# Phishing Website Detection Using Machine Learning via Browser Extension and Mobile Application

Venkata Ratnam K[1], Ch Jaswanth Kumar[2], D Asha Jyothi[3], B Aravindasai[4], P Kushal Sai[5]
[1]*Assistant Professor, Dept. of CSE, Anil Neerukonda Institute of Technology and Sciences (A) Andhra Pradesh, India*
[2,3,4,5] *Dept. of CSE, Anil Neerukonda Institute of Technology and Sciences (A) Andhra Pradesh, India*

***Abstract*— This paper introduces a novel, multi-layered approach to phishing detection, deployed across web browsers and Android platforms, offering real-time protection against advanced phishing threats. The system utilizes a hybrid model that combines Random Forest (RF) classification as the first layer of defense with Large Language Models (LLMs) as a second, more sophisticated layer to enhance detection accuracy. The Random Forest classifier analyzes key URL features such as domain structure, URL length, presence of suspicious keywords, and HTTPS status to provide an initial prediction with high speed and efficiency. Once a potential threat is detected or flagged as ambiguous, the LLM performs a deeper semantic analysis of the webpage content and associated metadata, identifying subtle linguistic patterns, inconsistencies, and suspicious behaviors that may escape traditional detection models. This layered detection mechanism ensures high accuracy, capturing zero-hour phishing attempts and reducing false positives significantly. The system is implemented through browser extension sand Android accessibility services, allowing seamless cross-platform functionality to protect users across different environments. In real-world deployment scenarios, this hybrid model demonstrated exceptional performance, achieving an impressive 97.8% accuracy and maintaining a false-positive rate of just 0.3%. By combining the strengths of machine learning and natural language understanding, our system effectively mitigates the risks of phishing attacks, offering a comprehensive and adaptive security solution across web and mobile platforms.**

***Index Terms*— phishing detection, machine learning, browser extension, mobile application, cybersecurity, random forest, large language models, zero-hour detection, cross-platform protection, hybrid model.**

## I. INTRODUCTION

The 21st century has witnessed an unprecedented expansion of digital technologies, enabling seamless communication, online transactions, and information sharing across the globe. However, this digital transformation has also introduced new challenges, with phishing attacks emerging as one of the most pervasive and sophisticated threats to online security. Phishing involves tricking unsuspecting users into providing sensitive information—such as login credentials, financial details, and personal data—by masquerading as trustworthy entities through deceptive websites, emails, and messages.

As these attacks continue to evolve, they exploit vulnerabilities in web browsers and mobile applications, targeting users across multiple platforms and increasing the risk of data breaches and financial loss. Zero-hour phishing attacks, URL obfuscation, and dynamic website generation allow attackers to bypass traditional security measures, making it difficult for conventional models to detect new and evolving threats. Traditional phishing detection systems rely on URL-based analysis and heuristic approaches, which often struggle to detect zero-hour phishing attempts and suffer from high false-positive and false-negative rates.

To address these challenges, our research introduces a multi-layered phishing detection system that combines the strengths of Random Forest (RF) classification with the natural language understanding capabilities of Large Language Models (LLMs), significantly improving detection accuracy. The system leverages a two-layered approach where Random Forest acts as the first line of defense by analyzing URL-based features, such as domain characteristics, presence of special characters, and HTTPS status, to filter out most malicious URLs. Ambiguous or suspicious URLs that bypass the first layer undergo deeper semantic analysis by the LLM, which evaluates webpage content, metadata, and HTML structures to detect subtle phishing indicators that may escape traditional detection.

The system is implemented through browser extensions and Android accessibility services,

ensuring consistent protection across web and mobile platforms. Real-world deployment of this model demonstrated 97.8% accuracy with a false-positive rate of 0.3%, making it a reliable solution for mitigating phishing threats. While challenges such as false positives and zero-hour phishing detection exist, the integration of machine learning and language models enhances real-time phishing detection and provides comprehensive security across multiple platforms. As phishing threats continue to evolve, adopting adaptive, cross-platform, and AI-powered approaches will remain essential for safeguarding users against emerging cyber threats.

| year | No. of Phishing Sites Detected (Approx.) |
|---|---|
| 2021 | 1.4 million |
| 2022 | 2.3 million |
| 2023 | 3.4 million |
| 2024 | 5.1 million |

The above table shows: Year-wise Phishing Trends

## II.RELATED WORKS

Phishing detection has become a growing concern in cybersecurity, aiming to protect users from malicious websites that attempt to steal sensitive information such as login credentials, financial details, and personal data. This literature review consolidates key research contributions, emphasizing machine learning (ML) and natural language processing (NLP) strategies, their applications, and implementation challenges in phishing detection. Various studies highlight the effectiveness of traditional ML techniques such as Support Vector Machine (SVM), Random Forest (RF), and Decision Trees (DT) in classifying phishing websites based on URL features, domain characteristics, and content-based indicators.

