Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

Real-Time AI/ML-Based Phishing Detection and Prevention System

[1] Neha Yogesh Nagdeve, [2]Gauri V. Naktode, [3]Rajat Rokade, [4]Kartiket kamdi

[1,2,3,4]Information Technology, St. Vincent Pallotti College of Engineering & Technology, Nagpur, India

Email: [1]Sighjass.singh@gmail.com, [2]gaurinaktode27@gmail.com, [3]rajatrokade185@gmail.com, [4] kartiketnkamdi@gmail.com

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Phishing, Artificial Intelligence (AI), Machine Learning, Phishing Detection, Cyber security

Abstract

Phishing attacks, which use phony emails, websites, and messages to trick users into disclosing private information like passwords or bank account information, have grown to be a significant cyber security concern. Phishing attacks are fraudulent attempts to obtain sensitive data by posing as a reliable individual or business. These attacks are growing more frequent and sophisticated. Antiquated phishing detection techniques that rely on set rules or recognized patterns frequently fall behind new phishing techniques. The potential of artificial intelligence (AI) to enhance phishing detection systems is discussed in this paper. AI detects and halts phishing more rapidly and precisely by using methods like machine learning, natural language processing, and pattern recognition. AI can identify subtle phishing indicators that traditional systems might overlook by analyzing vast volumes of data. The article also covers a variety of AI approaches, including deep learning, ensemble approaches, and supervised and unsupervised learning. It examines how well these AI systems function in practical settings and how well they adapt to novel phishing techniques. The study concludes by discussing the difficulties and potential advancements that will be required to address the ever-evolving and evolving phishing threats.

Introduction

Phishing has emerged as one of the most prevalent and hazardous cyber threats in the current digital age, aiming to steal private data, including financial information, login credentials, and personal information, from both individuals and organizations. The goal of this project is to create a real-time AI/ML-based phishing detection and prevention system that uses artificial intelligence models and machine learning algorithms to quickly detect and stop phishing attempts. The system can precisely identify questionable patterns and alert users in real time by analyzing different aspects of emails, URLs, and websites. This system's primary objective is to increase online security by employing intelligent, data-driven techniques that can swiftly pick up and adjust to new phishing techniques, assisting users in staying safe online.

In order to identify and anticipate potential cyberattacks, AI-based systems can process and analyze vast volumes of data, significantly enhancing cyber security protection. These automated and prompt systems can lower risks, preventing major harm and enabling people or organizations to carry on with their work in a safe manner. Businesses can better manage the intricate problems of cyber security by utilizing AI, which provides them with a robust and flexible defense against online dangers such as phishing. Phishing attempts and data leaks can be greatly decreased for both individuals and businesses when AI is incorporated into cyber security systems. Artificial intelligence (AI) tools in the real world use machine learning algorithms to analyze email content and structure, spot questionable trends, and identify phishing messages. This makes online communication safer by preventing malicious emails from getting to a user's inbox [1].

A. Scope

To develop an intelligent system that uses Artificial Intelligence (AI) and Machine Learning (ML) to detect and prevent phishing attacks in real time, ensuring safe and secure online communication.

B. Objective

1. To employ AI and ML techniques to analyze and detect phishing emails, messages, and websites.
2. To create a real-time detection model capable of immediately identifying and stopping phishing attempts.
3. To evaluate the effectiveness of different machine learning algorithms for precise phishing detection.
4. To develop a system that automatically learns and adjusts to novel phishing techniques.
5. To raise awareness of cyber security and lessen phishing-attack-related data theft.

C. Purpose

By creating an intelligent and automated phishing detection system, this project aims to safeguard users and organizations against online fraud. The system can swiftly identify questionable activity, block access to damaging content, and fortify overall cyber security defenses by utilizing AI and ML.

Literature Review

Phishing attack detection has been the subject of numerous studies, with a primary focus on email and website security. Blacklists and rule-based filtering were straightforward earlier approaches that were unable to keep up with evolving phishing tactics. To increase the accuracy of phishing detection, researchers have recently begun utilizing Artificial Intelligence (AI) techniques that integrate Natural Language Processing (NLP), Machine Learning (ML), and Computer Vision (CV). A major issue with cyber security is the increase in phishing attacks. Attackers deceive users with phony attachments, links, and messages. Because hackers regularly alter their techniques, conventional techniques like rule-based filtering and blacklisting frequently fall short in identifying novel or unknown (zero-day) phishing attacks. AI-based systems that employ ML, NLP, and CV have been created to address this issue and increase detection accuracy.

