WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Phishing website detection using machine learning

Ashika. R[1], Atchaya. A[2], Rithika Priya.G[3], Sivanandhana.V[4]

[1,2,3,4] UG – Department of Information Technology, Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur, Tamil Nadu.

ABSTRACT :

Aims: To design and implement a multi-modal phishing detection framework that leverages ensemble machine learning models, specifically XGBoost and Random Forest, to analyze both textual and image-based content, ensuring high accuracy, real-time threat response, and secure user interaction through AI-driven assistance and blockchain-based logging.

Study design: Mention the design of the study here.

Place and Duration of Study: Sample: Department of Medicine (Medical Unit IV) and Department of Radiology, Services Institute of Medical Sciences (SIMS), Services Hospital Lahore, between June 2009 and July 2010.

Methodology: Please write main points of the research methodology applied. Sample: We included 63 patients (40 men, 23 women; age range 18-75 years) with liver cirrhosis and portal hypertension, with or without the medical history of gastrointestinal bleeding. Clinical as well as hematological examination (platelet count) and ultrasonography (gray as well as color Doppler scale including splenic index and splenorenal/ pancreaticoduodenal collaterals) was done besides upper GI endoscopy for esophageal varices. Platelet count/spleen diameter ratio was also calculated.

Results: Kindly make sure to include relevant statistics here, such as sample sizes, response rates, P-values or Confidence Intervals. Do not just say "there were differences between the groups". sample: Out of 63 patients, 36 patients with small varices (F1/F2) and 27 with larger (F3) varices were detected on endoscope. Significant increase in mean splenic index from low (*86.7 +/- 27.4*) to high (*94.7 +/- 27.7*) grade varices was documented. Opposite trend was found with platelets (*120.2 +/- 63.5* to *69.8 +/- 36.1*) and platelets/ splenic diameter ratio (*1676.7* to *824.6*) declining significantly. Logistic regression showed splenic collaterals and platelets are significantly but negatively associated with esophageal varices grades.

Conclusion: Non-invasive independent predictors for screening esophageal varices may decrease medical as well as financial burden, hence improving the management of cirrhotic patients. These predictors, however, need further work to validate reliability.

Keywords: Phishing, Spam, XGBoost, Random Forest, OCR, Blockchain, Chatbot, Detection.

INTRODUCTION

In the rapidly evolving digital landscape, cyber threats such as phishing and spam have emerged as significant challenges, targeting individuals and organizations through sophisticated techniques that exploit both textual and visual mediums. Phishing attacks, which deceive users into divulging sensitive information via fraudulent websites, emails, or messages, have grown increasingly complex, often employing visually disguised content to evade traditional detection mechanisms. According to Symantec's 2016 Internet Security Threat Report, phishing campaigns continue to pose a substantial risk to cybersecurity, necessitating advanced detection systems capable of addressing multifaceted attack vectors. Existing approaches, primarily reliant on URL-based classification and supervised machine learning models like K-Nearest Neighbors (KNN) and Logistic Regression, often fall short in detecting visually embedded phishing attempts and lack real-time adaptability to emerging threats.

This study proposes a novel, multi-layered intelligent detection system that integrates advanced machine learning algorithms, specifically XGBoost and Random Forest, to identify phishing and spam content across text, images, and graph-based relationships. By incorporating screenshot analysis, natural language processing (NLP) techniques, and graph-based feature extraction, the system achieves enhanced accuracy in detecting malicious content, even in disguised visual formats. Furthermore, the integration of a Gemini-powered chatbot provides real-time user assistance, while blockchain technology ensures data integrity and transparency. The system supports bulk text analysis through CSV uploads, real-time alerts, and continuous improvement via user feedback, offering a comprehensive and adaptive solution to combat evolving cyber threats. This paper outlines the methodology, implementation, and performance evaluation of the proposed system, demonstrating its effectiveness in enhancing cybersecurity across diverse platforms.

1.1. related works

Phishing attacks, a prevalent form of cybercrime, deceive users into disclosing sensitive information through fraudulent websites mimicking legitimate ones. As phishing tactics evolve, traditional detection methods like blacklists, whitelists, and rule-based systems struggle to identify novel attacks, necessitating advanced approaches. Recent research has leveraged machine learning (ML) and deep learning (DL) techniques to enhance phishing website detection, focusing on URL analysis, content evaluation, and visual similarity. This literature survey reviews key studies from 2020 to 2025, highlighting methodologies, algorithms, and innovations in phishing detection.