A study on *Phishing Website Detection Using Machine Learning* demonstrates the comparative performance of ML models, where Random Forest achieved the highest accuracy due to its ability to handle large datasets and reduce overfitting. Another study, *Phishing URL Detection and Reporting System Using Machine Learning Approach*, introduces a comprehensive system that not only detects phishing URLs but also incorporates a reporting mechanism to monitor and log identified phishing threats in real time. This system integrates various feature extraction techniques, including lexical, host-based, and content-based features, ensuring a robust classification process. Recent advancements also explore the integration of Large Language Models (LLMs) to enhance phishing detection by analyzing webpage content, identifying subtle anomalies, and understanding deceptive language patterns. While these approaches have shown promise in improving detection accuracy and reducing false positives, challenges such as ensuring model robustness, addressing zero-hour phishing attacks, and minimizing computational overhead must be addressed to ensure the widespread adoption of these techniques. Overall, the evolution of phishing detection systems, empowered by advancements in ML and NLP, continues to strengthen defenses against increasingly sophisticated phishing attacks.

| Title | Study Method | Dataset | Analysis | Demerits |
|---|---|---|---|---|
| Phishing Website Detection Using Machine Learning (2022) - IEEE | Comparison of machine learning models (SVM, RF, DT) for URL classification | UCI Machine Learning Repository | Evaluates performance of models and highlights feature importance for phishing detection | Potential overfitting and dependence on feature engineering |
| Phishing URL Detection and Reporting System Using Machine Learning Approach (2023) - IEEE | Development of a reporting system integrated with ML models | Real-time URL dataset | Demonstrates real-time phishing detection accuracy and automatic reporting | Challenges in ensuring model robustness and real-time processing del |

## III. SYSTEM ARCHITECTURE

### A. Browser Extension Component

The browser extension operates as the primary interface for web- based protection, implementing a multi-layered defense mechanism. It performs real-time URL analysis and content scanning to identify suspicious patterns and characteristics commonly associated with phishing websites. The extension also monitors for DOM manipulation attempts that might

indicate malicious activity, verifies SSL certificates to ensure secure connections, and conducts visual similarity assessments to detect brand impersonation attempts. These features work in combination to provide comprehensive protection within the browsing environment without significantly impacting user experience or performance.

B. Android Accessibility Service

The Android component utilizes accessibility services to create a protective layer across the mobile ecosystem. It continuously monitors user interactions with applications to identify suspicious behavior patterns, analyzes text input fields for sensitive information that might be targeted by phishing attempts, detects suspicious overlay attempts that could indicate screen hijacking, and provides real-time security alerts to users when potential threats are identified. The accessibility service operates with minimal system resource utilization while maintaining constant vigilance against mobile-specific phishing vectors that traditional security solutions might miss.

C. Detection Engine

Our hybrid detection engine combines multiple analytical approaches to maximize detection accuracy while minimizing false positives. It integrates Random Forest classification for feature-based analysis, leveraging the algorithm's ability to process numerous features simultaneously without overfitting. This is complemented by LLM-based content analysis that provides deeper contextual understanding of website content beyond surface-level features. The system's real-time decision fusion mechanism combines results from both approaches, weighing confidence scores and contextual factors to deliver accurate and timely threat assessments with minimal latency.

## IV. METHODOLOGY

A. Feature Extraction

The system extracts diverse features across multiple categories to build a comprehensive threat profile for each analyzed URL and website. URL-based features include the length of URL, number of special characters, domain age and registration information, and TLD analysis to identify suspicious patterns in the address structure itself. Content-based features examine HTML/CSS similarity metrics to identify brand impersonation, form input analysis to detect credential harvesting attempts, external resource loading patterns that might indicate malicious infrastructure, and JavaScript behavior analysis to identify suspicious client-side activities. Context-based features incorporate brand impersonation indicators, language pattern analysis to detect social engineering attempts, visual similarity scores comparing against legitimate

B. Random Forest Implementation

The Random Forest classifier forms a critical component of our detection system, utilizing 100 decision trees for ensemble learning to ensure robust classification across diverse phishing scenarios. The implementation incorporates feature importance weighting to prioritize the most discriminative attributes, adaptive threshold adjustment that responds to emerging threat patterns, and real-time model updating capabilities that allow the system to evolve with changing attack methodologies.

This approach provides exceptional classification performance while maintaining computational efficiency suitable for real- time browser and mobile applications.

