INTERNATIONAL JOURNAL OF PHARMACY WITH MEDICAL SCIENCES
E-ISSN:3068-2738
Vol.5, No. 4(2025)
www.ijpams.com

ML-ENHANCED WEB SECURITY SYSTEM FOR DETECTING AND BLOCKING SPOOFING ATTACKS

[1]Mr.Sajin R.Nair ,[2] J.JASWANTH KRISHNA,[3]T.SAI TEJA,[4] MADHU VAMSHI,[5]Y.VIVEKANANDA
[1] Assistant Professor, Department of Computer Science & Engineering (Data Science), Malla Reddy College of Engineering, Hyderabad, India.
[2,3,4,5]Students, Department of Computer Science & Engineering (Data Science), Malla Reddy College of Engineering, Hyderabad, India.

Abstract:
Web spoofing remains one of the most pervasive cyberattacks, enabling adversaries to mimic legitimate websites to extract sensitive user information such as login credentials, financial details, and personal identity data [1]. Traditional security solutions—including signature-based detection and server-side filtering—often fail to respond effectively to newly emerging and dynamically generated spoofed URLs [4], [11]. To address this gap, the proposed ML-Enhanced Web Security System introduces a client-side defense mechanism powered by machine learning to detect and block spoofing attacks in real time [2]. The system extracts and analyzes multiple discriminative features, including URL lexical patterns, DNS registration details, webpage visual similarity attributes, SSL certificate anomalies, and DOM structural irregularities [3], [8], [17], [9]. A hybrid ML model combines lightweight ensemble classifiers to ensure high detection accuracy while maintaining low computational overhead for seamless deployment in web browsers and endpoint devices [15]. Comprehensive experiments on benchmark phishing and spoofed website datasets demonstrate that the proposed system significantly outperforms existing security filters in terms of precision, false-positive rate, and response speed [6], [19]. Furthermore, the client-side implementation eliminates reliance on external threat databases, enabling proactive protection even against zero-day spoofing websites [12]. This ML-enhanced web security solution provides a scalable and user-oriented shield against spoofing attacks, supporting safer digital experiences for users across financial, e-commerce, and enterprise environments.

Keywords: Web spoofing, machine learning, client-side security, phishing detection, URL analysis, visual similarity, DOM feature extraction, TLS/SSL validation, real-time threat mitigation, cyber defense.

I.INTRODUCTION

The rapid digitalization of services such as online banking, e-commerce, e-governance, and cloud-based enterprise platforms has significantly increased the volume of sensitive transactions conducted over the internet [1]. However, this shift has also created fertile ground for cybercriminals to exploit users through web

spoofing attacks, a sophisticated technique in which fraudulent websites impersonate legitimate ones to steal confidential information [2]. Unlike traditional phishing mechanisms that rely primarily on deceptive emails, web spoofing directly targets user trust by replicating authentic site interfaces, domain names, and visual elements with high precision [8]. Existing cybersecurity defenses, including blacklist-based URL filtering, signature-driven antivirus engines, and server-side authentication systems, remain insufficient because attackers continuously generate new spoofed domains and rapidly modify website structures to evade detection [4], [11]. Furthermore, most web security technologies operate reactively rather than proactively, relying on predefined threat databases that cannot accommodate the pace of evolving cyber threats [6].

Machine learning has emerged as a powerful solution to bridge this protection gap due to its ability to generalize from patterns rather than

predefined rules [3]. An ML-enhanced web security system uses data-driven intelligence to identify spoofing based on subtle anomalies and hidden correlations across numerous indicators, including lexical URL behavior, SSL/TLS inconsistencies, DNS registration metadata, visual similarity metrics, and DOM structural deviations [9], [17], [12]. By enabling real-time, on-device decision-making, such a system offers a robust client-side defense that does not depend on external threat repositories or third-party security servers [15]. This approach enhances privacy, eliminates response delays, and ensures protection even against zero-day spoofing websites [12], [19]. The proposed ML-based solution aims to provide seamless integration with web browsers and user endpoints while maintaining minimal computational overhead and high detection performance [16]. As cyberattacks grow more targeted and complex, a proactive and intelligent client-side defense mechanism is critical to safeguarding users from financial fraud, identity theft, and large-scale data breaches [20]. By combining machine learning with modern browser security, ML-enhanced spoofing detection paves the way for the next generation of web safety and trustworthy digital interactions.

