Impact Factor (QJIF): 8.38 E-ISSN: 2583-7206, P-ISSN: 3049-2432

International Journal of All Subject Research

Received: 17/February/2025 **IJASR:** 2026; 5(2):73-78 **Accepted:** 02/April/2026

Hybrid Phishing Detection Using Machine Learning

[1]Dr. K Karuppasamy, [*2]Mohammed Sami A, [3]Kali Dhanush S, [4]Nimesh Y and [5]Poovarasu T

[1]Head of the Department, Department Computer Science & Engineering, RVS College of Engineering & Technology, Coimbatore, Tamil Nadu, India.

[*2, 3, 4, 5]Research Scholar, Department Computer Science & Engineering, RVS College of Engineering & Technology, Coimbatore, Tamil Nadu, India.

Abstract

The rapid expansion of digital communication technologies and online service platforms has significantly increased exposure to phishing attacks, a form of cybercrime in which malicious actors use deceptive URLs to manipulate users into revealing confidential information such as login credentials, financial data, and personal records. Traditional phishing detection methods primarily rely on blacklist databases, signature matching, and static rule-based filtering. While these approaches are effective against previously identified threats, they are inadequate for detecting newly generated phishing URLs, polymorphic attacks, and zero-day exploits that continuously evolve to bypass conventional security mechanisms.

To address these challenges, this research proposes a Quantum-Enhanced Intelligent Phishing URL Detection system that integrates machine learning-based classification with quantum-inspired optimization techniques to improve detection accuracy and system adaptability. The proposed framework extracts and analyzes multiple structural, lexical, and behavioral features of URLs, including domain characteristics, URL length patterns, token distribution, and anomaly indicators. These features are processed using an intelligent learning model capable of distinguishing between legitimate and malicious links in real time.

A quantum-inspired optimization approach is incorporated to enhance feature selection and model parameter tuning, enabling improved convergence efficiency and predictive performance. This hybrid methodology reduces false positive rates while maintaining high detection sensitivity, making the system more robust against previously unseen phishing strategies.

Experimental evaluation demonstrates that the proposed framework achieves superior classification accuracy, improved computational efficiency, and enhanced scalability compared to traditional detection approaches. By providing adaptive learning capability and real-time threat identification, the system contributes to the development of secure and resilient cybersecurity infrastructures. The proposed solution supports scalable deployment across modern digital environments and offers an effective defense mechanism against emerging web-based threats.

Keywords: Phishing Detection, Machine Learning, Quantum-Inspired Optimization, Cyber Security, URL Classification.

1. Introduction

The widespread adoption of digital technologies and online services has transformed communication, commerce, and information exchange. However, this rapid digitalization has also created new opportunities for cybercriminal activities, among which phishing remains one of the most prevalent and damaging threats. Phishing attacks typically exploit deceptive web links and fraudulent websites to mislead users into disclosing sensitive information such as login credentials, financial data, and personal details. As internet usage continues to expand across individuals, organizations, and critical infrastructures, the need for reliable and intelligent phishing detection mechanisms has become increasingly important.

Traditional approaches for phishing detection largely rely on blacklist databases, heuristic rules, and signature-based filtering methods. While these techniques are effective for identifying previously known threats, they face significant limitations when dealing with newly generated phishing URLs and rapidly evolving attack strategies. Cyber attackers frequently modify URL structures, employ domain obfuscation techniques, and create short-lived malicious websites to bypass conventional security systems. As a result, static detection methods struggle to provide timely and accurate protection against emerging threats.

Machine learning has emerged as a promising solution for addressing these challenges by enabling systems to learn patterns and identify anomalies from data without relying

***Corresponding Author:** Mohammed Sami A

solely on predefined rules. By analyzing structural and behavioral characteristics of URLs, machine learning models can identify hidden relationships that distinguish malicious links from legitimate ones. These models offer improved adaptability, scalability, and predictive capability, making them well-suited for real-time phishing detection in dynamic online environments.

This research focuses on the development of an Intelligent Phishing URL Detection system using machine learning techniques to enhance cybersecurity performance. The proposed approach analyzes multiple features of URLs to accurately classify them as legitimate or phishing attempts. By leveraging data-driven learning and automated pattern recognition, the system aims to improve detection accuracy, reduce false positives, and provide robust protection against evolving phishing strategies.