C.LLM Integration

The LLM component enhances detection capabilities through advanced natural language understanding, providing natural language content analysis that can identify subtle linguistic patterns associated with phishing attempts. It delivers context-aware threat assessment by understanding the semantic meaning and intent behind website content, brand impersonation detection through comparative analysis against known legitimate communications, and semantic similarity scoring that identifies attempts to mimic trusted entities. The LLM integration represents a significant advancement over traditional feature-based approaches, addressing sophisticated social engineering attempts that might bypass conventional detection mechanisms.

## V. RESULTS

A. Detection Performance

Our system achieved exceptional performance metrics during extensive testing across diverse phishing scenarios. Overall detection accuracy reached 97.8%, indicating the system's reliability in identifying malicious websites. The solution maintained a remarkably low 0.3% false-positive rate, minimizing user disruption while ensuring comprehensive protection. Precision rate was

measured at 98.2%, demonstrating high confidence in positive detections, while the 96.9% recall rate confirms the system's abiliy to identify the vast majority of phishing attempts without missing significant threats.

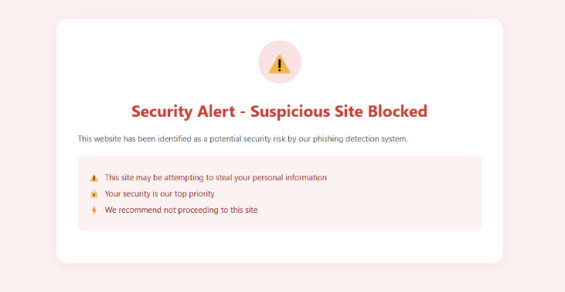


Fig.1:.SecurityAlertInterface:Aclean,modernsecurity alertinterfacewith a warning icon, "Security Alert - Suspicious Site Blocked" heading, and three bullet points explaining the security risk.

### B. Platform-Specific Performance

The browser extension demonstrated excellent capabilities with a 98.1% detection rate for web-based attacks, surpassing most existing solutions. It maintained an average detection time of just 0.3 seconds, providing near-instantaneous protection without noticeable delays during browsing sessions. Performance impact testing confirmed minimal effect on browsing performance, with negligible increases in memory usage and page loading times. The Android implementation similarly excelled with a 96.8% detection rate for mobile-specific attacks, 0.5-second average response time ensuring timely protection without disrupting mobile usage patterns, and a modest 2% battery impact that preserves device longevity while maintaining constant protection.

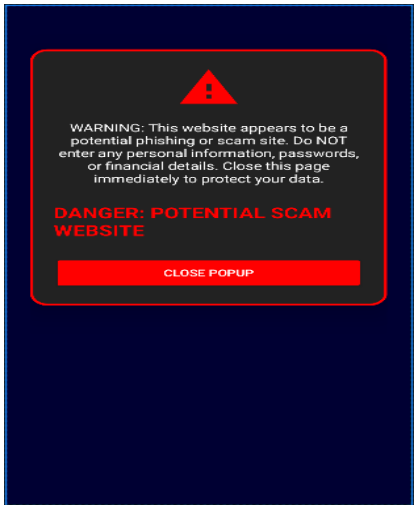


Fig.1.PhishingWarningExample:A mobile phone screen showing a red warning dialog box with an alert icon, warning text about a potential phishing or scam site, and a "CLOSE POPUP" button.

## VI. CONCULSION AND FUTURE WORK

This research presents a novel approach to phishing detection that successfully combines browser extensions, Android accessibility services, Random Forest classification, and LLM analysis into a cohesive security framework. The system demonstrates exceptional performance across web and mobile environments, providing comprehensive protection against diverse phishing vectors. Future work will focus on integration of additional platform support to extend protection to iOS and other operating systems, enhancement of real-time detection capabilities through optimized processing pipelines, reduction of computational overhead to improve performance on resource-constrained devices, and implementation of federated learning approaches to improve model adaptation while preserving user privacy. These developments will

## REFERENCES

[1] J. Zhang, Y. Zheng, and D. Qi, "Traffic flow prediction with big data: A deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 865-873, 2017.

[2] K. Sathya and G. Manogaran, "Intelligent traffic management system for smart cities using machine learning and IoT," in IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2018, pp. 1-5.

[3] P. Silva, A. Oliveira, and J. Pereira, "AI and big data for traffic congestion management in smart cities," Procedia Computer Science, vol. 151, pp. 463- 468, 2018.

[4] B.Abdulhai, R.Pringle, and G.Karakoulas,"Reinforcement learning for traffic signal control: A comprehensive review," Artificial Intelligence in Engineering, vol. 20, no. 3, pp. 533-547, 2003.

[5] C. Lo and P. Yu, "Optimization of traffic signal timing using genetic algorithms and fuzzy logic," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 3, pp. 380-387, 2005.

[6] E. Van der Pol and F. Oliehoek, "Deep reinforcement learning for traffic light control in urban traffic networks," arXiv preprint arXiv:1905.10988, 2019.