IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

A State-of-the-Art Review on Phishing Website Detection Techniques

Nirmal Prashant Vitthal[1], Prof. A. B. Jadhav[2], Prof. D. K. Gund[3]
[1] Student, [2] Guide, [3] Co-Guide
Department of Computer Engineering
Adsul Technical Campus, Chas, Ahilyanagar, India

Abstract: *Phishing websites are a major cybersecurity threat because they are designed to deceive users into sharing sensitive information, such as passwords, banking details, and personal data. These websites often imitate legitimate websites, making them difficult to identify through visual inspection alone. This paper presents a state-of-the-art review of techniques used to detect phishing websites. The review examines different approaches, including blacklist-based detection, heuristic-based methods, Machine Learning (ML), and Deep Learning (DL). It also discusses important website features, such as URL structure, domain information, webpage content, hyperlinks, and HTML characteristics, that help distinguish phishing websites from legitimate ones. Commonly used algorithms include Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN). The review further explores the applications, advantages, limitations, and evaluation metrics of these techniques. Existing research shows that AI-based detection methods have significant potential to improve phishing website identification and provide faster protection against online threats. However, challenges such as newly created phishing websites, changing attack strategies, imbalanced datasets, zero-day attacks, and model generalization remain important concerns. Future research should focus on real-time detection, explainable AI, updated and diverse datasets, lightweight detection systems, and hybrid approaches that combine multiple techniques to improve accuracy and reliability.*

Keywords: Phishing Website Detection, Cybersecurity, Machine Learning, Deep Learning, Artificial Intelligence, URL Analysis, Feature Extraction, Classification, Random Forest, XGBoost.

I. INTRODUCTION

Phishing websites have become a major cybersecurity concern as more people use the internet for online banking, shopping, education, social networking, and other digital services. These websites are created to look like genuine websites and trick users into revealing sensitive information, such as usernames, passwords, banking details, and personal data. Identifying phishing websites at an early stage is important to prevent financial losses, identity theft, and unauthorized access to user accounts. With the advancement of Artificial Intelligence (AI), automated techniques can analyze website characteristics and identify patterns that may indicate phishing activities.

Traditional phishing detection methods mainly rely on blacklists, manually defined rules, and website reputation checks. Although these methods are useful, they may fail to identify newly created phishing websites that have not yet been reported or added to a blacklist. Attackers frequently change website URLs, domain names, webpage designs, and other characteristics to avoid detection. AI-based phishing website detection provides an effective approach by analyzing website information and identifying suspicious patterns automatically. Machine Learning (ML) and Deep Learning (DL) techniques can process different website features and classify websites as legitimate or phishing.

Various techniques, including URL analysis, domain-based feature extraction, HTML content analysis, Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and Artificial Neural Networks (ANN), have been explored

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

for phishing website detection. These methods help identify differences between genuine and malicious websites by examining URL length, special characters, domain age, hyperlinks, webpage content, and other relevant features. Some advanced approaches also use Deep Learning to recognize complex patterns and improve detection performance. The main purpose of this paper is to present a state-of-the-art review of phishing website detection techniques, compare existing approaches, and examine their effectiveness, advantages, and limitations.

The review focuses on understanding how different detection methods identify phishing websites and how they can be improved to provide reliable protection against online threats. It also examines the role of feature selection, dataset quality, classification algorithms, and evaluation metrics in determining model performance. A comparison of existing techniques can help researchers understand the strengths and weaknesses of current approaches and identify suitable directions for future research.

However, existing phishing detection systems still face several challenges, including newly created phishing websites, frequently changing attack techniques, imbalanced datasets, misleading website designs, and difficulties in detecting previously unseen attacks. The quality and diversity of training data can significantly affect the performance of machine learning models.

Furthermore, some advanced models require considerable computational resources and may not clearly explain why a website has been classified as malicious. Therefore, future research should focus on real-time detection, explainable AI, updated datasets, hybrid detection methods, and improved generalization to emerging phishing attacks. These improvements can contribute to the development of more accurate, efficient, and reliable phishing website detection systems.