The remainder of this paper presents the system design, methodology, experimental evaluation, and performance analysis of the proposed detection framework, demonstrating its effectiveness in strengthening web security and supporting safer digital interactions

2. Literature Survey

Phishing detection has become a critical research area in cybersecurity due to the rapid increase in malicious web activities targeting online users. Several researchers have explored machine learning and intelligent computing techniques to improve the identification of phishing URLs beyond traditional blacklist and rule-based approaches.

Early phishing detection systems primarily relied on blacklist databases and heuristic filtering methods. Although these techniques were effective in identifying previously reported malicious websites, they lacked the ability to detect newly generated phishing URLs and zero-day attacks. This limitation motivated researchers to investigate data-driven and adaptive detection methods capable of learning patterns from URL characteristics.

Verma and Gupta (2020) explored machine learning-based phishing detection using ensemble classification techniques. Their study demonstrated that combining multiple classifiers improves detection accuracy and enhances robustness compared to individual models. The ensemble approach was shown to reduce misclassification and improve reliability in identifying suspicious URLs.

Sharma *et al*. (2021) proposed a hybrid phishing detection framework that integrates natural language processing techniques with feature selection methods to support real-time URL analysis. Their work emphasized the importance of extracting meaningful lexical and structural features from URLs to improve classification performance. The study reported improved detection efficiency in dynamic web environments.

Li and Wang (2022) investigated deep neural network-based models for large-scale phishing URL classification. Their approach utilized advanced representation learning to identify hidden patterns within URL structures. The experimental results indicated strong scalability and high accuracy when trained on large datasets, demonstrating the effectiveness of deep learning in cybersecurity applications.

Kumar and Singh (2023) introduced an artificial intelligence-driven phishing detection framework that incorporates enhanced feature engineering and adaptive learning strategies. Their system focused on minimizing false positive rates while maintaining high detection sensitivity. The study highlighted the importance of intelligent feature optimization in improving overall system performance.

Despite these advancements, existing approaches still face challenges related to computational efficiency, optimal feature selection, and adaptability to rapidly evolving phishing strategies. Many models require extensive training data or exhibit limitations in real-time deployment scenarios. These gaps highlight the need for improved optimization techniques that enhance model performance while maintaining scalability.

To address these limitations, the present work proposes an Intelligent Phishing URL Detection system that integrates machine learning classification with quantum-inspired optimization. By combining adaptive learning with optimized feature selection, the proposed approach aims to achieve higher detection accuracy, reduced false positives, and improved real-time performance in identifying malicious URLs.

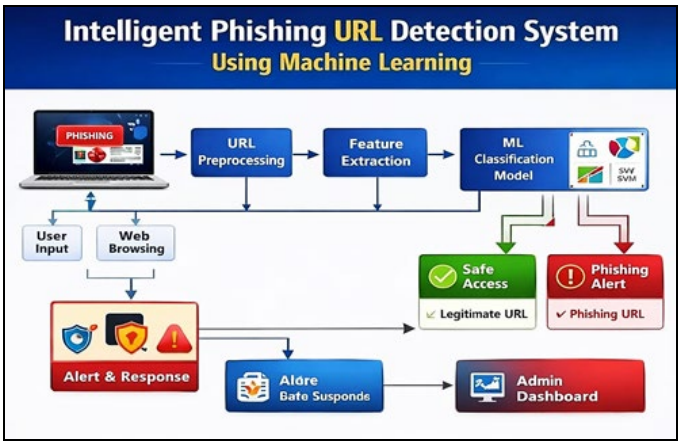
3. Proposed methodology

i). **URL Data Collection and Preprocessing Module:** This module is responsible for gathering URL data from various sources such as user input, web browsers, email links, and phishing datasets. The collected URLs undergo preprocessing to ensure data consistency and reliability for analysis. Preprocessing operations include normalization of URL formats, removal of redundant characters, tokenization, and cleaning of noise elements. The output of this module is a structured and standardized dataset prepared for feature extraction and classification.

ii). **Phishing Detection Model Module:** This module serves as the analytical core of the system. It extracts structural and lexical features from the preprocessed URLs, such as URL length, domain characteristics, presence of HTTPS, special character frequency, and suspicious keyword patterns. These features are converted into numerical representations and processed using supervised machine learning algorithms. The trained model classifies URLs as phishing or legitimate based on learned patterns from labeled training data. Optimization techniques are applied to improve feature selection and enhance model accuracy.

