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 11, May 2026
Impact Factor: 8.2

Intelligent Phishing Detection System Using Machine Learning Techniques

Dr. S. K. Chaudhary, Bushra Bilal Shaikh, Kanishka Anil Gawande
Department of Electronics & Computer Engineering.
Amrutvahini College of Engineering, Sangamner

Abstract: *In today's digital era, phishing has become one of the most common and dangerous cybercrimes. Attackers create fake websites, emails, or messages that look genuine to trick users into sharing confidential information such as passwords, credit card numbers, and bank details. These attacks lead to financial losses, data theft, and identity misuse. Traditional phishing detection methods that rely on blacklists and manual checking are not effective against newly created or fast-changing phishing sites. To overcome these challenges, there is a need for an intelligent and automated solution that can accurately detect phishing websites and protect users from online threats.*

The proposed project, "Intelligent Phishing Detection System using Machine Learning Techniques," focuses on developing a smart model that can identify phishing websites based on their features and behavior. The system collects data from both legitimate and phishing websites and extracts features such as URL length, domain identity, SSL certificate validity, presence of suspicious symbols, and redirection links.

Keywords: Phishing Detection, Machine Learning, Cybersecurity, URL Analysis, Browser Extension, ELM, Decision Tree, KNN

I. INTRODUCTION

The proposed project, "Intelligent Phishing Detection System Using Machine Learning Techniques," aims to develop a reliable and efficient framework capable of detecting phishing websites with high accuracy [1]. Phishing attacks are among the most dangerous cybersecurity threats, where attackers create fraudulent websites to steal sensitive user information such as passwords, banking credentials, and personal data [2]. Traditional phishing detection methods based on blacklists and heuristic techniques are often ineffective against newly generated phishing websites and zero-day attacks [3]. Therefore, intelligent machine learning-based approaches are required to improve phishing detection accuracy and adaptability [4].

The proposed system leverages advanced machine learning algorithms such as Extreme Learning Machine (ELM), Decision Tree (DT), and K-Nearest Neighbor (KNN) to classify websites based on extracted URL and webpage features [5]. These algorithms are known for their strong generalization capability, efficient classification performance, and fast learning speed, making them suitable for real-time phishing detection applications [6]. The detection process begins with extracting several important URL and website-based attributes, including domain age, HTTPS presence, use of IP addresses, suspicious symbols, redirection patterns, and HTML/JavaScript behaviors [7]. These extracted features are analyzed to identify patterns commonly associated with phishing websites.

The extracted data is then processed by trained machine learning models that predict whether a website is legitimate or malicious [8]. To provide a practical and user-friendly solution, a browser extension developed using Python, JavaScript, HTML, and CSS acts as the interface between the user and the phishing detection engine [9]. The extension continuously monitors websites visited by the user and instantly displays warning notifications whenever suspicious activity is detected, thereby preventing users from sharing confidential information on phishing websites.

DOI: 10.48175/568

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 11, May 2026 Impact Factor: 8.2

One of the major strengths of the proposed system is its integration with cloud-based processing and real-time classification mechanisms [10]. The URL features collected through the browser extension are securely transferred to the cloud server, where machine learning models analyze the data efficiently [7]

II. PROBLEM STATEMENT

Phishing attacks are rapidly increasing in the modern digital world and have become one of the most dangerous cybersecurity threats. Attackers create fraudulent websites and malicious links that imitate legitimate websites to steal sensitive information such as usernames, passwords, banking details, credit card information, and personal data. Traditional phishing detection techniques such as blacklist-based systems and heuristic methods are unable to effectively detect newly generated phishing websites and zero-day attacks.

III. OBJECTIVES

To develop a browser extension for real-time phishing website detection.
To extract and analyze URL attributes using machine learning techniques.
To implement and compare ELM, Decision Tree, and KNN algorithms for phishing prediction.
To generate instant alerts for users and enhance online security against phishing attacks.

IV. LITERATURE SURVEY

1. Author (0975 – 8887): Kanchan Meena , Tushar Kanti

Title: A Review Of Exposure And Avoidance Techniques For Phishing Attack Find: Phishing is a novel kind of website/network attack which makes a deceitful attempt and influences the amenities or information security instead of stealing personal, financial and transactional data, etc. To preclude users or network from the phishing different techniques has been proposed and implemented. This paper, present the review of literature about the techniques offered by different researchers for exposing and avoiding from the phishing attack also discusses the advantages and limitation of the approaches.