II. PROBLEM STATEMENT

Phishing websites have become a serious cybersecurity threat because they are designed to imitate legitimate websites and deceive users into sharing sensitive information, such as login credentials, banking details, and personal data. Identifying these websites at an early stage is important to prevent online fraud, financial loss, and identity theft. However, detecting phishing websites can be challenging because attackers frequently change website URLs, domain names, and webpage designs to make malicious websites appear genuine. Traditional detection methods, such as blacklists and rule-based techniques, may also fail to identify newly created phishing websites that have not yet been reported.

Artificial Intelligence (AI) provides a promising approach to identifying phishing websites by analyzing website features and detecting suspicious patterns. Machine Learning (ML) and Deep Learning (DL) techniques can examine URL structures, domain information, HTML content, and other website characteristics to classify websites as legitimate or phishing. However, challenges related to changing attack patterns, dataset quality, detection accuracy, and previously unseen phishing websites still need to be addressed. Therefore, this paper reviews existing phishing website detection techniques, compares their strengths and limitations, and identifies potential improvements for developing more accurate and reliable detection systems.

III. RESEARCH GAP

Existing Approach	Limitation	Proposed Improvement
Traditional blacklist-based detection	May fail to detect newly created phishing websites that are not yet listed.	Develop AI-based detection techniques that can identify previously unseen phishing websites.
URL-based phishing detection	Attackers can modify URLs to make malicious websites appear legitimate.	Combine URL analysis with domain and webpage features.
Single-feature detection methods	Relying on only one type of feature may lead to incorrect classification.	Use multiple features, including URL structure, domain information, HTML content, and hyperlinks.
Traditional Machine Learning methods	Performance depends on feature selection, dataset quality, and the chosen	Improve feature extraction and compare different ML algorithms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

	algorithm.	
Static website analysis	May struggle to identify phishing websites that frequently change their structure and content.	Explore dynamic analysis and real-time detection techniques.

IV. OBJECTIVES

1. To study how Artificial Intelligence (AI) can be used to detect phishing websites and distinguish them from legitimate websites.
2. To explore Machine Learning (ML) and Deep Learning (DL) techniques for identifying phishing websites based on their characteristics.
3. To understand how URL analysis, domain information, HTML content, and feature extraction can improve phishing website detection.
4. To review existing phishing website detection techniques and compare their performance using evaluation metrics such as accuracy, precision, recall, and F1-score.

V. PROPOSED SYSTEM

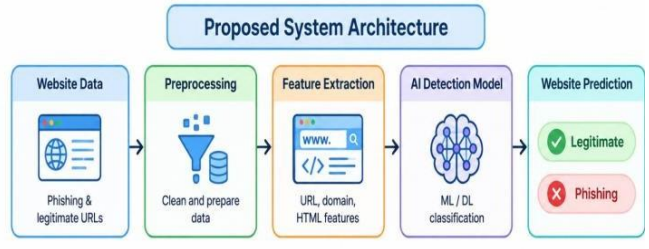
The proposed system focuses on reviewing different Artificial Intelligence (AI)-based techniques used to detect phishing websites and distinguish them from legitimate websites. In this approach, website information such as URL structure, domain details, HTML content, hyperlinks, and other relevant features is analyzed to identify suspicious activities. The collected website data is processed, and important features are selected to help detection models recognize patterns associated with phishing attacks.

Machine Learning (ML) and Deep Learning (DL) techniques, including Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and Artificial Neural Networks (ANN), are reviewed to understand their effectiveness in phishing website detection. The performance of these techniques is compared using evaluation metrics such as accuracy, precision, recall, and F1-score. The review also considers factors such as dataset quality, changing phishing techniques, computational requirements, and the ability of models to identify newly created phishing websites, as these factors can influence detection performance.

This approach may help researchers understand the strengths and limitations of existing phishing detection methods and identify opportunities for improvement. Advanced techniques such as hybrid detection, explainable AI, and real-time website analysis may further improve the reliability of phishing detection systems. These techniques could be useful in browser security tools, online banking protection, and other cybersecurity applications.

VI. PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture describes how different Artificial Intelligence (AI) techniques can be used to identify phishing websites and distinguish them from legitimate websites. The process begins with collecting website data from suitable phishing and legitimate website datasets. The collected data is then preprocessed to remove irrelevant information, handle missing values where necessary, and prepare the data for analysis.