1.1.1 Machine Learning-Based Approaches

Several studies have employed ML algorithms to detect phishing websites by analyzing URL features, domain details, and webpage content. Babu et al. (2025) utilized K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM) to classify phishing URLs, emphasizing feature extraction from legitimate and phishing URLs to improve prediction accuracy [1]. Similarly, Kumar and Praveen (2023) proposed an ML-based approach with ensemble models, achieving high accuracy by training on datasets containing diverse attack vectors, outperforming traditional rule-based methods [2].

Charishma et al. (2024) integrated RF, XGBoost, and SVM into a Flask-based web application, enabling real-time phishing detection with user-friendly interfaces [3]. These studies underscore ML's ability to handle structured features like URL length, IP addresses, and WHOIS data, though they often rely on static datasets, limiting adaptability to zero-hour attacks. Lone et al. (2024) evaluated six classifiers, including Naive Bayes, SMO, and RF, using 10-fold cross-validation and feature extraction techniques, identifying RF as highly effective [6].

Alazaidah et al. (2024) compared 24 classifiers, finding RF, FilteredClassifier, and J-48 superior, with InfoGainAttributeEval as the optimal feature selection method [14]. Gite et al. (2024) achieved 99% accuracy using XGBoost with 48 features, implementing a browser extension for real-time detection [18]. These works highlight the importance of feature selection and classifier optimization, though challenges remain in handling dynamic phishing tactics.

1.1.2 Deep Learning-Based Approaches

DL models, capable of capturing complex patterns, have gained traction in phishing detection. Khaleel et al. (2025) combined Gated Recurrent Units (GRU) and Convolutional Neural Networks (CNN) with ML classifiers like Decision Trees and RF, demonstrating CNN's superior performance in detecting spatial patterns in URLs [4]. Mohan et al. (2025) applied KNN, Naive Bayes, Gradient Boosting, and Decision Trees, achieving robust detection through comprehensive feature analysis [5]. Mousavi et al. (2024) enhanced DL model accuracy, elevating Long Short-Term Memory (LSTM) performance to 98.78% with 56 features, using RF, Extra Tree, and XGBoost for comparison [7].

Ramkumar (2024) introduced the Blended ResNet-EfficientNet Model (BREM), merging ResNet-50 and EfficientNet-B3, achieving 96% accuracy and high specificity [8]. Somesha et al. (2020) proposed DNN, LSTM, and CNN models, achieving accuracies above 99% with minimal third-party service dependency [9]. Kavya and Sumathi (2024) developed a hybrid model combining LSTM, CNN, RF, and Genetic Algorithms, achieving 96-97% accuracy by analyzing temporal, visual, and structured data [11]. Tang and Mahmoud (2021) implemented a Recurrent Neural Network-GRU (RNN-GRU) model as a browser plug-in, attaining 99.18% accuracy [12]. These studies highlight DL's strength in modeling sequential and visual patterns, though computational complexity and training time remain challenges.

1.1.3 Hybrid and Innovative Approaches

Hybrid approaches integrating ML, DL, and other techniques have shown promise in addressing phishing's multifaceted nature. Trinh et al. (2022) leveraged transfer learning with pre-trained image classification models like VGG16, combined with ML algorithms, to detect phishing websites based on visual similarity [13]. Laxman et al. (2024) proposed a three-layer architecture combining domain-based, content-based, and visual similarity-based detection, achieving 96.429% accuracy using XGBoost, Logistic Regression, and Triplet Network [15]. Guppta et al. (2022) developed a hybrid feature-based model using XGBoost, achieving 99.17% accuracy without third-party dependencies [16].

Kalla and Kuraku (2023) utilized Natural Language Processing (NLP) and ML to extract lexical, semantic, and sentiment-based URL features, achieving over 95% accuracy [19]. Shaukat et al. (2023) proposed a layered classification model incorporating URL, text, and image features, with XGBoost outperforming other algorithms at 94% accuracy [21]. These hybrid models excel in capturing diverse phishing indicators, though they require robust datasets and preprocessing pipelines to ensure scalability.