Prior phishing detection techniques examined elements like text content, domain age, and message patterns. To increase the accuracy of URL classification, ensemble models that combine Random Forest and XGBoost have been employed. Rule-based methods still have trouble with emerging phishing patterns, though. This problem has led to the development of AI-based models that can learn from changing threats. One such model is deep reinforcement learning, which can adjust to new phishing tactics based on user interactions, significantly enhancing real-time detection performance [3]. Extensive research into different detection techniques has been spurred by the sharp rise in phishing attacks, underscoring the necessity for systems that can successfully adjust to changing threats. The use of artificial intelligence (AI) to improve the precision and responsiveness of phishing detection has gained attention in recent years. The development of phishing detection methods is examined in this literature review, with a focus on the use of AI technologies to increase detection effectiveness and flexibility in the face of new phishing tactics [2].

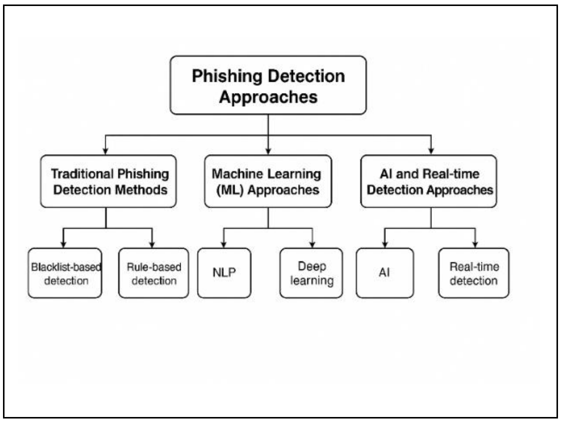
Phishing Detection Approaches
Traditional Phishing Detection Methods
Machine Learning (ML) Approaches
AI and Real-time Detection Approaches
Blacklist-based detection
Rule-based detection
NLP
Deep learning
AI
Real-time detection

Figure 1: Phishing Detection Approaches

Phishing detection has evolved from static, rule-based systems to intelligent, adaptive, and real-time AI-driven models, as illustrated in Figure 1. The various methods for spotting phishing attempts are compiled in the "Phishing Detection Approaches" diagram. It divides the detection of phishing into three primary categories:

1. Traditional Methods: These include rule-based and blacklist-based detection, which depend on pre-established rules and recognized dangerous URLs. These are straightforward but useless against fresh or developing attacks.
2. Machine Learning (ML) Techniques: Learn phishing trends from data by using algorithms. To automatically identify suspicious activity, methods like Deep Learning and Natural Language Processing (NLP) examine text, URLs, and webpage structures.

3. AI and Real-Time Detection Techniques: To quickly identify new phishing threats and adjust to evolving attack tactics, combine cutting-edge AI with real-time monitoring.

A. Traditional Phishing Detection Methods

Since hackers utilize phony messages, links, and attachments to fool users into disclosing private information, phishing attacks remain a major danger to cyber security. Long employed to safeguard consumers are conventional detection techniques like rule-based filtering, blacklists, heuristic analysis, and email content scanning. These methods use known malicious patterns to compare emails or URLs in order to identify phishing. These techniques work well for dangers that have already been discovered, but they are unable to identify zero-day phishing attempts [5].

B. Machine Learning Approaches

In order to get over conventional constraints, researchers started utilizing Machine Learning (ML) methods. ML models don't just rely on predetermined rules; they also learn patterns from massive datasets. Emails and URLs are categorized as safe or phishing using algorithms like XGBoost, Random Forest, Decision Trees, and Support Vector Machines (SVM) [6]. XGBoost and Random Forest are examples of ensemble models that combine several models to improve accuracy. To be effective against emerging phishing tactics, machine learning systems need to be updated frequently, have high-quality datasets, and carefully choose their features [8].

1. Data Collection: Compile website information, URLs, and emails.
2. Feature Extraction: Determine important attributes such as domain age or URL length.
3. Feature Selection: To improve accuracy, pick the most helpful attributes.
4. Model Training: To identify phishing, employ supervised or unsupervised learning.
5. Model Evaluation: Use precision and accuracy to assess performance.
6. Real-Time Detection: Quickly identify fresh emails or URLs as either phishing or safe.