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

VII. EXISTING SYSTEM

In recent years, different techniques have been developed to detect phishing websites and protect users from online threats. Traditional phishing detection systems mainly rely on blacklists, predefined rules, and website reputation checks to identify suspicious websites. Although these methods can detect many known phishing websites, they may not identify newly created websites that have not yet been reported. As phishing attacks continue to evolve, attackers frequently modify website URLs, domain names, and webpage designs to make malicious websites appear genuine. To overcome these limitations, researchers have explored Artificial Intelligence (AI)-based approaches using Machine Learning (ML) and Deep Learning (DL). These systems analyze website characteristics, such as URL structure, domain information, HTML content, and hyperlinks, to distinguish phishing websites from legitimate ones. Some existing approaches use algorithms such as Support Vector Machine (SVM), Decision Tree, Random Forest, and XGBoost, while advanced methods use neural networks to identify more complex patterns.

However, the performance of existing systems can be affected by limited or outdated datasets, imbalanced data, changing phishing techniques, and difficulties in detecting previously unseen attacks. Some models may also require significant computational resources or provide limited explanations for their predictions. Therefore, reviewing existing phishing website detection techniques is important to understand their effectiveness, identify their limitations, and explore possible improvements. This review focuses on comparing current approaches and examining how AI-based techniques can contribute to more accurate and reliable phishing website detection.

VIII. ADDITIONAL FEATURES OF THE SYSTEM

The proposed review focuses on Artificial Intelligence (AI)-based techniques for detecting phishing websites and distinguishing them from legitimate websites. Several additional features can help improve the reliability of phishing detection and the overall performance of AI-based models.

1. URL Structure Analysis: The system examines website URLs for suspicious characteristics, such as unusual domain names, excessive URL length, special characters, and misleading subdomains. These features can help identify potentially fraudulent websites.

2. Website Feature Extraction: Relevant information, including domain details, HTML content, hyperlinks, and other website characteristics, is extracted for further analysis. These features help AI models recognize patterns associated with phishing attacks.

3. Machine Learning-Based Detection: The system reviews machine learning algorithms such as Support Vector Machine (SVM), Decision Tree, Random Forest, and XGBoost to understand how effectively they distinguish phishing websites from legitimate ones.

4. Phishing Website Classification: The selected AI model classifies a website into one of two categories: phishing or legitimate. This classification can help users recognize potentially dangerous websites before sharing sensitive information.

Phishing Detection Accuracy = (Correctly Classified Websites / Total Websites Tested) × 100

For example, if a model correctly classifies 85 out of 100 tested websites, its classification accuracy is 85%. However, accuracy alone may not fully represent the effectiveness of a phishing detection model. Precision, recall, and F1-score are also important, particularly when evaluating how effectively a model identifies phishing websites without incorrectly flagging legitimate websites.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

IX. NEW FEATURES FOR BETTER DATA SHARING



Fig. 1. Proposed Data Sharing Improvement Process

The proposed review focuses on using Artificial Intelligence (AI) techniques to detect phishing websites and distinguish them from legitimate websites. It covers different stages of phishing detection, including website data collection, preprocessing, feature extraction, classification, and performance evaluation. These features help explain how existing AI-based techniques identify suspicious websites and where further improvements may be needed.

1. Website Data Collection:

The process begins with collecting data from available phishing and legitimate website datasets. The data may include URLs, domain information, HTML content, and other website characteristics. Using a variety of website samples is important because it helps researchers understand how well detection models perform on different types of websites.

2. Website Data Preprocessing:

Before analyzing the collected data, it is prepared for further processing. This may involve removing duplicate records, handling missing values, correcting inconsistencies, and converting website information into a suitable format. Proper preprocessing helps improve data quality and supports more reliable model evaluation.

3. Feature Extraction:

Before analyzing the collected data, it is prepared for further processing. This may involve removing duplicate records, handling missing values, correcting inconsistencies, and converting website information into a suitable format. Proper preprocessing helps improve data quality and supports more reliable model evaluation.

4. AI-Based Website Classification:

The extracted features are analyzed using Machine Learning (ML) or Deep Learning (DL) techniques. Common algorithms include Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and Artificial Neural Networks (ANN). These models learn patterns from labeled website data and classify websites as phishing or legitimate.

5. Classification Result and Feedback:

After analyzing a website, the model produces a predicted category. The prediction can be compared with the actual label in the dataset to identify correct classifications and errors. This feedback helps researchers understand the model's weaknesses and determine which types of phishing websites are difficult to detect.