iii). **Alert and Decision System Module:** The Alert and Decision System generates real-time responses based on the prediction produced by the detection model. When a URL is identified as malicious, the system immediately issues a warning notification and blocks access to the suspicious website. Legitimate URLs are allowed to proceed without interruption. This module ensures proactive protection by minimizing user exposure to phishing threats and providing clear decision outcomes.

iv). **Admin Dashboard Module:** The Admin Dashboard provides centralized monitoring and management of the phishing detection system. It enables administrators to track detected malicious URLs, evaluate system performance metrics, and generate analytical reports. The dashboard supports system supervision, data management, and performance evaluation, ensuring reliable operation and continuous improvement of the detection framework.



A) System Architecture Explanation

The diagram illustrates the workflow of the Intelligent Phishing URL Detection System using Machine Learning. It shows how URLs are collected, analyzed, classified, and monitored to protect users from phishing attacks in real time.

i). Input Layer

The process begins with two sources of URL data:

- **User Input** — URLs manually entered by users.
- **Web Browsing Activity** — Links accessed during normal browsing.

These inputs provide raw URLs that may be legitimate or malicious.

ii). URL Preprocessing

The collected URLs are forwarded to the preprocessing stage, where the system prepares them for analysis. This stage includes:

- Removing unnecessary characters
- Normalizing URL structure
- Cleaning noise and formatting inconsistencies

Preprocessing ensures that the data is consistent and suitable for feature analysis.

iii). Feature Extraction

After preprocessing, the system extracts meaningful characteristics from each URL. These features help identify suspicious patterns commonly found in phishing links. Typical extracted attributes include:

- URL length and structure
- Domain-related information
- Presence of HTTPS
- Special character frequency
- Suspicious keywords or tokens

The extracted features are converted into numerical values for machine learning processing.

iv). Machine Learning Classification Model

The feature vectors are provided to the machine learning classification module. This model is trained using labeled datasets to distinguish between phishing and legitimate URLs. Based on learned patterns, the model predicts whether a given URL is safe or malicious.

v). Decision Output

Depending on the model's prediction, the system produces one of two outcomes:

- **Safe Access** — The URL is legitimate and access is allowed.
- **Phishing Alert** — The URL is identified as malicious and blocked.

This step ensures immediate protection for users.

vi). Alert and Response System

When a phishing URL is detected, the alert module generates warnings and prevents the user from accessing the harmful website. This provides proactive security against online threats.

vii). Data Storage and Monitoring

All detection results and system activities are stored for analysis. The stored data is used for performance evaluation, tracking phishing attempts, and improving system accuracy.

viii). Admin Dashboard

The Admin Dashboard provides centralized monitoring and control of the system. It allows administrators to:

- View detected phishing URLs
- Monitor system performance
- Analyze detection statistics
- Manage system data

This module supports continuous monitoring and system management.

B) URL Input Page Explanation

The figure represents the user interface of the Intelligent Phishing URL Detection System. This page allows users to manually submit a web link for security analysis and phishing detection.

i). **User Interaction Interface:** The page provides a simple and user-friendly interface where users can enter a URL they want to verify. This input may come from suspicious emails, messages, or unknown websites encountered during browsing.

ii). **URL Entry Field:** A text input box is provided to enter the complete web address. The system accepts the URL in standard format and prepares it for analysis. This step serves as the starting point of the detection process.

iii). **URL Verification Process:** After entering the link, the

user clicks the "Check URL" button. The system then forwards the entered URL to the backend processing module where preprocessing, feature extraction, and machine learning classification are performed.

iv). **Security Purpose:** This page acts as the access point for real-time phishing detection. It enables users to verify website authenticity before visiting the link, thereby preventing exposure to malicious web pages and protecting sensitive information.

v). **System Integration:** The URL Input Page is connected to the detection model, alert mechanism, and monitoring system. Based on analysis results, the system either allows safe access or generates a phishing warning.