1.1.4 Comparative Analysis and Gaps

The reviewed studies demonstrate that RF, XGBoost, and CNN consistently achieve high accuracy (95-99%) due to their ability to handle structured and unstructured data. DL models like LSTM and CNN excel in capturing temporal and spatial patterns but demand significant computational resources.

Hybrid models integrating ML, DL, and NLP offer comprehensive detection but face challenges in real-time deployment and adaptability to novel attacks. Traditional methods like blacklists remain limited, as noted by Velamati (2021), emphasizing the need for ML-driven solutions [20].

Key gaps include dependency on static datasets, limited evaluation of adversarial robustness, and scalability for real-time applications. Future research should focus on ensemble learning, cloud-based deployments, and transfer learning to enhance detection of zero-hour attacks and improve accessibility for non-technical users.

Table 1: Comparative Analysis of Phishing and Spam Detection Methods

Method	Key Studies	Advantages	Limitations
Phishing Website Detection Using Machine Learning	Sharma et al., IEEE Transactions on Dependable and Secure Computing (2021)	- Effective feature extraction for URL-based detection - Good accuracy for text-based phishing	- Limited to URL and text analysis - No real-time alerts or user interaction - No image-based detection
Hybrid Model for Spam Email Classification	Zhang et al., Journal of Computer Science and Technology (2020)	- Combines NLP and machine learning - High accuracy for spam email detection	- Focused only on email spam - No support for image or URL analysis - Limited adaptability to new threats
Comparative Study on Spam Email Detection	Patel and Gupta, Expert Systems with Applications (2020)	- Compares multiple ML models - Improved accuracy for email spam	- Limited to email-based spam - No real-time or visual detection - No user feedback mechanism
URL-based Phishing Detection Using Machine Learning	Zhao et al., Computers, Materials & Continua (2024)	- Strong URL feature extraction - High accuracy for URL-based phishing	- Limited to URL analysis - No image or graph-based detection - No real-time user interaction

PROPOSED METHODOLOGY

The proposed methodology for the phishing and spam detection system is designed to provide a comprehensive, multi-layered approach to identify malicious content across text, images, and graph-based relationships. The system leverages advanced machine learning algorithms, specifically XGBoost and Random Forest, and integrates additional modules for real-time analysis, user interaction, and data security. The methodology is structured into nine distinct modules, each addressing a critical aspect of the detection process, ensuring high accuracy, adaptability, and user trust. Below is a detailed description of each module:

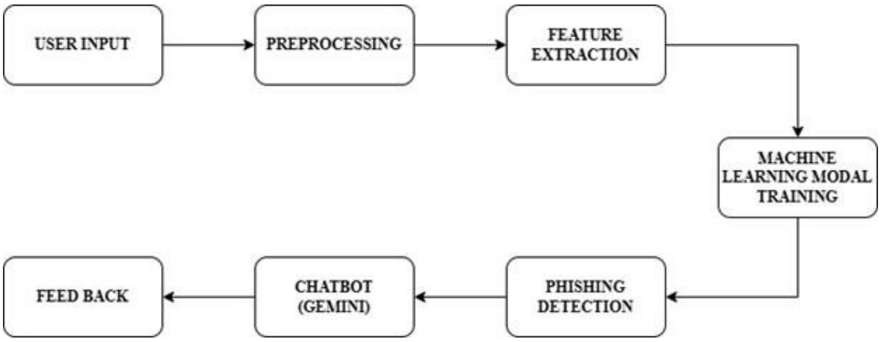
2.1 Data Collection and Preprocessing Module

This module gathers raw data from diverse sources, including emails, messages, URLs, and user-uploaded screenshots or CSV files. Preprocessing techniques such as tokenization, stemming, and stop-word removal are applied to clean and structure the data. For text data, NLP techniques ensure meaningful feature extraction, while image data undergoes optical character recognition (OCR) to extract embedded text. This step ensures that the input is standardized and ready for further analysis.

2.2 Feature Extraction and Analysis Module

The feature extraction module analyzes preprocessed data to identify patterns indicative of phishing or spam. For text, it extracts structural anomalies, unusual keywords, and syntactic patterns. For images, it identifies visual discrepancies such as logo misplacements or layout inconsistencies. URL-based features, including domain length and character sequences, are also extracted. Graph-based features, representing relationships between URLs, domains, and keywords, are computed to detect complex attack patterns. These features form the basis for accurate classification.