6. Model Evaluation and Improvement:

The performance of detection techniques is evaluated using measures such as accuracy, precision, recall, and F1-score. These metrics help researchers compare different models and understand their effectiveness. The review also considers challenges such as outdated datasets, changing phishing strategies, false alarms, and difficulties in detecting previously unseen attacks. These findings can guide future improvements in phishing website detection.

IX. REVIEW METHODOLOGY

This study follows a review-based approach to understand the different techniques used for phishing website detection. The main focus is on existing methods, including traditional detection approaches, Machine Learning (ML), and Deep Learning (DL), and how effectively they distinguish phishing websites from legitimate ones. First, research papers related to phishing website detection, cybersecurity, URL analysis, feature extraction, and AI-based classification were identified from relevant research databases, such as IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

Greater attention was given to studies discussing recent developments, detection accuracy, and methods for identifying newly created phishing websites.

After collecting the research papers, their titles and abstracts were reviewed to identify studies relevant to the topic. Papers that focused on unrelated cybersecurity applications or did not discuss phishing website detection techniques were excluded from further consideration. Preference was given to studies that explained their detection methods, datasets, algorithms, and evaluation results.

The selected studies were then examined based on important factors, including the datasets used, website features analyzed, preprocessing methods, feature selection techniques, AI models, classification results, and evaluation metrics. Different approaches were compared using measures such as accuracy, precision, recall, and F1-score, wherever these results were available. The limitations of each approach were also considered, particularly challenges related to outdated datasets, changing phishing techniques, false positives, computational requirements, and difficulties in detecting previously unseen attacks.

X. PRIVACY AND SECURITY

Phishing website detection systems analyze information such as URLs, domain details, website content, and hyperlinks to identify potentially harmful websites. Although this information can help improve detection accuracy, it must be handled carefully to protect users' privacy and security. In some cases, website analysis may involve browsing information or other data that could reveal details about a user's online activities. To address these concerns, phishing detection systems should collect only the information required for analysis and avoid storing unnecessary personal data. Any collected information should be stored securely, and access should be restricted to authorized users. Appropriate security measures, such as data encryption, access control, and secure data handling practices, can help reduce the risk of unauthorized access or misuse.

XI. AI-BASED COUGH SOUND RECOGNITION ALGORITHM

Input: Website URL or Website Dataset and a Trained AI Model

Output: predicted Website Category (Phishing or Legitimate) and Classification Result.

1. The system collects website URLs and related data from available phishing and legitimate website datasets or suitable website sources.
2. The collected website data is preprocessed to remove duplicate records, handle missing values, and prepare the information for analysis.
3. Important website features are extracted, including URL structure, domain information, HTML content, hyperlinks, and other relevant characteristics.
4. The extracted features are provided to a trained Machine Learning (ML) or Deep Learning (DL) model for analysis.
5. The AI model examines the website features and compares them with patterns learned during the training process.
6. The model classifies the website into one of two categories: phishing or legitimate.
7. The predicted website category is generated as the model's output to indicate whether the website may be malicious or genuine.
8. The prediction is compared with the actual website label, when available, to determine whether the classification is correct.
9. The model's performance is evaluated using measures such as accuracy, precision, recall, and F1-score.
10. Classification errors and limitations are analyzed to understand where the detection model may need improvement.

XI. EXPECTED RESULTS

The proposed review is expected to provide a better understanding of how Artificial Intelligence (AI) techniques can be used to detect phishing websites and distinguish them from legitimate websites. Machine Learning (ML) and Deep

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

Learning (DL) approaches are expected to help identify suspicious patterns in URLs, domain information, HTML content, hyperlinks, and other website features.

The review is expected to show how data preprocessing, feature extraction, and suitable classification algorithms can contribute to effective phishing website detection. Comparing different AI techniques using evaluation measures such as accuracy, precision, recall, and F1-score will help researchers understand their strengths and limitations. This comparison may also help identify suitable approaches for detecting phishing websites more reliably.

The review is also expected to highlight common challenges, including outdated datasets, changing phishing strategies, imbalanced data, false positive predictions, and difficulties in identifying previously unseen attacks. Understanding these limitations can help researchers identify research gaps and explore future improvements, such as real-time detection, explainable AI, hybrid detection techniques, and more adaptable models.

