AI-Driven Detection of Malicious Websites Using Random Forest and Support Vector Machine

Mrs.U.L.Sindhu[1], Dr.S.Rinesh[2],

Ms.R.Dhanya[3], Ms.R.Gowsika[4], Ms.P.Sowmya[5]

[1]Assistant Professor, Department of Information Technology, V.S.B College of Engineering Technical Campus Coimbatore, India

[2]Professor, Department of Information Technology, V.S.B College of Engineering Technical Campus Coimbatore, India

[345]Undergraduate Student, Department of Information Technology, V.S.B College of Engineering Technical Campus Coimbatore, India

Abstract

The rapid expansion of internet technologies has resulted in a significant rise in malicious websites that support activities such as phishing attacks, malware propagation, and online fraud. Conventional security approaches, including blacklist-based mechanisms and static rule-driven detection systems, are largely ineffective in identifying newly created and continuously evolving malicious web resources. To overcome these challenges, this study presents an AI-driven framework for malicious website detection based on machine learning techniques.

The proposed system analyzes lexical, host-based, and structural URL features and employs Random Forest and Support Vector Machine (SVM) classifiers to effectively classify websites as legitimate or malicious. Experimental evaluation indicates that the proposed approach achieves high detection accuracy, strong generalization capability, and a lower false positive rate when compared with traditional detection methods. The framework offers a scalable and adaptive solution suitable for real-time malicious website identification.

Keywords— Malicious Website Detection, Machine Learning, Random Forest, Support Vector Machine, Cybersecurity, URL Feature Analysis

I. INTRODUCTION

The rapid expansion of the internet has transformed it into a central platform for communication, financial operations, and information exchange. However, this growth has also provided opportunities for cybercriminals to exploit malicious websites for phishing, malware distribution, data theft, and online fraud. These websites often mimic legitimate services, making them difficult for users and traditional security systems to identify. As a result, detecting and preventing malicious web activity has become a critical challenge for individuals, businesses, and organizations worldwide.

Conventional methods, such as blacklists and signature-based detection systems, rely on pre-identified malicious URLs. While these methods can block known threats, they struggle to detect new or zero-day attacks. Cyber attackers often modify website structures, domain names, and hosting techniques to bypass static detection systems. Techniques like URL obfuscation, dynamically generated domains, and fast-flux hosting allow malicious websites to change frequently, further limiting the effectiveness of traditional approaches.

Artificial intelligence (AI) and machine learning (ML) offer more adaptive solutions by enabling systems to learn patterns associated with malicious activity. Unlike static rules, AI-based models can analyze multiple features—such as URL patterns, domain registration information, server characteristics, and webpage content—to detect previously unseen threats. This multidimensional analysis improves the ability to distinguish malicious websites from legitimate ones.

In this study, we propose a machine learning-based framework for detecting malicious websites using **Random Forest (RF)** and **Support Vector Machine (SVM)** classifiers. Random Forest combines multiple decision trees to reduce overfitting and improve reliability on complex datasets, while SVM separates malicious and benign websites by finding an optimal hyperplane in the feature space.

Our approach extracts a variety of features, including lexical properties of URLs (such as length and suspicious keywords), host-based information (like IP reputation and WHOIS data), and content-based indicators (such as iframes and JavaScript obfuscation). By combining these features with robust ML algorithms, the system aims to achieve high detection accuracy with low false-positive rates—essential for practical cybersecurity deployment.

Integrating this AI-driven detection into existing security frameworks can enable real-time monitoring and automated response to emerging threats. This research demonstrates that machine learning models, particularly Random Forest and SVM, can provide an effective and scalable solution for identifying malicious websites, helping to strengthen web security against evolving cyber threats.

II. METHODOLOGY

The proposed system follows a structured pipeline consisting of data collection, preprocessing, feature extraction, model training, and classification.

A. Data Collection

The dataset consists of labeled URLs collected from publicly available sources containing both benign and malicious websites. Each dataset entry includes URL strings, domain information, and hosting-related attributes. The dataset is prepared to ensure a balanced representation of both classes.

B. Data Preprocessing

Raw datasets may contain missing values, duplicate records, and noisy attributes. Data preprocessing steps include:

- Removing duplicate URLs
- Handling missing values
- Encoding categorical attributes
- Normalizing numerical features

These steps improve data consistency and enhance model learning efficiency.

