Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1271

AI-BASED PHISHING WEBSITE DETECTION

Nishanshu Deshmukh[1], Nishant Sahu[2], Sahil Darokar[3], Sangam Jain[4], Prof. Ravi Kumar Mohane[5]

[5]Guide

Abstract:
Phishing remains a major cyber threat, with traditional defences like blacklists proving ineffective against rapidly evolving attacks. This review highlights the shift toward Artificial Intelligence (AI) and Machine Learning (ML) as adaptive solutions. It outlines a methodology involving feature engineering—using URL, domain, and content features—and applying classifiers such as Decision Trees, Random Forests, and Support Vector Machines (SVM). Comparative analysis shows ensemble models like Random Forest offer superior accuracy and generalisation. The paper surveys advancements from basic ML to modern Deep Learning and hybrid systems, identifies gaps such as vulnerability to adversarial attacks, and emphasizes the need for real-time, continuous learning detection mechanisms.

1.INTRODUCTION

In today's digitally connected world, the internet is essential for communication, commerce, and governance—but this dependence exposes users to rising cyber threats, notably phishing. Phishing, which deceives users into revealing sensitive information by imitating legitimate entities, has become a leading cause of identity theft and financial loss. Traditional detection methods like blacklists and whitelists are reactive and ineffective against zero-day phishing sites that exploit victims before being flagged. Attackers also evade these defences through URL obfuscation, typo squatting, and rapid domain changes.

To address these challenges, cybersecurity has shifted toward Artificial Intelligence (AI) and Machine Learning (ML). These techniques proactively identify hidden patterns in URLs, domain metadata, and webpage content, enabling real-time detection. This research reviews current AI-based phishing detection systems, analysing methodologies, algorithms, and performance metrics. It provides a comparative evaluation of common ML approaches and outlines future directions for building robust, adaptive, and real-time phishing detection frameworks.

2.LITERATURE SURVEY

The body of research on phishing detection has undergone a rapid evolution, moving from rudimentary manual checks to highly sophisticated, automated AI systems. This survey traces the key methodological milestones that have defined this progression, categorising the approaches by their underlying detection principles.

1. Aburrous et al. (2010) - This paper presents an intelligent system to protect e-banking users from phishing. Unlike basic blacklisting, it uses a rule-based expert system that analyses URL, domain, and content features to calculate a risk score. This heuristic approach achieved higher accuracy and laid the groundwork for later feature-based, machine learning–driven detection methods.

2. Jain and Gupta (2020) - This paper reviews and compares Machine Learning (ML) techniques for phishing detection, evaluating classifiers like Decision Tree, Random Forest, SVM, and Naïve Bayes using metrics such as accuracy, precision, and recall. It identifies the most effective models for different feature

Advanced International Journal of Multidisciplinary Research

E-ISSN: 2584-0487
editor@aijmr.com

Volume 4, Issue 2, March-April 2026

CrossRef DOI: 10.62127/aijmr.2026.v04i02.1271

sets (URL-based, content-based, etc.) and serves as a key guide for selecting optimal ML algorithms in phishing detection.

3. Shahzad and Aman (2024) - This paper evaluates the effectiveness of AI-based algorithms for detecting modern, sophisticated phishing attacks. It compares traditional ML and deep learning models, showing that advanced approaches like CNNs and ensemble methods achieve superior detection rates. The study guides the selection of suitable AI models for developing robust, current anti-phishing systems.

4. Islam et al. (2024) - This paper presents a comprehensive review of phishing detection systems that rely solely on URL-based Machine Learning approaches. It examines methods for extracting lexical and structural URL features, compares ML model performance on standard datasets, and highlights trends, challenges, and future directions for lightweight, real-time phishing detection through URL analysis.

5. UCI Machine Learning Repository - The Phishing Websites Dataset is a widely used public resource containing labelled data for training and testing phishing detection models. It includes thousands of websites—legitimate or phishing—described by features from URLs and source code. This dataset supports reproducible research and comparative studies in AI-based cybersecurity.

6. PhishTank - PhishTank is a community-driven, continuously updated database of verified phishing websites. It provides real-time blacklisting services and serves as a key source of fresh, ground-truth data for researchers, helping train and test ML models against the latest phishing attacks.

7. Do et al. (2022) - This paper presents a comprehensive review of Deep Learning (DL) techniques for phishing detection, categorizing architectures like CNNs and RNNs and analysing their strengths and limitations. It highlights challenges such as data requirements and adversarial vulnerabilities, offering recommendations and future research directions for DL-based anti-phishing systems.