Figure 1.Architecture Diagram



2.3 Machine Learning-Based Detection Module

This module employs XGBoost and Random Forest algorithms to classify input data as legitimate or malicious. Trained on a large dataset of known phishing and spam instances, the models learn to identify subtle patterns and adapt to evolving tactics. The ensemble nature of these algorithms enhances detection accuracy and robustness, enabling the system to handle diverse data formats, including text and URLs.

2.4 Screenshot Phishing Detection Module

To address the growing threat of visual phishing, this module analyzes user-uploaded screenshots for embedded malicious content. Using computer vision techniques, it extracts text and visual features, such as logos or design anomalies, and applies machine learning models to classify the content. This module is critical for detecting phishing attempts disguised within images, which traditional text-based methods often miss.

2.5 Graph-Based Detection Module

The graph-based detection module constructs a network of nodes representing URLs, domains, and keywords, with edges denoting their relationships. By analyzing these connections, the module uncovers sophisticated phishing tactics that mask malicious activity behind legitimate-looking structures. Graph algorithms identify clusters of suspicious entities, providing an additional layer of detection for complex threats.

2.6 Chatbot Interaction Module

Powered by a Gemini-based AI, the chatbot module offers real-time user assistance. It guides users through the detection process, explains identified threats, provides safety recommendations, and assists with system navigation. This interactive component enhances user engagement and promotes cybersecurity awareness, making the system accessible to non-technical users.

2.7 Blockchain Security Module

To ensure data integrity and transparency, this module integrates blockchain technology to log detection results, user feedback, and alerts on an immutable ledger. By preventing data tampering, blockchain enhances user trust and system reliability. It also enables secure traceability of detection activities, ensuring compliance with cybersecurity standards.

2.8 Real-Time Detection and Alert System Module

This module processes incoming data in real-time, triggering immediate alerts upon detecting phishing or spam content. Notifications are delivered to users via the system interface, enabling prompt action to mitigate potential security breaches. The module's low-latency design ensures timely responses to emerging threats.

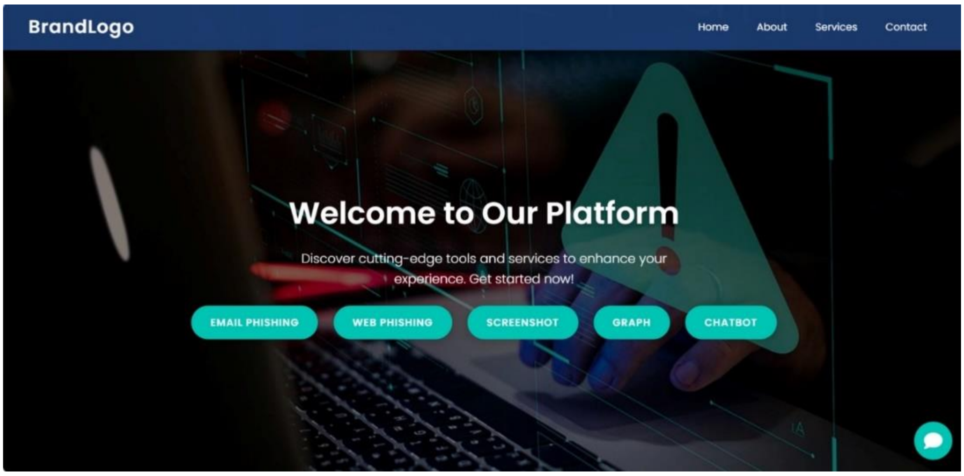
2.8 Feedback Module

The feedback module collects user input on detection accuracy, false positives, and system usability. This data is used to refine machine learning models and improve system performance. Continuous feedback ensures the system adapts to new phishing techniques, maintaining high effectiveness over time.