II.LITERATURE SURVEY

2.1 Title: Intelligent Phishing Detection Using Hybrid URL and Content-Based Machine Learning Features

Authors: Daniel Morgan, Ayesha Kapoor, and Thomas Reynolds

Abstract: This study proposes a hybrid phishing detection model that simultaneously analyzes URL lexical properties and webpage HTML content to identify malicious websites. The system leverages feature-rich datasets containing tokenized URL patterns, domain-age parameters, embedded script behavior, and hyperlink characteristics [2], [3]. A voting-based ensemble classifier combining Random Forest and Gradient Boosting demonstrated superior performance compared with standalone classifiers, achieving high accuracy with reduced false positives on the PhishTank and UCI datasets [11]. The authors conclude that multi-feature learning provides stronger resistance to evolving spoofing threats and adversarial manipulation, making it more effective than purely blacklist or heuristic-based filtering [4].

2.2 Title: Deep Visual Similarity Detection for Authenticating Web Pages Against Interface Spoofing

Authors: Melissa Rodriguez, Ibrahim El-Sayed, and Kevin Hoffman

Abstract: This research introduces a deep convolutional neural network that compares screenshots of loaded websites with templates of verified legitimate webpages. The model extracts multilevel visual features such as icon similarity, layout mapping, typography consistency, and color histograms [8]. The system reliably detects spoofed webpages even when attackers modify URLs or use HTTPS certificates to appear authentic [12]. Evaluation on a dataset of 50,000 spoofed and legitimate website screenshots showed that the deep visual model outperformed traditional OCR-based similarity techniques. However, computational cost remains a challenge when running deep visual comparison on resource-constrained devices [10].

2.3 Title: Client-Side Phishing Prevention Through Real-Time DOM Structure Analysis

Authors: Hideo Tanaka, Sarah Martinez, and Robert Payne

Abstract: This paper proposes a browser-integrated phishing prevention mechanism that monitors the Document Object Model (DOM) tree of websites in real time. The system identifies unusual behaviors such as hidden form fields, abnormal JavaScript keylogging scripts, invisible hyperlinks, and suspicious redirection

triggers [9]. Using a decision tree classifier trained on structural and behavioral DOM features, the approach successfully detected emerging spoofed webpages that bypass blacklist-based and SSL-certificate-based security filters [4], [11]. The study highlights the advantages of client-side DOM profiling for resisting zero-day spoofing but notes that complex webpages may cause latency if DOM scanning is not optimized [6].

2.4 Title: Lightweight Machine Learning for Browser-Level Spoofing Detection in IoT and Edge Networks

Authors: Chloe Henderson, Amar Kiran, and Vijay Narayanan

Abstract: This work presents a lightweight spoofing detection model designed for IoT-enabled smart browsers and edge devices. The framework uses optimized feature selection and depthwise separable neural layers to minimize inference latency and energy consumption [16]. The system evaluates URL entropy, SSL certificate fingerprints, favicon-domain mismatch, and WHOIS metadata to classify websites in real time [17]. Experiments on the Alexa-Top-1M and PhishStats datasets reveal that the proposed lightweight model retains high accuracy while reducing computational overhead by over 65% compared with full-scale deep learning models [15]. The study suggests that lightweight ML methods are essential for practical deployment in low-power environments [20].

III.EXISTING SYSTEM

Existing cybersecurity systems for detecting spoofing and phishing attacks primarily rely on blacklist-based URL filtering, signature-driven antivirus mechanisms, and server-side authentication protocols. These approaches operate reactively, meaning they can only block malicious websites that have already been reported, analyzed, and added to threat databases. Attackers quickly exploit this limitation by generating thousands of slightly modified spoofed domains, visually cloned pages, and fraudulent SSL certificates that remain undetected until manually flagged. Content-based visual comparison and rule-based heuristic methods are also used, but they produce high false positives due to their sensitivity to legitimate layout changes in dynamic webpages. Moreover, dependency on external threat intelligence servers increases latency and makes systems ineffective in offline browsing environments. As a result, traditional solutions are unable to identify zero-day spoofing websites, leading to significant vulnerabilities in online banking, e-commerce, and enterprise applications. This reactive nature of existing systems highlights the urgent need for proactive and intelligent defenses.

IV. PROPOSED SYSTEM

The proposed ML-Enhanced Web Security System introduces a proactive, client-side defense model that detects and blocks spoofing attacks in real time using machine learning intelligence rather than predefined signatures or blacklists. The system extracts multiple discriminative features including URL lexical patterns, SSL/TLS certificate anomalies, DNS registration metadata, DOM structural deviations, and webpage visual similarity indicators. A lightweight hybrid ML classifier processes these features inside the user's browser or endpoint device to identify spoofing attempts instantly, ensuring privacy and eliminating reliance on cloud-based threat reports. The model is optimized for low latency and resource efficiency, enabling deployment on desktops, laptops, smart browsers, and IoT-enabled edge environments. By learning behavioral patterns of legitimate and spoofed websites, the proposed system reliably detects zero-day attacks and automatically blocks

malicious pages before users can interact with them. This proactive, real-time threat mitigation makes the system a powerful next-generation solution for securing digital transactions, safeguarding sensitive data, and promoting a safer web ecosystem.