8. Al-Sarem et al. (2021) - This paper introduces an optimized stacking ensemble model for phishing website detection, combining multiple classifiers through a meta-classifier for improved accuracy. The approach outperforms individual models, achieving higher accuracy and lower false positives, demonstrating the effectiveness of advanced ensemble learning for robust detection.

3.METHODOLOGY

The methodology for developing the detection system is systematically structured into four core phases, ensuring high-quality data and optimised model performance:

1. **Data Collection and Preprocessing:** Raw datasets, such as those from the UCI Repository, Kaggle, and Phish Tank, serve as the foundation. The data, containing labelled legitimate and phishing URLs, undergoes rigorous cleaning. This includes normalisation, handling missing values, and ensuring a balanced distribution of both classes to prevent model bias.
2. **Feature Extraction:** This phase transforms raw data (the URL) into quantifiable inputs for the ML model. Critical features considered include:

- **URL Structure:** Length, presence of suspicious characters (@, // after the protocol), use of IP addresses instead of domain names, and the count of sub-domains.
- **Domain Features:** Domain registration age, expiration date, and Whois records (often obscured or very recent for phishing sites).
- **Security and Content Features:** Presence and validity of an SSL/TLS certificate, the use of pop-up windows, and the presence of deceptive embedded scripts.

3. **Model Training and Selection:** Supervised Machine Learning algorithms (Decision Tree, Random Forest, SVM) are trained using the extracted feature vector. Hyperparameter tuning is essential in this phase to optimise each model's internal settings for the best possible accuracy and generalisation.

4. **Evaluation and Testing:** Model performance is assessed using standard classification metrics: Accuracy, Precision, Recall, and F1-score. High Recall is critical, as it minimises False Negatives (failing to identify a phishing site), while high Precision ensures a low False Positive rate (avoiding flagging a legitimate site as malicious). The best-performing model, often an ensemble technique, is then selected for deployment.

In this project we use the different algorithm and check the accuracy. The algorithm are Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) to classify websites as legitimate or malicious. This selection balances the need for interpretability with high predictive accuracy.

Decision Tree (DT)

Decision Tree is a supervised machine learning algorithm that is widely used for classification and prediction tasks. In this project, the Decision Tree algorithm is used to classify websites as legitimate or phishing based on extracted features such as URL length, presence of HTTPS, domain age, and HTML attributes.

A Decision Tree works by creating a tree-like structure where each internal node represents a condition or test on a feature, each branch represents the outcome of that test, and each leaf node represents the final classification result. The model splits the dataset into smaller subsets based on feature values, selecting the most important features using metrics like Information Gain or Gini Index.

One of the major advantages of Decision Trees is their simplicity and interpretability. The decision-making process can be easily visualized, making it understandable even for non-technical users. However, Decision Trees can suffer from overfitting, especially when the tree becomes too deep and starts memorizing the training data.

In this project, Decision Tree provides a baseline model for comparison. Although its accuracy is slightly lower than ensemble methods, it helps in understanding how different features contribute to phishing detection and provides a clear rule-based classification approach.

Random Forest (RF)

Random Forest is an advanced ensemble learning algorithm that improves the performance of Decision Trees by combining multiple trees to produce a more accurate and stable prediction. In this project, Random Forest is used to classify websites as phishing or legitimate by analyzing various extracted features.

The algorithm works by creating multiple Decision Trees during training and combining their outputs using a voting mechanism. Each tree is trained on a random subset of the dataset and uses a random subset of features, which reduces overfitting and improves generalization. This technique is known as bagging (Bootstrap Aggregation).

Random Forest is highly effective for phishing detection because it can handle complex relationships between features such as URL structure, domain characteristics, and HTML content. It is also robust to noise and outliers in the dataset.

E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1271

One of the key advantages of Random Forest is its high accuracy and reliability compared to a single Decision Tree. It also provides feature importance scores, helping to identify which features contribute most to classification.

In this project, Random Forest achieved better performance than Decision Tree and provided more consistent results. Its ability to reduce overfitting and improve prediction accuracy makes it a strong model for detecting phishing websites.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for classification tasks. In this project, SVM is applied to classify websites as legitimate or phishing based on extracted features from URLs, domains, and webpage content.