XII. COMPARISON WITH EXISTING SYSTEM

Point	Existing System	Our Approach
Phishing Website Detection	Traditional systems mainly depend on blacklists, predefined rules, and website reputation.	Reviews AI-based techniques for identifying both known and previously unseen phishing websites.
Website Classification	Some methods may detect only known phishing websites or rely on limited classification rules.	Focuses on classifying websites into phishing and legitimate categories using machine learning and deep learning.
URL Analysis	Basic methods examine URLs using fixed rules and may miss cleverly designed or newly generated URLs.	Reviews URL-based features such as URL length, special characters, suspicious keywords, and domain structure.
Feature Extraction	Some systems analyze only URLs or a limited set of website characteristics.	Examines multiple features, including URLs, domain information, HTML content, and hyperlinks.
AI Techniques	A system may use a single machine learning algorithm or predefined detection rules.	Compares different approaches, including Decision Tree, Random Forest, SVM, XGBoost, and deep learning models.

XII. ADVANTAGES

• Early Phishing Detection:

AI-based phishing website detection can help identify suspicious websites before users share sensitive information, such as passwords, banking details, or personal data.

• Detection of Multiple Phishing Patterns:

Machine learning and deep learning models can analyze different website characteristics to classify websites as phishing or legitimate, depending on the quality of the training data.

• Faster Website Analysis:

Automated detection can analyze website URLs, domain information, HTML content, and hyperlinks more quickly than manual inspection of every website

• Improved Detection Performance:

Data preprocessing, feature extraction, and the use of suitable AI algorithms can help identify suspicious website patterns and may improve classification performance.

• Performance Evaluation:

Metrics such as accuracy, precision, recall, and F1-score help researchers compare different detection techniques and understand their strengths and limitations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

XIII. LIMITATIONS

The use of Artificial Intelligence for phishing website detection has several limitations that need to be considered. The accuracy of an AI model may depend on the quality, quantity, and freshness of the data used for training. Since phishing websites are frequently created and modified, a model trained on older data may fail to recognize new attack patterns. Some phishing websites are also designed to look very similar to genuine websites, making them difficult to identify.

Another limitation is the availability of reliable and diverse datasets. If the training data contains limited examples or an imbalance between phishing and legitimate websites, the model may not perform equally well in all situations. Different machine learning and deep learning algorithms can also produce different results depending on the selected features, dataset, and training process.

In addition, some detection methods analyze only website URLs, which may not provide enough information to identify every phishing attempt. Combining URL details, domain information, HTML content, and hyperlinks may improve detection, but it can increase processing requirements. Complex deep learning models may also require more computational resources and can be difficult to interpret. False positives are another concern, as legitimate websites may sometimes be incorrectly classified as phishing websites.

Therefore, further research is needed to improve the reliability and adaptability of AI-based phishing detection techniques. Testing models on diverse, updated, and previously unseen datasets can help evaluate their real-world performance. These techniques should be developed and evaluated carefully so that they can support cybersecurity systems and help users identify potentially harmful websites more effectively.

XIV. FUTURE SCOPE

In the future, AI-based phishing website detection can be improved by using larger, more diverse, and regularly updated datasets containing examples of both phishing and legitimate websites. This can help researchers develop models that identify new phishing techniques more reliably and perform better when analyzing websites that were not included in the training data.

The detection process can also be improved by exploring advanced machine learning and deep learning techniques. Combining different website features, such as URLs, domain information, HTML content, and hyperlinks, may help models identify suspicious patterns more effectively. Comparing different AI algorithms and using explainable AI techniques can also help researchers understand why a website is classified as phishing or legitimate.

Another possible development is the integration of AI-based phishing detection into web browsers, browser extensions, and mobile applications. Such tools could analyze suspicious websites and warn users before they enter sensitive information, such as passwords or banking details. Integration with email security systems could also help identify phishing links received through emails and messages.

XV. CONCLUSION

Phishing websites are a major cybersecurity concern because they can trick users into sharing sensitive information, such as login credentials, banking details, and personal data. Detecting these websites at an early stage is important for reducing the risk of online fraud and protecting users. Artificial Intelligence (AI) offers a promising approach to phishing website detection by identifying suspicious patterns and distinguishing phishing websites from legitimate ones. This review focuses on existing machine learning and deep learning techniques used for detecting phishing websites.