C. Natural Language Processing (NLP) in Phishing Detection

Natural Language Processing (NLP) focuses on understanding and analysing the text used in phishing messages, emails, or websites. NLP detects suspicious patterns like urgent tone, misspellings, fake requests, or unusual grammar. Models such as BERT (Bidirectional Encoder Representations from Transformers) analyse meaning and context, identifying deceptive or manipulative language [10]. NLP combined with ML enhances accuracy but requires strong pre-processing and large text datasets to train effectively. Natural Language Processing (NLP) is used in phishing detection in following manner.

1. **Text Data –** The system collects text from emails, messages, or websites.
2. **Feature Extraction and Pre-processing–** The text is cleaned and converted into features (like keywords, tone, or structure).
3. **Phishing Classifier –** A trained model analyzes the features to detect suspicious patterns.
4. **Phishing Detection –** The system classifies the message as *phishing* or *safe*.

D. Artificial Intelligence (AI) Techniques and Real-Time Detection

Deep Learning (DL) techniques have been applied to detect phishing websites, images, and visual features. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) automatically learn complex patterns from data [4]. CNNs are used to analyse website layouts and logos, identifying fake websites or scam screenshots, while RNNs can process email text sequences. These approaches improve detection accuracy but require large datasets and powerful hardware, making real-time use challenging.AI Techniques for Phishing Detection. Deep Learning Approaches process.

1. **Neural Network** – The core structure that processes data.
2. **Feature Learning** – The network automatically identifies important patterns or features.
3. **Model Training** – The system learns from large datasets to recognize phishing examples.
4. **Prediction** – The trained model predicts whether new data is *phishing* or *legitimate*.

E. Supervised and Unsupervised Learning in Phishing Detection

Recently, researchers have combined AI, ML, NLP, and Computer Vision (CV) to develop real-time functioning, sophisticated phishing detection systems. AI models combine word, image, and URL analysis to more precisely identify scams. For instance, CNNs and OpenCV techniques identify fake visuals, while NLP models detect suspicious text. Real-time systems also use secure communication protocols like WebSockets**,** JWT**,** and OAuth for instant detection in chat and messaging platforms [7]. These hybrid systems improve adaptability but still face challenges with false alarms and high processing times. AI detects phishing in real time. Data from emails or websites is collected, cleaned, and analysed using AI models like ML, NLP, or deep learning. The system then classifies content as phishing or safe and immediately alerts users to prevent attacks.

In phishing detection systems, two prominent machine learning models that are frequently used are supervised learning and unsupervised learning. Models are trained using labelled datasets in supervised learning, where each occurrence is linked to a known outcome, such as valid or phishing. The model gains knowledge about the relationships and patterns that exist between the labels and the input features (such as the email header, text content, domain age, and URL structure) during training. The class of new, unseen data can be accurately predicted by the model once it has been trained. The supervised learning techniques Support Vector Machines (SVM), Random Forests, Decision Trees, and Logistic Regression are frequently employed in phishing detection. When big, precisely labeled datasets are available, these models work well and can detect well-known phishing patterns with high accuracy [10].When they are exposed to new or changing phishing attempts that are not included in the training data, their performance might, however, suffer.

Labeled data is not necessary for unsupervised learning. Rather, it examines input data to find abnormalities, patterns, or hidden structures. Unsupervised learning is especially helpful in phishing detection when it comes to identifying zero-day attacks or previously undiscovered phishing strategies. Techniques like DBSCAN, Auto encoders, and K-Means Clustering can detect anomalous behaviors that depart from typical patterns or group comparable data points [11]. These algorithms improve flexibility and are able to identify new phishing techniques that evade conventional filters. They typically need greater processing power for real-time analysis, though, and can potentially result in false positives. It has been demonstrated that improving detection performance through the use of supervised and unsupervised techniques in a hybrid or ensemble framework allows systems to learn from both known and undiscovered phishing threats.

Modern AI-driven phishing protection systems rely heavily on this kind of integration since it facilitates adaptive learning, ongoing model refinement, and real-time threat identification. The progression of phishing detection from conventional rule-based systems to sophisticated AI-driven models is illustrated in Table 1. Blacklists and heuristics are the foundation of traditional techniques, which are straightforward and efficient against known threats but ineffective at spotting emerging or novel assaults. By using labelled datasets to train algorithms like Random Forests and Decision Trees, machine learning (ML) techniques overcome this constraint and increase accuracy and adaptability. By analyzing the language and contextual indicators in phishing messages, Natural Language Processing (NLP) approaches improve detection, particularly when paired with machine learning (ML). Lastly, AI-based real-time solutions give quick, multi-modal phishing detection by combining machine learning, natural language processing, and computer vision. This allows for quick, flexible responses to ever-changing cyberthreats.