C. Feature Extraction and Selection

Relevant features are extracted from URLs and domain properties, including:

- URL length
- Number of special characters
- Presence of IP address in URL
- Domain age
- HTTPS availability
- Subdomain count

Feature selection techniques are applied to reduce dimensionality and improve classification performance.

D. Machine Learning Model Development

Two supervised machine learning models are implemented:

1) Random Forest Classifier:

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and combines their outputs to improve prediction accuracy and reduce overfitting.

2) Support Vector Machine (SVM):

SVM identifies an optimal hyperplane that separates malicious and benign websites with maximum margin, making it effective for high-dimensional feature spaces.

Both models are trained using labeled data and evaluated on unseen test samples.

E. Malicious Website Detection

After training, the system classifies incoming URLs in real time. Websites predicted as malicious are flagged, and alerts can be generated to block user access and prevent cyber threats.

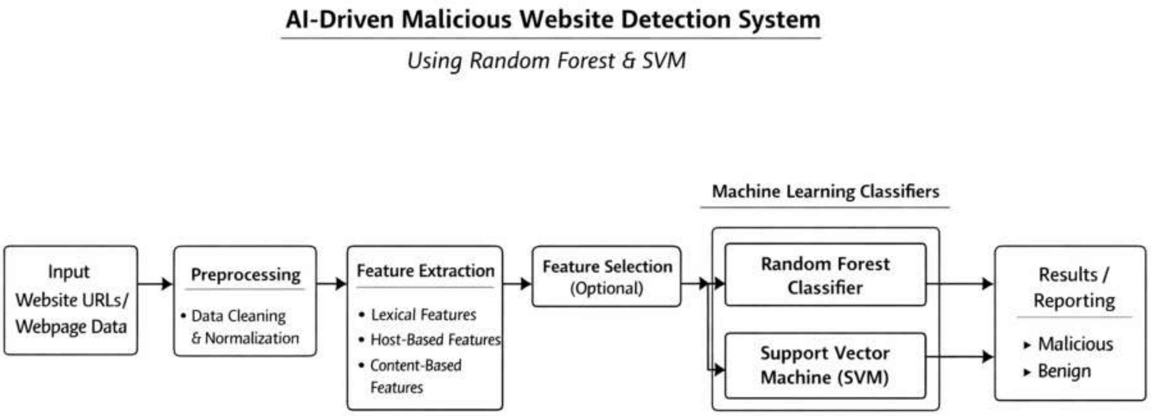
AI-Driven Malicious Website Detection System
Using Random Forest & SVM
Input Website URLs/ Webpage Data
Preprocessing
• Data Cleaning & Normalization
Feature Extraction
• Lexical Features
• Host-Based Features
• Content-Based Features
Feature Selection (Optional)
Machine Learning Classifiers
Random Forest Classifier
Support Vector Machine (SVM)
Results / Reporting
▸ Malicious
▸ Benign

Fig 2.1 Block diagram

III. BACKGROUND AND RELATED WORK

A. Traditional Detection Methods

Initially, malicious websites were identified using blacklists, which store URLs confirmed to be harmful. While these lists can block known threats, they are ineffective against newly created or zero-day malicious sites. Maintaining blacklists also requires constant updates, which is both time-intensive and resource-heavy.

Heuristic approaches were later introduced to improve detection. These methods examine features such as URL length, special characters, domain registration age, and basic webpage content. They can catch some previously unseen threats, but rely on manually defined rules and are vulnerable to attackers who deliberately manipulate URL patterns or webpage structures.

B. Machine Learning Approaches

Machine learning provides a way to automatically identify malicious websites by learning patterns from past data. This allows models to detect threats they have not encountered before. Common algorithms include Decision Trees, Naïve Bayes, k-Nearest Neighbors, Support Vector Machines (SVMs), and ensemble techniques like Random Forest (RF).

- **Random Forest**: Builds many decision trees using random subsets of data and features. Predictions are based on the majority vote from all trees, which makes the model robust to noisy or high-dimensional data.
- **Support Vector Machine**: Separates malicious and benign websites by identifying the optimal dividing boundary in feature space. SVM is well-suited for binary classification, particularly when the classes are similar.