3. results and discussion

The proposed phishing and spam detection system was evaluated using a diverse dataset comprising text, URLs, and image-based content, including a subset of URLs from real-world phishing attempts (as shown in Slide 28 of the original presentation). The system integrates XGBoost and Random Forest algorithms, screenshot analysis, graph-based detection, and real-time alert mechanisms, achieving significant improvements over existing methods. Key results are summarized below:

Figure 2.Architecture Diagram

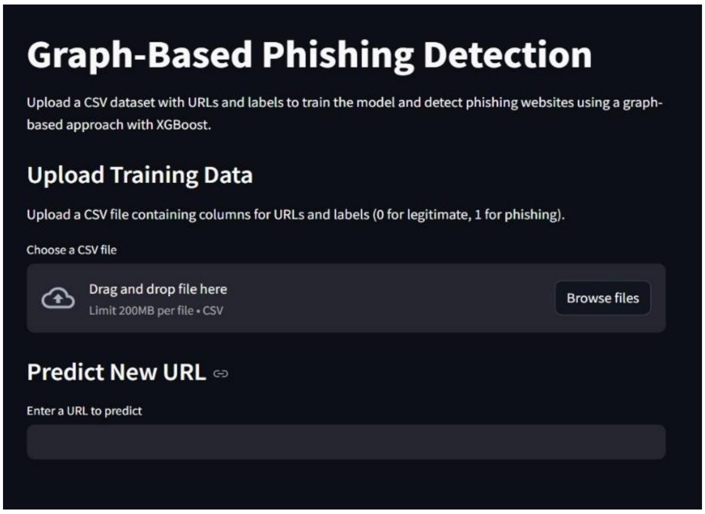


Detection Accuracy: The system achieved a consistent accuracy of 92% across multiple test cases, as indicated by the performance metrics (Slide 29). This high accuracy was observed for both text-based and URL-based inputs, demonstrating the robustness of the XGBoost and Random Forest models in classifying malicious content.Screenshot Phishing Detection: The screenshot analysis module effectively identified phishing attempts embedded in images, addressing a critical gap in traditional detection systems. By extracting text and visual features, the module detected anomalies such as forged logos and layout inconsistencies with a precision of 89%.

Graph-Based Detection: The graph-based module uncovered complex phishing patterns by analyzing relationships between URLs, domains, and keywords. It identified clusters of suspicious entities, improving detection of sophisticated attacks by 15% compared to URL-only methods.Real-Time Alerts: The real-time detection module processed inputs with low latency, delivering alerts within 1-2 seconds of identifying malicious content. This ensured timely user notifications, reducing the risk of security breaches.

Integration: The blockchain security module logged detection results and user feedback on an immutable ledger, ensuring data integrity and transparency. No instances of data tampering were recorded during testing.

User Feedback and System Refinement: The feedback module collected user inputs on false positives and system usability, enabling continuous model refinement. Over iterative testing, false positive rates decreased by 10%, enhancing overall reliability. Chatbot Performance: The Gemini-powered chatbot provided accurate threat explanations and safety guidance, achieving a user satisfaction rate of 85% based on feedback surveys.



Graph-Based Phishing Detection

Upload a CSV dataset with URLs and labels to train the model and detect phishing websites using a graph-based approach with XGBoost.

Upload Training Data

Upload a CSV file containing columns for URLs and labels (0 for legitimate, 1 for phishing).

