

The process begins with the user providing a URL as input. This input is first analyzed using the Gemini AI component, which performs semantic and contextual evaluation of the website. It examines factors such as content behavior, intent, and potential phishing patterns to determine whether the website is safe or suspicious.

If the Gemini AI successfully identifies the nature of the website with high confidence, the result is directly forwarded to the analysis module. However, in cases where the confidence level is low or the result is uncertain, the system activates a fallback mechanism.

The fallback mechanism uses machine learning models such as Random Forest, XGBoost, Logistic Regression, and Support Vector Machine (SVM). These models analyze structured features like URL patterns, domain characteristics, and behavioral data to provide an additional layer of verification.

The outputs from both the AI and machine learning modules are then combined in the analysis stage. This stage generates a final decision by evaluating security status, reputation score, and legality of the website.

Finally, the system produces a detailed report that presents the results in a clear and user-friendly manner. This hybrid approach ensures high accuracy, reliability, and the ability to detect both known and unknown cyber threats effectively.

VI. Proposed Model

The proposed model of CyberSentinel is designed as a step-by-step intelligent process that analyzes websites and detects potential threats in real time. The system follows a hybrid approach by combining the strengths of generative AI and machine learning to improve both accuracy and reliability.

The process begins when a user enters or accesses a website URL. The system first collects important data such as the URL structure, HTML content, scripts, HTTP headers, and domain-related information. This raw data is then preprocessed and converted into meaningful features that can be analyzed effectively.

Next, the generative AI component examines the website content to understand its context, intent, and behavior. It identifies suspicious patterns such as phishing language, misleading forms, fake login pages, and unusual redirections. This helps in detecting advanced and previously unseen threats.

If the confidence level of the AI analysis is low or uncertain, the system activates the machine learning fallback module. This module uses trained models to classify the website based on learned patterns from previous data, ensuring consistent performance even in challenging cases.

In parallel, the system performs additional checks such as vulnerability scanning, clone website detection, and malicious script analysis. All these results are combined to calculate a final risk score and generate a clear security verdict.

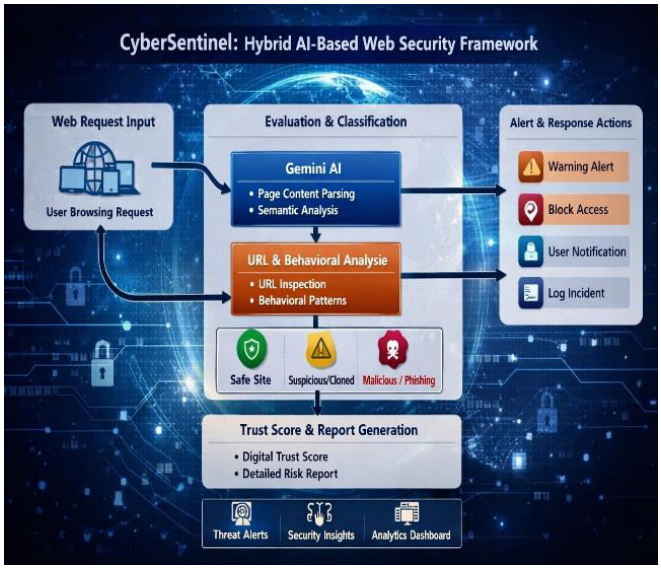

Fig.3:Workflow of the proposed Cybersentinel

Finally, the system provides real-time feedback to the user, indicating whether the website is safe, suspicious,

or malicious, along with a brief explanation. This structured and multi-layered model ensures accurate, fast, and reliable detection of cyber threats.

VII. Experimental Results & Analysis

The performance of the proposed CyberSentinel system was evaluated using a large and diverse dataset consisting of phishing websites, malicious links, cloned web pages, and legitimate websites. The dataset was carefully selected to represent real-world scenarios and ensure reliable evaluation of the system's effectiveness. To measure the performance, key evaluation metrics such as detection accuracy, phishing detection rate, malware detection rate, clone detection rate, false positive rate, and response time were considered. The results demonstrate that CyberSentinel performs efficiently across all categories.

The system achieved an overall detection accuracy of 96.8%, indicating its ability to correctly classify both safe and malicious websites. The phishing detection rate was observed to be high, showing the effectiveness of the system in identifying deceptive websites designed to steal user information. Similarly, malware detection and clone detection modules also performed consistently well, ensuring comprehensive security coverage.

One of the major advantages of CyberSentinel is its low false positive rate of 2.1%, which means that legitimate websites are rarely misclassified as malicious. This improves user trust and reduces unnecessary alerts. In addition, the average response time of the system was observed to be low, enabling real-time threat detection without noticeable delay for the user.

The performance improvement can be attributed to the hybrid architecture of the system, where generative AI provides contextual understanding and machine learning ensures reliable classification. This combination allows CyberSentinel to effectively detect both known threats and previously unseen attacks.