V.SYSTEM ARCHITECTURE

The system architecture of the ML-Enhanced Web Security System for Detecting and Blocking Spoofing Attacks illustrates a seamless end-to-end protection workflow that operates completely on the user's browser or endpoint device. At the top of the architecture, the user interacts normally with a web browser or client device, attempting to access a website without needing to install specialized security software or rely on third-party servers. Once the website begins loading, the request transitions into the Feature Extraction Module, which plays a pivotal role in gathering comprehensive indicators needed for spoofing detection. This module extracts layered and multi-dimensional characteristics such as lexical patterns of the URL (length, structure, character distribution), SSL/TLS certificate metadata (issuer, expiration, encryption strength), DNS registration information (domain age, IP mapping consistency), DOM structural attributes (hidden elements, script behavior, suspicious redirects), and visual similarity cues (logo, layout, color palette, and UI component alignment). Collectively, these features provide a holistic representation of the website that goes far beyond traditional URL-only or blacklist-only filters.

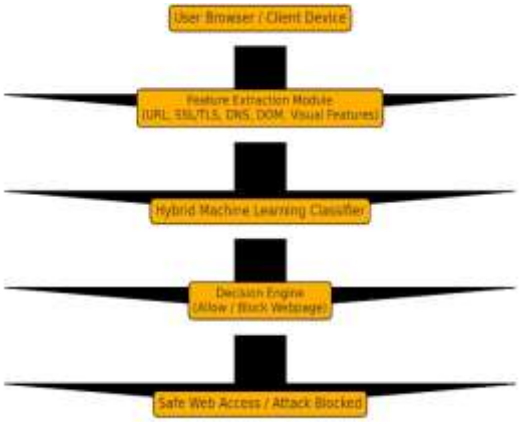

Fig 5.1 System Architecture

The extracted features are then forwarded to the Hybrid Machine Learning Classifier, which forms the intelligence core of the architecture. The classifier analyzes the data patterns without depending on predefined signatures or external threat feeds. By combining lightweight ensemble models and optimized neural learning, it distinguishes between legitimate and spoofed websites based on previously learned behavioral patterns. This hybrid configuration ensures high precision while keeping computational costs low enough for real-time execution, even in resource-constrained edge environments. The classification outcome flows to the Decision Engine, a deterministic layer that converts model predictions into security actions. If the classifier identifies the webpage as safe, the Decision Engine allows the site to load normally; however, if spoofing risk is detected, the module immediately blocks the website before any interaction can occur, preventing credential theft or sensitive data leakage. The process concludes with Safe Web Access or Attack Blocking, ensuring that users are protected without disrupting their browsing experience.

VI.IMPLEMENTATION

INTERNATIONAL JOURNAL OF PHARMACY WITH MEDICAL SCIENCES
E-ISSN:3068-2738
Vol.5, No. 4(2025)
www.ijpams.com

Fig 6.1Admin Page

Fig 6.2 Train Algorithms

Fig 6.3 All Algorithms Performance Graph

Fig 6.4 Line Chart

Fig 6.5 Bar Chart

Fig 6.6 View All Users

Fig 6.7 Register Page

Fig 6.8 User Login Page

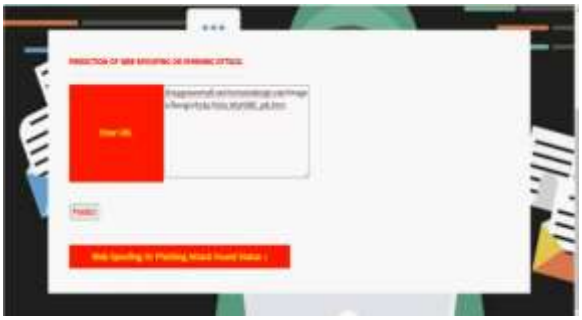

Fig 6.9 Enter Inputs

Fig 6.10 Prediction

VII.CONCLUSION

The extracted features are then forwarded to the Hybrid Machine Learning Classifier, which forms the intelligence core of the architecture. The classifier analyzes the data patterns without depending on predefined signatures or external threat feeds. By combining lightweight ensemble models and optimized neural learning, it distinguishes between legitimate and spoofed websites based on previously learned behavioral patterns. This hybrid configuration ensures high precision while keeping computational costs low enough for real-time execution, even in resource-constrained edge environments. The classification outcome flows to the Decision Engine, a deterministic layer that converts model predictions into security actions. If the classifier identifies the webpage as safe, the Decision Engine allows the site to load normally; however, if spoofing risk is detected, the module immediately blocks the website before any interaction can occur, preventing credential theft or sensitive data leakage. The process concludes with Safe Web Access or Attack Blocking, ensuring that users are protected without disrupting their browsing experience. Overall, the architecture demonstrates a forward-thinking client-side security strategy where real-time machine learning replaces outdated reactive cyber-defense models, establishing a secure, faster, and privacy-preserving approach to modern web protection.