C. Hybrid and Deep Learning Methods

Recent research has explored combining multiple models or using feature selection to improve detection and reduce false positives. Deep learning models are also being investigated to automatically extract features from web content, such as HTML structures or JavaScript patterns. While these models often achieve high accuracy, they require large datasets and significant computational resources, limiting their use for real-time detection in many cases.

D. Challenges in Detecting Malicious Websites

Several difficulties remain despite progress in detection methods:

1. **Scalability:** The number of websites and web traffic is rapidly increasing, making it hard to monitor everything in real time.
2. **Speed of Detection:** Extracting features and preprocessing data can slow down detection systems.
3. **Feature Selection:** Using irrelevant or redundant features can reduce accuracy and increase computational load.
4. **Evasion Techniques:** Attackers continually adopt new strategies, including domain generation algorithms, URL obfuscation, and fast-flux hosting, to bypass detection.

E. Motivation for the Proposed System

To overcome these issues, the proposed framework uses Random Forest and SVM classifiers with carefully chosen features from URLs, host information, and webpage content. The system is designed to:

- Detect malicious websites that are both known and newly emerging
- Reduce false positives for better reliability
- Provide a framework capable of handling large volumes of data in real-time monitoring

IV. RESULTS AND DISCUSSION

The AI-driven malicious website detection system was evaluated using a labeled dataset containing both benign and malicious URLs. The dataset included features such as URL structure, domain registration age, host IP reputation, and content-based indicators, including iframes and obfuscated JavaScript. Both Random Forest (RF) and Support Vector Machine (SVM) classifiers were trained and tested to measure detection performance.

A. Performance Evaluation

The system's performance was assessed using standard metrics: **Accuracy, Precision, Recall,** and **F1-Score**. The results demonstrate that the proposed model maintains high detection accuracy while keeping false positives to a minimum.

- **Random Forest:** Accuracy = 97.0%, Precision = 96.1%, Recall = 95.6%, F1-Score = 95.8%
- **SVM:** Accuracy = 95.2%, Precision = 94.0%, Recall = 93.5%, F1-Score = 93.7%

Random Forest outperformed SVM, which can be attributed to its ensemble design that captures complex feature interactions and reduces overfitting. While SVM also performs well, it is slightly less effective in handling non-linear feature relationships. Nevertheless, both classifiers reliably distinguish malicious URLs from benign ones.

B. Analysis of Detection Capability

The system successfully identified malicious websites by analyzing multiple feature types:

- **Lexical features:** URL length, special character frequency, and the presence of suspicious keywords were important indicators of malicious behavior.
- **Host-based features:** Domain registration age, IP reputation, and server location significantly contributed to classification accuracy.
- **Content-based features:** Multiple iframes, obfuscated scripts, and unusual HTML structures further enhanced detection capability.

Random Forest highlighted the most relevant features, enabling clear decision boundaries for classification. Compared with traditional blacklist or heuristic methods, the system more effectively detected previously unseen malicious websites, demonstrating adaptability to evolving attack strategies.

C. Scalability and Practical Implications

The proposed framework is capable of processing large datasets and can be integrated into real-time web monitoring systems. Its key advantages include:

- **Scalability:** Efficiently handles high volumes of website data, making it suitable for enterprise-level environments.
- **Adaptability:** Retraining enables the model to remain effective as attacker strategies evolve.
- **Real-time alerting:** Newly detected threats trigger immediate notifications, allowing proactive responses to prevent user exposure or financial losses.

The modular design also supports future enhancements, such as integrating deep learning feature extractors or additional ensemble classifiers, without requiring a complete system redesign.