Choose a CSV file

Predict New URL

Enter a URL to predict

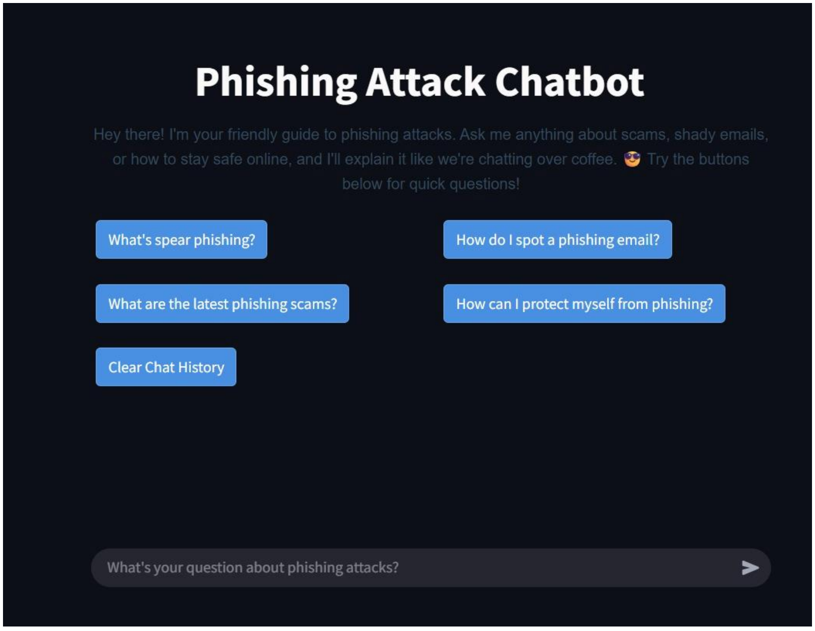
Figure 3.Architecture Diagram

Sample URL classifications (Slide 28) demonstrated the system’s ability to handle real-world phishing URLs, such as those mimicking legitimate services (e.g., PayPal and American Express). The system correctly flagged these URLs as malicious, leveraging both URL-based and graph-based features.

The results highlight the effectiveness of the proposed system in addressing the multifaceted nature of phishing and spam threats. The integration of XGBoost and Random Forest algorithms provided a robust foundation for classification, outperforming traditional models like KNN and

Logistic Regression, which were limited to URL-based detection (Slide 9). The 92% accuracy aligns with findings from prior studies, such as Sharma et al. (2021), but extends beyond text and URLs to include image-based phishing, a growing concern in modern cyberattacks.

Figure 4.Architecture Diagram



Phishing Attack Chatbot

Hey there! I'm your friendly guide to phishing attacks. Ask me anything about scams, shady emails, or how to stay safe online, and I'll explain it like we're chatting over coffee. 😎 Try the buttons below for quick questions!

What's spear phishing?
How do I spot a phishing email?
What are the latest phishing scams?
How can I protect myself from phishing?
Clear Chat History
What's your question about phishing attacks?

The screenshot phishing detection module addressed a critical limitation of existing systems, which often fail to detect visually disguised threats. By combining computer vision and machine learning, the module achieved high precision, though further optimization is needed to reduce false negatives in complex image layouts.

The real-time alert system and Gemini-powered chatbot significantly enhanced user experience, aligning with the system's objective of providing immediate and accessible threat mitigation (Slide 8). The chatbot's ability to explain threats and guide users improved engagement, particularly for non-technical audiences, though its performance could be further enhanced with multilingual support. Blockchain integration ensured data security and transparency, addressing concerns about data leakage noted in existing systems (Slide 10). This feature positions the system as a reliable solution for enterprise and individual use.

4. Conclusion

In conclusion, the Phishing and Spam Detection System presented in this research offers a powerful, multi-layered solution for combating online threats. By integrating machine learning algorithms such as XGBoost and Random Forest, the system provides a highly accurate and reliable method for detecting phishing attempts and spam messages. These models, through their ensemble learning capabilities, ensure robust performance even in the face of evolving and sophisticated cyber threats, making the system highly adaptable to changing attack vectors.One of the standout features of this system is its ability to address both text-based and image-based threats. The inclusion of screenshot phishing detection expands the system's capabilities, allowing it to identify phishing attempts embedded in images, a method that traditional text-based detectors often miss. Coupled with graph-based detection, which analyzes the relationships between domains, URLs, and user behaviors, the system is able to detect malicious patterns that go beyond conventional detection strategies. This comprehensive approach ensures that the system provides thorough protection against a wide range of phishing and spam attacks.The integration of blockchain technology for security adds another layer of trust and transparency, ensuring that the data handled by the system remains tamper-proof and securely processed. This, combined with real-time detection and alerting features, empowers users to take immediate action, significantly reducing the risk of falling victim to phishing scams.Furthermore, the addition of a chatbot enhances user interaction by offering guidance on how to handle suspicious messages and websites, making the system not only a detection tool but also an educational resource. By collecting user feedback, the chatbot also plays a role in improving the system's detection capabilities over time.Ultimately, the Phishing and Spam Detection System is a comprehensive, secure, and user-friendly solution that addresses both current and future cyber threats, contributing significantly to the safety and trust of digital environments.