C) Phishing Report and Monitoring Interface Explanation

The figure shows the reporting and monitoring interface of the Intelligent Phishing URL Detection System. It presents the results of phishing detection and provides administrative control for system supervision and response management.

i). Phishing Detection Records

The left panel displays a list of URLs that have been identified as suspicious or malicious by the detection model. Each record includes:

- Unique ID for identification
- Detected phishing URL
- Detection date and time
- Option to view detailed analysis

This section helps track phishing attempts detected by the system over time.

ii). Phishing Analysis and Reasons

The right panel provides a detailed phishing report that explains why a URL was classified as malicious. The report includes:

- URL identified as phishing
- Detection date
- Reason for classification
- Action options

Common detection reasons shown include absence of HTTPS, presence of suspicious keywords, and invalid security certificates. This improves transparency and supports security analysis.

iii). Action and Response Control

The system allows administrators to take immediate actions on detected threats. Available actions include:

- Blocking malicious URLs
- Viewing detailed reports
- Managing detection records

This ensures proactive threat mitigation and system control.

iv). Report Export Function

Both panels include an export option that allows administrators to download phishing detection reports for documentation, auditing, or further analysis.

v). Role in System Workflow

This interface represents the Admin Dashboard and Monitoring Module of the system. It supports performance tracking, threat analysis, and decision-making, ensuring continuous system supervision and improved cybersecurity management.

D) Detection Result Page Explanation

The figure represents the Detection Result Page of the Intelligent Phishing URL Detection System. This page displays the final classification outcome after the machine

learning model analyzes the submitted URL.

i). **Detection Status Display:** The system prominently shows the classification result. In this case, the URL is identified as a phishing link. The alert message informs the user that the website is potentially dangerous and should not be trusted.

ii). **Analyzed URL Information:** The page displays the submitted URL along with the analysis summary. This provides transparency by showing the exact link evaluated by the system.

iii). **Reason for Classification:** The system explains why the URL was classified as malicious. The reasons may include:

- Suspicious or misleading domain structure
- Absence of secure HTTPS protocol
- Excessive redirections or abnormal parameters

Providing reasons improves user awareness and helps understand phishing characteristics.

iv). **Security Warning Message:** A warning message informs the user that the detected website may attempt to steal personal or sensitive information. This supports proactive cybersecurity awareness.

v). **User Action Options**

The interface provides two response options:

- Block Website — Prevents access to the malicious link
- Ignore Warning — Allows the user to proceed at their own risk

This ensures user control while maintaining system protection.

vi). **Role in System Workflow**

This page represents the output stage of the phishing detection pipeline. It connects the classification model with the alert system and user response mechanism, ensuring real-time decision support.

E) Admin Dashboard Explanation

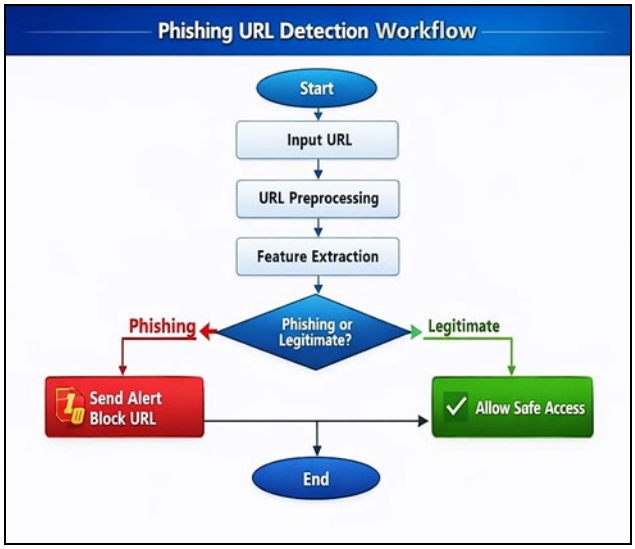
The figure presents the Admin Dashboard of the Intelligent Phishing URL Detection System. This interface provides centralized monitoring, analysis, and management of phishing detection activities within the system.

The dashboard displays key performance indicators that summarize system operation. It shows the total number of URLs analyzed, the number of legitimate URLs identified, the number of detected phishing URLs, and the overall detection accuracy of the machine learning model. These indicators help administrators quickly evaluate system effectiveness and detection performance.

The interface also includes navigation sections such as URL logs, phishing reports, user management, and model performance monitoring. The URL logs section provides a structured record of analyzed web links along with their classification status and detection timestamps. Each entry allows administrators to view detailed analysis results for further inspection.