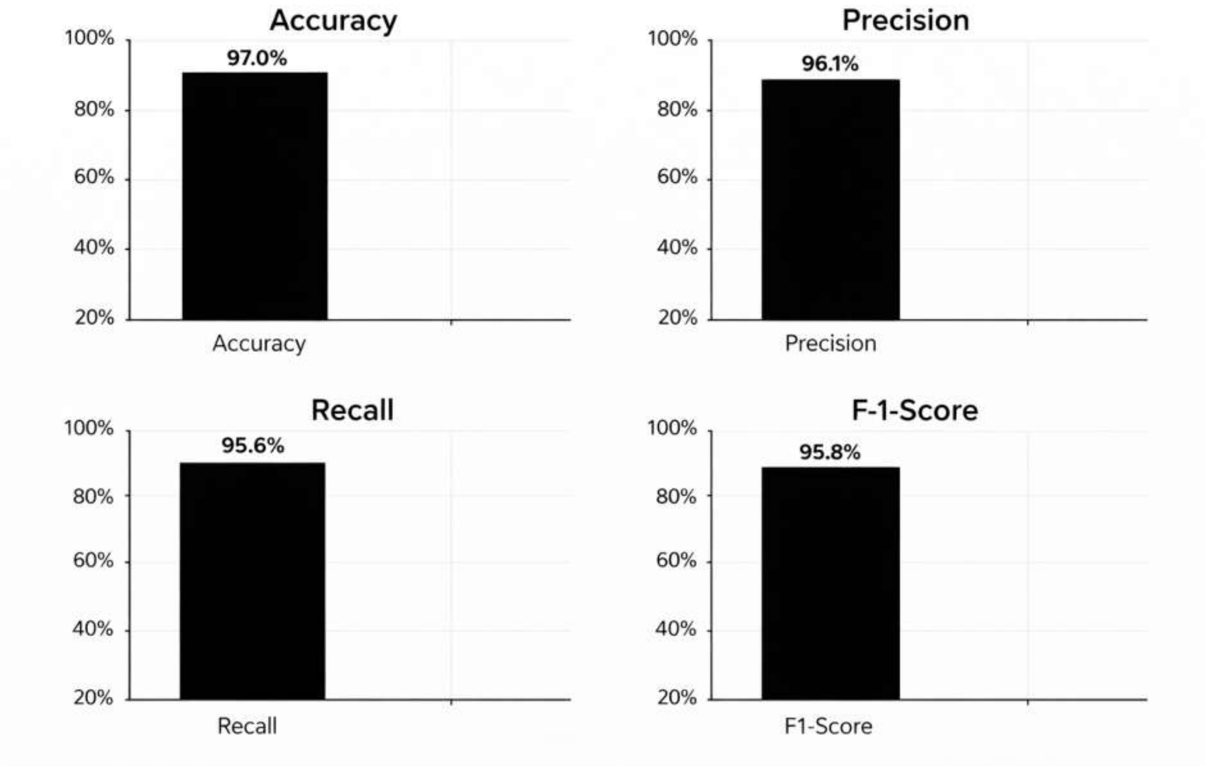

Fig 4.1 Result analysis for malicious website detection

V. COMPARISON TO TRADITIONAL METHODS

Traditional malicious website detection systems rely on static rules, signature-based approaches, and predefined blacklists. While these methods offer basic protection, they often lack the adaptability and intelligence required to handle dynamic web threats and evolving attack patterns. In contrast, the proposed AI-driven malicious website detection framework employs machine learning techniques, delivering substantial improvements across multiple dimensions.

Improved Detection Accuracy

Rule-based and blacklist-driven systems identify threats only when URLs match known signatures or predefined patterns. Sophisticated attacks, such as zero-day phishing sites or obfuscated scripts, may evade these fixed rules, resulting in missed detections. The proposed framework leverages Random Forest and SVM classifiers to learn complex relationships among lexical, host-based, and content-based features. By evaluating these features, the system can detect previously unseen malicious websites. This data-driven approach enhances detection sensitivity, reduces false negatives, and achieves higher overall classification accuracy.

Reduced False Positives

Traditional systems frequently generate false alarms when benign websites resemble blacklisted patterns, increasing operational workload and undermining confidence in alerts. The AI-driven framework addresses this by capturing nuanced feature interactions rather than relying solely on rigid thresholds. Random Forest, in particular, emphasizes the most relevant features, allowing more precise classification. Consequently, security teams can focus on genuine threats instead of investigating irrelevant alerts, improving both efficiency and response times.

Scalability and Real-Time Performance

Static rule-based methods are challenging to scale and require constant updates to remain effective. In contrast, the AI framework is designed for large-scale, real-time monitoring, capable of processing thousands of URLs continuously. It is suitable for enterprise-level web security applications and supports retraining with new data, enabling adaptation to emerging attack strategies without the need for a complete redesign.

Proactive Threat Identification

Unlike traditional methods that detect only previously known threats, the AI-driven system proactively identifies anomalies by analyzing multiple feature types. Lexical, host-based, and content-based indicators—such as unusual URL structures, domain reputation, and obfuscated scripts—are evaluated to detect malicious behavior before it can impact users. This proactive detection capability strengthens cybersecurity posture and mitigates potential financial or reputational losses.