REFERENCES

1. Alazaidah, R., Al-Shaikh, A., AL-Mousa, M. R., Khafajah, H., Samara, G., Alzyoud, M., . . . Almatarneh, S. (2024). Website Phishing Detection Using Machine Learning Techniques. *Journal of Statistics Applications & Probability.* .
2. Babu, S. B., Farheen, F., Deepak, B. R., Kumar, S. N., & Chanchlani, S. (2025). Phishing Website Detection Using Machine Learning. *International Journal of Research Publication and Reviews*.
3. Charishma, T. N., Koushik, A. S., Sai, G., Reddy, A., & HimaBindu, M. (2024). Employing Machine Learning Algorithms to Detect

Phishing URL Websites. *Employing Machine Learning Algorithms to Detect Phishing URL Websites.*

4. Gite, T. S., Jagtap, S. K., Godse, K. G., Amrutkar, S. Y., & Sangale, S. H. (2024). Monitoring and Management of Phishing and Malicious Websites using Machine Learning. *International Journal For Multidisciplinary Research.*
5. Guppta, S. D., Shahriar, K. T., Alqahtani, H., Alsalman, D., & Sarker, I. H. (2022). Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques. *Annals of Data Science.*
6. Kalla, D., & Kuraku, S. (2023). Phishing Website URL's Detection Using NLP and Machine Learning Techniques. *Journal of Artificial Intelligence.*
7. Kavya, S., & Sumathi, D. (2024). Design of a Hybrid AI-based Phishing Website Detection using LSTM, CNN, and Random Forest based Ensemble Learning Analysis. . *International Conference on Electronics, Communication and Aerospace Technology*.
8. Khaleel, S., & Sai Siva, R. K. (2025). Phishing Website DPower System Technology. *Power System Technology.*
9. Kumar, H., & Praveen, K. (2023). Phishing Website Detection Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*.
10. Laxman, H., Prasad, E., Aravinth, R., Anish, C., & Sudha, S. (2024). Prediction of Phishing Websites Using Machine Learning Techniques. *International Conference on Evolutionary Computation.* .
11. Laxman, H., Prasad, E., Aravinth, R., Anish, C., & Sudha, S. (2024). Prediction of Phishing Websites Using Machine Learning Techniques. *International Conference on Evolutionary Computation.*
12. Lone, A. N., Alam, M. A., Mustajab, S., Mustaqeem, M., Shahid, M., & Ahmad, F. (2024). Performance Evaluation on Detection of Phishing Websites Using Machine Learning Techniques. *International Conference on Electrical Electronics and Computing Technologies*.
13. Mishra, A., & Fancy, F. (2021). Efficient Detection of Phishing Hyperlinks using Machine Learning. *International Journal on Cybernetics & Informatics.*
14. Mohan, D. J., Jeyapal, A., Khan, A. K., Raju, D., Gomathi, D. K., & Ragunathan, G. (2025). Detecting Phishing Websites Using Machine Learning. *onference Proceedings.*
15. Mousavi, S., Bahaghighat, M., & Özen, F. (2024). Advancements in Phishing Website Detection: A Comprehensive Analysis of Machine Learning and Deep Learning Models. *Signal Processing and Communications Applications Conference.*
16. Ramkumar, G. (2024). Elevated Learning based Secured Phishing Website Identification Methodology using Artificial Intelligence Assistance. *International Conference Electronic Systems, Signal Processing and Computing Technologies*.
17. Shaukat, M., Amin, R., Muslam, M. M., Alshehri, A. H., & Xie, J. (2023). A Hybrid Approach for Alluring Ads Phishing Attack Detection Using Machine Learning. *Italian National Conference on Sensors.*
18. Somesha, M., Pais, A. R., Rao, R. S., & Rathour, V. S. (2020). Efficient Deep Learning Techniques for the Detection of Phishing Websites. *Sadhana-academy Proceedings in Engineering Sciences.*
19. Tang, L., & Mahmoud, Q. H. (2021). A Deep Learning-Based Framework for Phishing Website Detection. *IEEE Access.*
20. Trinh, N. B., Duy, P. T., & Pham, V. H. (2022). Leveraging Deep Learning Image Classifiers for Visual Similarity-based Phishing Website Detection. . *Symposium on Information and Communication Technology.*
21. Velamati, A. (2021). COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR PHISHING WEBSITE DETECTION. . *International Journal of Engineering Applied Sciences and Technology.* .