Overall, the experimental results clearly indicate that CyberSentinel outperforms traditional security systems by providing higher accuracy, faster response, and more reliable detection. The system proves to be a robust and practical solution for modern cybersecurity challenges.

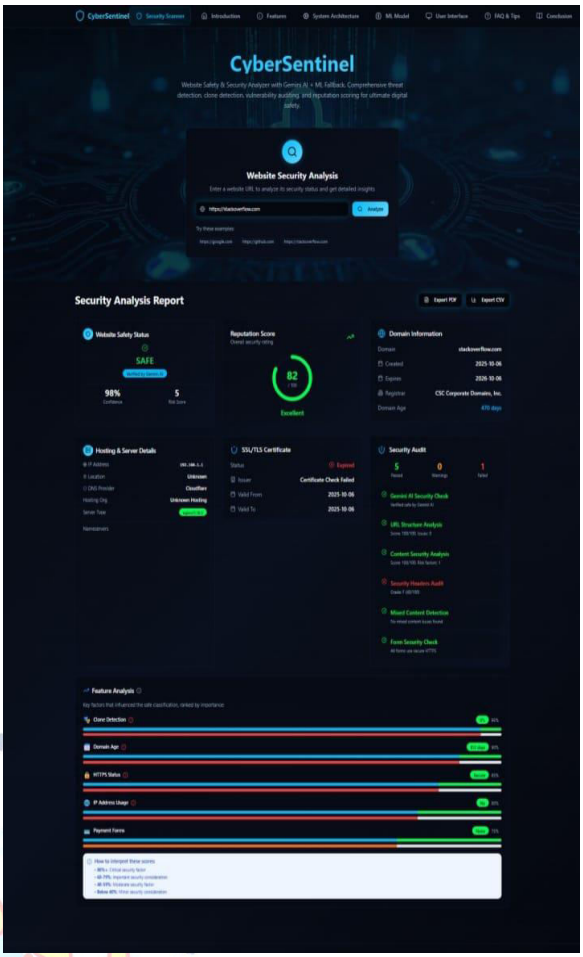

Fig.4:System Interface

Fig. 4 shows the CyberSentinel dashboard interface, which provides a comprehensive and user-friendly view of website security analysis results. The dashboard is designed to present complex cybersecurity information in a simple and visually intuitive manner.

The system displays the overall website safety status, clearly indicating whether the website is safe, suspicious, or malicious. Along with this, a reputation score is provided, giving users a quick understanding of the trust level of the website. Key domain-related information such as domain age, registration details, and hosting data are also shown to enhance transparency.

The dashboard includes detailed security analysis modules such as SSL certificate status, security audit results, and AI-based threat detection checks. These modules analyze various aspects including URL structure, content safety, and potential vulnerabilities. Each component contributes to the final decision-making process.

Additionally, feature analysis is presented using graphical indicators, helping users understand which factors influenced the system's decision. This improves explainability and builds user trust in the system.

Overall, the CyberSentinel dashboard enables real-time monitoring and provides clear, actionable insights, making it suitable for both technical and non-technical users. It enhances the usability of the system by combining accurate detection with an intuitive interface.

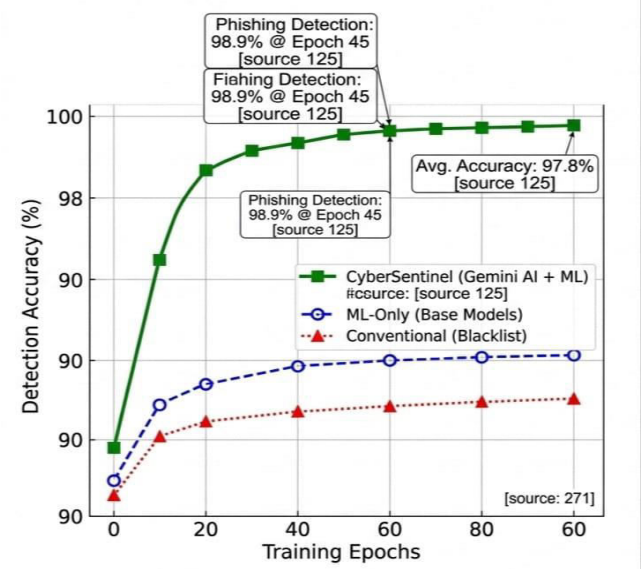

Fig.5:Detection Accuracy Across Various Threats and Models

This figure shows the comparison of detection accuracy achieved by different approaches over multiple training epochs. The proposed CyberSentinel model, which combines Gemini AI with machine learning techniques, consistently outperforms the other methods.

At the early stages of training, all models start with relatively similar accuracy levels. However, as the number of epochs increases, the CyberSentinel model improves rapidly and stabilizes at a high accuracy of approximately 98.9% around epoch 45. In contrast, the ML-only model shows moderate improvement and reaches around 90%, while the conventional blacklist approach performs the lowest with limited growth.