Table 1: Comparative Analysis of different phishing detection approaches

Title	**Research Problem**	**Methodology**	**Key Contributions**
Traditional Phishing Detection Methods [5]	Existing rule-based and blacklist systems fail to detect zero-day and evolving phishing attacks.	Uses rule-based filtering, blacklists, heuristic analysis, and email content scanning.	Simple implementation and effective against known threats but lack adaptability.
Machine Learning (ML) Approaches [6] [8]	Traditional methods cannot adapt to new phishing patterns.	ML algorithms such as Decision Trees, Random Forest, SVM, and XGBoost trained on labeled phishing datasets.	Improved detection accuracy through learning patterns from data; ensemble models enhance robustness.
Natural Language Processing (NLP) Approaches [10]	Text-based phishing emails use deceptive language and tone to bypass filters.	Uses NLP techniques like tokenization, feature extraction, and BERT model to analyze text semantics.	Detects phishing through contextual and linguistic patterns; improved accuracy when combined with ML.

Deep Learning (DL) Approaches [4]	Manual feature extraction in ML models limits accuracy for complex data.	Employs CNNs and RNNs to automatically learn visual and textual phishing features.	Enhances detection accuracy for websites, logos, and text; reduces dependency on manual feature design.
AI Techniques and Real-Time Detection [7] [9]	Need for fast, adaptive, and multi-modal phishing detection systems.	Integrates AI, ML, NLP, and CV with real-time communication protocols (e.g., WebSockets, JWT).	Enables instant phishing detection with hybrid models analyzing text, image, and URL data together.

Challenges And Future Scope

The quick evolution of attack methods, which frequently makes current detection models outdated in a short amount of time, makes phishing detection a crucial cyber security concern. Static or rule-based systems find it challenging to keep up with the constant changes made by attackers to URLs, website content, and communication methods. Furthermore, there aren't many labeled datasets available, and many of them are unbalanced, with fewer phishing samples than authentic ones. This imbalance, especially for uncommon or unique phishing behaviors, results in biased model performance and decreased detection accuracy. The fact that deep learning models frequently function as "black boxes" is another significant worry. Despite their high accuracy, it is challenging to understand or interpret their internal decision-making processes. Particularly in delicate areas like banking and e-commerce, this lack of openness creates problems with responsibility and confidence. Furthermore, because high-speed examination of text, photos, and URLs necessitates a significant amount of processing power and efficient algorithms, real-time phishing detection has major technical and latency hurdles. Zero-day attacks, which use phishing techniques that haven't been seen before, are particularly challenging to detect since they don't have historical data to train on.

A number of upcoming enhancements are suggested in order to get around these restrictions. To ensure ongoing relevance, AI models that are flexible and self-learning should be created so they can automatically update themselves as new threats appear. Explainable AI (XAI) and hybrid techniques can improve interpretability and performance, allowing users to comprehend the reasoning behind a detection conclusion. By utilizing common knowledge across other domains, transfer learning and federated learning can further increase accuracy and privacy without disclosing private information. Furthermore, creating lightweight, optimized models can lower processing requirements, enabling effective real-time detection on mobile platforms and edge devices.

Lastly, encouraging cross-platform cooperation and data exchange among cyber security firms helps broaden threat coverage, resulting in more robust and complete phishing defensive systems.

The thorough review of previous research shows that phishing detection techniques have advanced significantly, moving from conventional rule-based and blacklist-based systems to advanced Artificial Intelligence (AI)-driven solutions.

Traditional methods are not flexible or resilient to zero-day phishing assaults, even when they are good at detecting known dangers. Models are now able to recognize intricate patterns and evaluate language clues in phishing content thanks to the advent of Machine Learning (ML) and Natural Language Processing (NLP) approaches. Additionally, Deep Learning techniques like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) have shown exceptional performance by automatically extracting high-level features from enormous datasets. Notwithstanding these developments, issues with data quality, model interpretability, and computational complexity still exist. As a result, current studies focus on hybrid and real-time detection systems that include adaptive feedback mechanisms, secure communication frameworks, and machine learning, natural language processing, and deep learning. Continuous learning, dynamic responsiveness, and increased detection accuracy are made possible by such systems. To sum up, AI-based real-time phishing detection frameworks offer a viable path toward developing resilient, scalable, and intelligent cyber security solutions that can handle the quickly changing phishing threat field.