The phishing reports section enables administrators to review identified threats and understand the reasons behind classification decisions. The dashboard supports system supervision by allowing administrators to monitor detection trends, manage stored data, and maintain operational control.

Overall, the Admin Dashboard functions as the control and monitoring center of the phishing detection system, ensuring transparency, performance evaluation, and effective cybersecurity management.



The diagram illustrates the operational workflow of the Intelligent Phishing URL Detection System. The process begins when a user submits a URL for verification. The system first performs preprocessing, where the input link is cleaned, normalized, and structured for analysis. After preprocessing, important structural and lexical features are extracted from the URL. These features capture patterns that commonly indicate phishing behavior, such as suspicious domain structure, abnormal length, and absence of secure protocols.

The extracted features are then evaluated by a trained machine learning classification model that determines whether the URL is legitimate or malicious. If the URL is identified as phishing, the system generates a warning alert and blocks access to the harmful website to protect the user. If the URL is classified as legitimate, safe access is allowed without interruption. This workflow enables accurate detection, real-time decision-making, and proactive protection against phishing threats.

4. Result and Discussion

The proposed Intelligent Phishing URL Detection System was

evaluated using labeled phishing and legitimate URL datasets. The machine learning classifiers were trained on extracted structural and lexical features and tested on unseen data to measure performance. The system achieved high classification accuracy with strong precision and recall values, indicating its effectiveness in correctly identifying malicious URLs while minimizing false alarms. The real-time evaluation demonstrated that the model can process incoming URLs efficiently and generate immediate decisions, making it suitable for practical deployment in user-facing environments.

The results show that feature-based machine learning significantly improves detection capability compared to traditional blacklist and rule-based methods, particularly for newly generated phishing URLs. The alert mechanism successfully prevented access to harmful links, while the administrative dashboard provided clear monitoring of detection outcomes and performance metrics. Overall, the system demonstrates reliable threat identification, reduced false positives, and scalable operation, confirming its effectiveness as an intelligent cybersecurity solution for phishing prevention.

The abc deep learning methodologies to improve the precision and efficacy of recognizing persons inside a criminal database. The system exhibits superior performance in photo matching and video surveillance by employing Convolutional Neural Networks (CNNs) for feature extraction and matching. The comprehensive preprocessing techniques guarantee that diverse input abc safety and improving security protocols. This model attains an accuracy of 98.5%, precision of 98%, true positive rate of 97.5%, and abc capabilities to boost its utility in diverse security situations.

5. Acknowledgment

The authors would like to express their sincere gratitude to Dr. K. Karuppasamy, Head of the Department of Computer Science and Engineering, RVS College of Engineering and Technology, for his continuous guidance, encouragement, and valuable suggestions throughout the development of this project. His support and motivation played a crucial role in the successful completion of the work.

We would like to thank all the faculty members of the Department of Computer Science and Engineering for their cooperation and support. We are also grateful to the management of RVS College of Engineering and Technology for providing the necessary facilities and resources to carry out this project successfully.

Finally, we would like to express our sincere appreciation to our friends and family members for their encouragement and moral support, which helped us stay motivated throughout the project duration.

Conclusion

This research presented an Intelligent Phishing URL Detection System that applies machine learning techniques to identify malicious web links in real time. By combining URL preprocessing, feature extraction, and supervised classification, the system effectively distinguishes phishing URLs from legitimate ones based on structural and behavioral characteristics. The proposed framework overcomes the limitations of traditional blacklist and rule-based approaches by enabling adaptive detection of previously unseen threats while maintaining high accuracy and low false positive rates.

The integration of an alert mechanism and administrative monitoring interface ensures practical usability, proactive protection, and continuous system evaluation. The results demonstrate that the proposed approach provides a reliable and scalable solution for enhancing web security and protecting users from phishing attacks. Future enhancements may include advanced deep learning models, expanded feature analysis, and integration with browser-based security tools to further strengthen detection capability.

References

1. Ian Goodfellow *et al*., *Deep Learning*, MIT Press, 2016.
2. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer
3. https://www.kaggle.com (Phishing URL Datasets)
4. https://www.sciencedirect.com
5. https://ieeexplore.ieee.org
6. https://scikit-learn.org
7. https://www.tensorflow.org.