SVM works by finding the optimal boundary, known as a hyperplane, that separates data points into different classes. The goal is to maximize the margin between the two classes, ensuring better generalization on unseen data. For complex datasets where data is not linearly separable, SVM uses kernel functions such as linear, polynomial, or radial basis function (RBF) to transform the data into higher dimensions.

One of the major strengths of SVM is its ability to perform well on high-dimensional data and handle complex patterns effectively. It is particularly useful in phishing detection where multiple features interact in non-linear ways.

However, SVM can be computationally expensive and requires careful tuning of parameters such as kernel type and regularization.

In this project, SVM provided strong classification performance and helped improve detection accuracy. It is especially effective in distinguishing subtle differences between phishing and legitimate websites, making it a valuable model in the system.

4.RESULTS:

The developed system for AI-based phishing website detection was successfully implemented and tested using multiple website URLs. The results demonstrate the effectiveness of the model in identifying whether a website is legitimate or phishing. The system achieved an accuracy of approximately 97.4%, indicating high reliability and precision in classification.

The model was trained on a well-structured dataset collected from trusted sources such as the UCI Machine Learning Repository (Phishing Websites Dataset) and PhishTank, which provide real-world phishing and legitimate URLs. These datasets include various features related to URL structure, domain properties, and

HTML characteristics, enabling the model to learn and detect complex phishing patterns.

Overall, the high accuracy and use of real-world datasets ensure that the system performs efficiently in detecting even previously unseen phishing websites, making it a reliable solution for enhancing cybersecurity.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 4, Issue 2, March-April 2026
CrossRef DOI: 10.62127/aijmr.2026.v04i02.1271

Overall Performance

The AI-based model shows good performance in distinguishing between legitimate and phishing websites. The system provides fast results and a clear output, making it useful for real-time applications. The user-friendly interface combined with accurate predictions makes the solution effective for enhancing online security.

CONCLUSION

Phishing threats demand intelligent, continuous defence, and this review confirms AI and Machine Learning as the most effective tools. By replacing reactive lists with proactive, pattern-based models, detection systems now better counter evolving attacks. Feature engineering across URL, domain, and content layers, combined with robust classifiers—especially ensemble methods like Random Forest—has proven highly accurate for real-world use. Though challenges remain, such as adversarial resilience and real-time deployment, the future lies in hybrid, continuous learning systems and lightweight deep learning models, ensuring stronger, adaptive protection against phishing.

The system offers a simple and user-friendly interface, allowing users to easily input URLs and receive quick results. This makes it suitable for real-time usage, even for non-technical users. The use of artificial intelligence improves the efficiency and reliability of detection compared to traditional methods.

In the future, the system can be enhanced by integrating advanced machine learning algorithms, real-time threat intelligence, and browser extensions for automatic detection. Overall, the proposed system is an effective solution for improving online security and protecting users from phishing attacks.

REFERENCES:

1. Aburrous, M., M.A. Hossain, K. Dahal, and F. Thabtah. "Intelligent Phishing Detection System for E-banking." *Expert Systems with Applications*, vol. 37, no. 12, 2010, doi:10.1016/j.eswa.2010.02.068.
2. Jain, A.K., and B.B. Gupta. "Phishing Detection: Analysis of Machine Learning Techniques." *Security and Privacy*, vol. 3, no. 5, 2020, doi:10.1002/spy2.93.
3. Shahzad, T., & Aman, K. (2024). "Unveiling the Efficacy of AI-based Algorithms in Phishing Attack Detection." *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 116–133.
4. Islam, Jahirul, et al. "Phishing URL Detection via Machine Learning: A Comprehensive Survey." *International Journal on Artificial Intelligence Tools*, vol. 33, no. 5, 2024.
5. *UCI Machine Learning Repository - Phishing Websites Dataset*. Kaggle, The University of California, Irvine. Web.
6. *PhishTank - Community-driven Phishing Database*. OpenDNS, LLC. Web.
7. Do, N. Q., et al. "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions." *IEEE Access*, vol. 10, 2022, pp. 36431-36449.
8. Al-Sarem, M., et al. "An Optimized Stacking Ensemble Model for Phishing Websites Detection." *Electronics*, vol. 10, no. 11, 2021, p. 1285.
9. Patil, D. R., Wagh, R. B., Punjabi, V. D., & Pardeshi, S. M. (2024). *Enhanced Phishing URL Detection using Feature Selection and Machine Learning. IJWMT, 14*(6), 48–67. doi:10.5815/ijwmt.2024.06.04