The review discusses important stages of phishing detection, including data collection, preprocessing, feature extraction, website classification, and model evaluation. It also examines different website features, such as URLs, domain information, HTML content, and hyperlinks. In addition, the review highlights challenges that can affect detection performance, including outdated datasets, changing phishing techniques, false positives, and difficulties in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

identifying previously unseen attacks. Comparing existing approaches helps researchers understand their strengths, limitations, and possible areas for improvement.

Overall, AI-based phishing website detection has the potential to strengthen cybersecurity by supporting faster and more effective identification of suspicious websites. However, its performance depends on the quality of the data, the selected algorithms, and the ability to adapt to new threats. Traditional security methods, such as blacklists and rule-based detection, can still provide useful protection when combined with AI-based approaches. Further research using diverse and updated datasets, explainable AI techniques, and real-time detection methods is needed to improve reliability and practical usefulness. In conclusion, AI-based phishing website detection can contribute to safer online activities when developed carefully and used alongside other cyber security measures.

XVI. ACKNOWLEDGMENT

I would like to express my sincere gratitude to my project guide for providing valuable guidance, suggestions, and support throughout this research work. The advice and feedback received helped me understand the topic of AI-based phishing website detection and improve my knowledge of how Artificial Intelligence can be used to identify and prevent phishing attacks.

I am also thankful to my department and college for providing the necessary facilities and a supportive environment to complete this research paper. I would like to thank my classmates and friends for their helpful discussions, encouragement, and support during the preparation of this work. I am especially grateful to my family for their continuous motivation and support throughout the research and writing process. Finally, I would like to acknowledge the researchers and authors whose published studies on Artificial Intelligence, machine learning, deep learning, and phishing website detection helped me understand existing techniques, identify research gaps, and prepare this review paper on “A State-of-the-Art Review on Phishing Website Detection Techniques.”

REFERENCES

[1] W. Li, S. Manickam, Y.-W. Chong, W. Leng, and P. Nanda, “A State-of-the-Art Review on Phishing Website Detection Techniques,” IEEE Access, vol. 12, pp. 187976–188012, 2024, doi: 10.1109/ACCESS.2024.3514972.

[2] KR. Zieni, L. Massari, and M. Calzarossa, “Phishing or Not Phishing? A Survey on the Detection of Phishing Websites,” IEEE Access, vol. 11, pp. 18499–18519, 2023, doi: 10.1109/ACCESS.2023.3247135.

[3] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, “A Comprehensive Survey of AI-Enabled Phishing Attacks Detection Techniques,” Telecommunication Systems, vol. 76, no. 1, pp. 139–154, 2021, doi: 10.1007/s11235-020-00733-2.

[4] A Systematic Literature Review on Phishing Website Detection Techniques,” Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 2, pp. 590–611, 2023, doi: 10.1016/j.jksuci.2023.01.004.

[5] Phishing Webpage Detection: Unveiling the Threat Landscape and Investigating Detection Techniques,” IEEE Communications Surveys & Tutorials, vol. 27, no. 2, pp. 974–1007, 2025, doi: 10.1109/COMST.2024.3441752.

[6] URL-Based Phishing Websites Detection via Machine Learning,” in Proc. International Conference on Data Analytics for Business and Industry (ICDABI), 2021, doi: 10.1109/ICDABI53623.2021.9655851.

[7] Phishing Websites Detection Using Machine Learning with URL Analysis,” in Proc. IEEE World Conference on Applied Intelligence and Computing (AIC), 2022, doi: 10.1109/AIC55036.2022.9848895.

[8] Survey on Detection and Prevention of Phishing Websites Using Machine Learning,” in Proc. International Conference on Advanced Computing and Innovative Technologies in Engineering (ICACITE), 2021, doi: 10.1109/ICACITE51222.2021.9404714.

[9] Phishing URL Detection Using Machine Learning and Deep Learning,” in Proc. IEEE World AI IoT Congress (AIIoT), 2024, doi: 10.1109/AIIoT61789.2024.10579005.

[10] Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review,” IEEE Access, vol. 13, pp. 192080–192104, 2025, doi: 10.1109/ACCESS.2025.3630334.