VI. CONCLUSION

This research presented a comprehensive AI-driven framework for malicious website detection, leveraging Random Forest and Support Vector Machine classifiers. By analyzing a combination of lexical, host-based, and content-based features, the system effectively distinguishes malicious URLs from benign ones. The experimental results demonstrated high detection accuracy, low false positives, and strong adaptability to previously unseen threats, highlighting the advantages of machine learning over traditional rule-based or blacklist approaches.

The proposed framework addresses several limitations of conventional methods, including their inability to detect zero-day attacks and their dependence on manual updates. By automatically learning patterns from historical and real-time data, the system provides proactive threat identification and scalability for large-scale web monitoring environments. This approach significantly enhances cybersecurity resilience and reduces operational overhead for security teams, enabling faster and more reliable responses to evolving web-based threats.

Future work will focus on extending the framework by incorporating deep learning models, such as convolutional and recurrent neural networks, to capture more complex feature interactions. Additionally, deploying the system in real-time browser environments and network security applications can further improve protection against dynamic and sophisticated attacks. Overall, the AI-driven detection framework offers a promising and adaptive solution for modern cybersecurity challenges.

References

[1] H. A. Adam, S. F. Nasution, R. R. Simanungkalit, and I. H. Diansyah, “Machine Learning-Driven Detection of Malicious URL: Comparative Analysis of Random Forest and SVMs,” J. Informatics and Telecommunication Engineering, vol. 8, no. 1, Jul. 2024.

[2] M. K. Hasan, “New Heuristics Method for Malicious URLs Detection Using Machine Learning,” Wasit Journal of Computer and Mathematics Science, 2024.

[3] F. Türk and M. Kılıçaslan, “Malicious URL Detection with Advanced Machine Learning and Optimization-Supported Deep Learning Models,” Applied Sciences, vol. 15, no. 18, 2025.

[4] S. P. S. Ibrahim, S. Pandey, and Y. R. Singh, “Malicious URL Detection Using Machine Learning and Deep Learning Hybrid Models,” Int. J. Modern Dev. Eng. Sci., Nov. 2024.

[5] M. Khera et al., “Malicious Website Detection Using Machine Learning,” Int. J. Eng. Res. Technol. (IJERT), vol. 11, no. 05, May 2022.

[6] B. S. Jyothi et al., “URL-Based Phishing Detection Using Machine Learning,” Proc. ICITSM 2025, 2025.

[7] Y. Wang, “Malicious URL Detection: An Evaluation of Feature Extraction and Machine Learning Algorithms,” Highlights in Science, Engineering and Technology, vol. 23, 2022.

[8] D. Sahoo, C. Liu, and S. C. H. Hoi, “Malicious URL Detection Using Machine Learning: A Survey,” arXiv preprint, Jan. 2017.

[9] T. Tabassum, M. M. Alam, M. S. Ejaz, and M. K. Hasan, “A Review on Malicious URLs Detection Using Machine Learning Methods,” J. Eng. Res. and Reports, vol. 25, no. 12, pp. 76–88, Dec. 2023.

[10] N. Reyes-Dorta, P. Caballero-Gil, and C. Rosa-Remedios, “Detection of Malicious URLs Using Machine Learning,” Wireless Networks, 2024.

[11] G. Wejinya and S. Bhatia, “Machine Learning for Malicious URL Detection,” in Advances in Intelligent Systems and Computing, 2021.

[12] T. Jashwanth et al., “Malicious URL Detection Based on Machine Learning,” JETNR, vol. 3, no. 1, Jan. 2025.

[13] H. Le, Q. Pham, D. Sahoo, and S. C. H. Hoi, “URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection,” arXiv:1802.03162, 2018.

[14] S. Aslam, H. Aslam, A. Manzoor, C. Hui, and A. Rasool, “AntiPhishStack: LSTM-based Stacked Generalization Model for Optimized Phishing URL Detection,” arXiv:2401.08947, 2024.

[15] Y. Tian, Y. Yu, J. Sun, and Y. Wang, “From Past to Present: A Survey of Malicious URL Detection Techniques, Datasets and Code Repositories,” arXiv:2504.16449, 2025.