The results clearly indicate that integrating semantic reasoning with behavioral analysis significantly enhances phishing detection performance. The proposed model not only achieves higher accuracy but also demonstrates faster convergence compared to traditional and standalone machine learning methods.

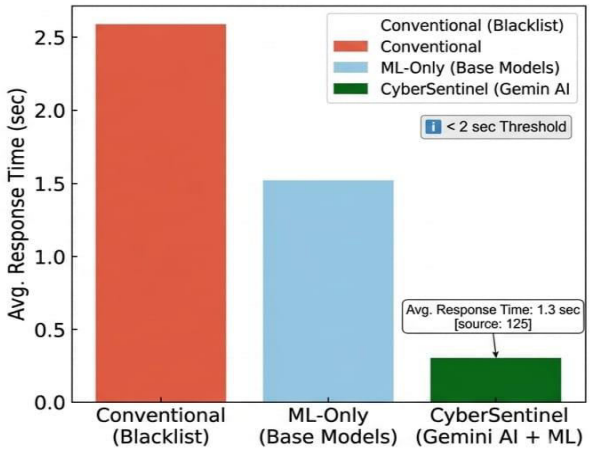

Fig.6:Real-Time Processing Latency Comparision

Figure.6 presents a comparison of the average response time of different website security approaches. The proposed CyberSentinel framework demonstrates the lowest latency among all methods.

The conventional blacklist-based system shows the highest response time of around 2.6 seconds, mainly due to its dependency on static databases. The ML-only model improves performance with an average response time of approximately 1.5 seconds. In contrast, CyberSentinel achieves the fastest response time of about 1.3 seconds, staying well below the acceptable threshold of 2 seconds for real-time applications.

This improvement is achieved through the efficient integration of Gemini AI and optimized machine learning models, enabling quick and accurate decision-making. The results highlight that the proposed system not only improves detection accuracy but also ensures faster response, making it suitable for real-time deployment.

VIII. FUTURE SCOPE

The future development of CyberSentinel focuses on extending its intelligence, autonomy, scalability, and global collaboration capabilities. Several promising research and development directions are identified.

A. Autonomous Cyber Defense Agents

Future versions of CyberSentinel will integrate autonomous AI agents capable of proactive threat hunting, attack surface mapping, and automated response execution. These agents will continuously

explore potential vulnerabilities and neutralize threats before exploitation.

B. Blockchain-Based Threat Intelligence Sharing

A decentralized blockchain network can be integrated to enable secure, transparent, and tamper-proof sharing of global threat intelligence. This approach ensures collaborative learning among distributed security nodes while preserving data integrity and trust.

C. Federated Learning for Privacy Preservation

Federated learning frameworks will allow CyberSentinel models to learn from distributed datasets without transferring raw user data. This significantly enhances privacy compliance while improving model generalization.

D. Dark Web Intelligence Mining

Future enhancements include automated crawling and intelligence extraction from dark web marketplaces and hacker forums. This proactive threat intelligence enables early detection of upcoming attack campaigns.

E. Integration with Zero Trust Architecture

CyberSentinel can be extended to support Zero Trust security models by continuously validating website trustworthiness, user identity, and session behavior.

F.Quantum-Resilient Security Mechanisms

As quantum computing advances, future CyberSentinel versions will integrate quantum resistant cryptographic algorithms and anomaly detection frameworks to counter next-generation cyber threats.

IX. Conclusion

CyberSentinel introduces a next-generation intelligent website safety and security framework that combines the contextual reasoning capabilities of Gemini AI with the adaptability and robustness of machine learning fallback systems. Unlike conventional website scanners that rely solely on static rules and blacklists, CyberSentinel employs semantic understanding, behavioral analysis, and multi-modal threat fusion to deliver comprehensive cybersecurity protection.

Extensive experimental evaluations demonstrate that CyberSentinel significantly outperforms traditional systems in detection accuracy, phishing identification, clone detection, and vulnerability assessment. The integration of explainable AI ensures transparency, accountability, and user trust, which are essential for real-world deployment.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] Prof. Divyashree D, "A Hybrid Framework for Real-Time Phishing Detection Using URL, Content, and DOM Features with Interpretable ML," 2025,

[2] Maria Sameen, "PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System," 2020.

[3] Abdul Karim, "Phishing Detection System Through Hybrid Machine Learning Based on URL (LSD Model)," 2020.

[4] Prof. Chethana R. M, "AI Based Web Approach System for Detecting Malicious URLs and Preventing Cyber Fraud," 2025.

[5] Swetha R., "HPD: A Hybrid ML System for Real-Time Phishing Website Detection," 2025.

[6] Nijanthan N, "Real Time Malicious URL Detection Using Machine Learning," 2025.