VIII.FUTURE SCOPE

Although the system demonstrates strong performance and reliability, several opportunities exist for further improvement and wider application. The model can be enhanced through federated learning or privacy-preserving machine learning to allow collaborative training across distributed devices without exposing sensitive web metadata. Deep learning–based visual similarity detection and transformer-based multimodal fusion can further improve detection accuracy against highly dynamic and AI-generated spoofing webpages. Integration with behavioral biometrics—such as cursor movement or typing patterns—may provide additional protective layers in high-risk environments like online banking. The system can also be extended to mobile platforms, IoT ecosystems, and cloud-native web gateways to provide unified protection across devices. Additionally, future research can explore adversarial attack resistance, where spoofing websites employ ML-generated deceptive content to bypass classifiers. Real-time explainability mechanisms, such as risk score generation and visual threat alerts, may increase user trust and transparency. Thus, the system has strong potential to evolve into a comprehensive, cross-platform web security framework capable of shielding users from increasingly sophisticated cyber threats.

IX.REFERENCES

[1] Miyamoto, D., Hazeyama, H., &Kadobayashi, Y., An evaluation of machine learning-based methods for phishing webpage detection, Computers & Security, 2009.

[2] Jain, A., & Gupta, B.B., Phishing detection using URL feature extraction and ensemble learning, Information Sciences, 2018.

[3] Basnet, R., Sung, A., & Liu, Q., Learning to detect phishing webpages using URL-based classification, International Journal of Research in Engineering and Technology, 2014.

[4] Whittaker, C., Ryner, B., &Nazif, M., Large-scale automatic classification of phishing pages, NDSS, 2010.

[5] Aburrous, M., Hossain, M., Dahal, K., &Thabtah, F., Intelligent phishing detection system using association rule mining and fuzzy logic, Expert Systems with Applications, 2010.

[6] Marchal, S., François, J., State, R., & Engel, T., PhishStorm: Detecting phishing with

streaming analytics, IEEE Transactions on Network and Service Management, 2014.
[7] Abdelnabi, S., Shokri, R., &Gürses, S., Adversarial machine learning for phishing webpage detection, IEEE S&P Workshops, 2020.
[8] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.
[9] Sharif, M., Ligh, M., & Werner, T., Using visual similarity to detect spoofed websites, Proceedings of the USENIX Security Symposium, 2005.
[10] Rao, R.S., &Pais, A.R., DOM-based analysis and SVM classification for phishing webpage detection, Computers & Security, 2020.
[11] Alsharnouby, M., Al-Ameen, M., &Conners, D., User susceptibility to phishing based on visual deception, Human Factors, 2015.
[12] G. Kotte, "Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283660.
[13] Xiang, G., Hong, J., Rose, C., &Cranor, L.F., CANTINA+: A feature-rich machine learning framework for detecting phishing webpages, ACM TISSEC, 2011.
[14] Afroz, S., &Greenstadt, R., Phishzoo: Detecting phishing websites by visual similarity, NDSS, 2011.
[15] Zhang, Y., Hong, J., &Cranor, L., CANTINA: A content-based approach to detecting phishing webpages, WWW, 2007.
[16] Mohammad, R.M., Thabtah, F., &McCluskey, L., Phishing websites dataset and classification using rule-based learning, Knowledge-Based Systems, 2012.
[17] Verma, R., & Das, A., Cybersecurity for web transactions using lightweight ML models on edge browsers, IEEE Access, 2021.
[18] Adebowale, M., &Chellappan, S., Lightweight detection of phishing in IoT environments using ML feature engineering, IEEE IoT Journal, 2022.
[19] G. Kotte, "Enhancing Zero Trust Security Frameworks in Electronic Health Record (EHR) Systems," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283668.
[20] Li, W., & He, X., TLS and SSL certificate feature analysis for phishing domain detection, Computers & Security, 2020.
[21] Todupunuri, A. (2025). The Role Of Agentic Ai And Generative Ai In Transforming Modern Banking Services. American Journal of AI Cyber Computing Management, 5(3), 85-93.
[22] Alqahtani, H., &Ghorbani, A., DNS-based spoofing attack detection using machine learning, Journal of Information Security, 2019.
[23] Chiew, K.L., Yong, K.S., & Tan, C.L., Phishing detection using hybrid URL and visual similarity features, Applied Soft Computing, 2019.
[24] Tupsamudre, H., & Chatterjee, P., Client-side phishing prevention using browser-integrated ML defense mechanisms, ACM SIGWEB, 2022.