References

Zeinab Shahbazi, Rezvan Jalali and Maryam Molaeevand, “AI-Based Phishing Detection and Student Cybersecurity Awareness in the Digital Age”,Research Environment of Computer Science

(RECS), Kristianstad University, 29188 Kristianstad, Sweden.

Obaloluwa Ogundairo, Peter Broklyn, "**AI-Driven Phishing Detection Systems",** Department of Computer and Systems Science, Stockholm University, 16440 Stockholm, Sweden; reja5738@student.su.se (R.J.); mamo7654@student.su.se (M.M.).

R B Aarthinivasini, Sridevi S, Subashini M, Roshini P, Shanjana P, "AI-POWERED PHISHING DETECTION SYSTEM IN WHATSAPP WEB CLONE" , International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 12 Issue: 05 | May 2025 www.irjet.net p-ISSN: 2395-0072.

Jumde, A., Hazarika, I., & Akre, V. (2023). *Challenges and opportunities in integrating rapidly changing technologies in business curriculum*. In *Proceedings of the 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 203–208). IEEE. https://doi.org/10.1109/ICCIKE58312.2023.10131683

Hazarika, I., Saoji, S., Bhandari, R. B., Jorvekar, G., Rao, P. H., & Porwal, T. (2025). *Mapping resilience pathways: A conceptual framework for portfolio risk management in microenterprise lending during economic shocks*. *Enterprise Development and Microfinance, 35*(1), 1–20. https://doi.org/10.3362/edm.v35i1.5

Akshatha, A. P., Chaithra, R., Madhura, S., Chandana, C. Sagar, & Hiriyanna, G. S. (2024). *Literature Review on Phishing Website Detection Using Deep Learning – IJSREM*.

Sharma, B. (2025). *Liability and virtual spaces: Examining legal responsibilities in Metaverse*. *National Journal of Cyber Security Law, 8*(2).

Jadhav, A., & Chandre, P. R. (2025). Survey and comparative analysis of phishing detection techniques: current trends, challenges, and future directions. *IAES International Journal of Artificial Intelligence (IJ-AI), 14*(2), 853-866.

Maddireddy, B. R., & Maddireddy, B. R. (2022). AI-Based Phishing Detection Techniques: A Comparative Analysis of Model Performance. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 13*(1), 652-674.

Sharma, B. (2023). *Impact of artificial intelligence on the legal industry: Advantages, challenges, and ethical implications*. *BioGecko, 12*(2), 3363–3374.

Popescul, D., & Radu, L. D. (2025). AI in phishing detection: a bibliometric review. *Frontiers in Artificial Intelligence, 8*, 1496580.

Patil, R. V., Gaidhani, V. A., Kashid, P. V., Hazarika, I., Mahadik, R. V., Poddar, G. M., & Patila, S. R. (2025). *Decentralized autonomous organizations as emerging economic entities in accounting and governance frameworks*. *International Journal of Accounting and Economics Studies, 12*(4), 166–177. https://doi.org/10.14419/1sy2j677

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2023). Phishing email detection using Natural Language Processing techniques: a literature survey. (University of Salford repository).

Yuan, Y., Apruzzese, G., & Conti, M. (2022). Multi-SpacePhish: Extending the Evasion-space of Adversarial Attacks against Phishing Website Detectors using Machine Learning. *arXiv preprint arXiv:2210.13660*.

Gupta, S., Sharma, R., & Patel, N. (2022). *Phishing Detection Using Natural Language Processing and Machine Learning*. IEEE Access.

Sospeter, B. K., & Odoyo, W. (2024). AI-Based Phishing Attack Detection and Prevention Using NLP. IC-ITECHS, 5(1).

Uzoaru, G. C., Odikwa, N. H., & Agbugba, A. A. (2024). Intelligent Phishing Website Detection Model Powered by Deep Learning Techniques. Asian Journal of Research in Computer Science, 17(1), 71-85.

Telecommunication Engineering Dept., Politeknik Negeri Sriwijaya. (2024). Implementation of Deep Learning for Detecting Phishing Attacks on Websites with CNN and LSTM. JuTif Jour