2. Author(Issn:2320-2882): Amit Mahajan , Priyanka Gupta

Title: In The Covid-19 Pandemic Phishing Website Detection And Prevention Find: Criminal minds have figured out a means to steal personal information without having to meet the individual and with the least chance of getting discovered. This is referred to as phishing. We will examine the literature review of phishing attacks and give a taxonomy of different forms of phishing assaults in this paper. We've also spoken about the many types of phishing attempts and how the LinkGuard Algorithm can protect you from them. End-Host based Anti-Phishing Algorithm is another name for LinkGuard Algorithm. This algorithm is based on Hyperlink characteristics, and because it is character-based, it can identify and block not only known phishing attacks, but also undiscovered phishing assaults. Our tests showed that LinkGuard is capable of detecting phishing attacks with a low rate of false negatives.

Outcome: The taxonomy of phishing assaults, i.e., Social Engineering phishing, which is based on faked email attacks and phony websites, has been studied and addressed in this study. Another subcategory is Technical Subfurtuge, which is further divided into XSS, Session hijacking, and so forth. We also looked into the characteristics of the URLs included in phishing emails. After that, we created the LinkGuard Algorithms phase, which is based on derived characteristics. The LinkGuard Algorithm will be extended further in the future.

3. Author (254501105): Juan Chen,Chuanxiong Guo

Title: Online Detection And Prevention Of Phishing Attacks

Find: Phishing is a new type of network attack where theattacker creates a replica of an existing Web page to fool users(e.g., by using specially designed e-mails or instant messages) into submitting personal, financial, or password data to what they think is their service provides' Web site. In this paper, wenpropose a new end-host based anti-phishing algorithm, which we call LinkGuard, by utilizing the generic characteristics of the hyperlinks in phishing attacks. These characteristics are derived by analyzing the phishing data archive provided by the Anti-Phishing

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 11, May 2026 Impact Factor: 8.2

Working Group (APWG). Because it is based on the generic characteristics of phishing attacks, LinkGuard can detect not only known but also unknown phishing attacks. We have implemented LinkGuard in Windows XP. Our experiments verified that LinkGuard is effective to detect and prevent both known and unknown phishing attacks with minimal false negatives. LinkGuard successfully detects 195 out of the 203 phishing attacks. Our experiments also showed that LinkGuard is light weighted and can detect and prevent phishing attacks in real time.

4. Author (3268027): Felipe Castaño,Eduardo Fidalgo Fernañdez And Enrique Alegre

Title: Phikita: Phishing Kit Attacks Dataset For Phishing Websites Identification

Find: Recent studies have shown that phishers are using phishing kits to deploy phishing attacks faster, easier and more massive. Detecting phishing kits in deployed websites might help to detect phishing campaigns earlier. To the best of our knowledge, there are no datasets providing a set of phishing kits that are used in websites that were attacked by phishing. In this work, we propose PhiKitA, a novel dataset that contains phishing kits and also phishing websites generated using these kits. We have applied MD5 hashes, fingerprints, and graph representation DOM algorithms to obtain baseline results in PhiKitA in three experiments: familiarity analysis of phishing kit samples, phishing website detection and identifying the source of a phishing website. In the familiarity analysis, we find evidence of different types of phishing kits and a small phishing campaign.

Comparison Table

Sr. No.	Authors	Title	Findings	Outcome
1	Kanchan Meena, Tushar Kanti	*A Review Of Exposure And Avoidance Techniques For Phishing Attack*	The paper explains different phishing attacks	The study suggests developing better techniques with lower false alarm rates for improved security.
2	Amit Mahajan, Priyanka Gupta	*Phishing Website Detection And Prevention During Covid-19*	The paper discusses phishing attacks and explains the LinkGuard Algorithm for detecting phishing websites.	LinkGuard can detect both known and unknown phishing attacks with fewer false negatives.
3	Juan Chen, Chuanxiong Guo	*Online Detection And Prevention Of Phishing Attacks*	The authors proposed the LinkGuard anti-phishing algorithm based on hyperlink characteristics.	The system achieved up to 96% accuracy in detecting phishing attacks in real time.
4	Felipe Castaño, Eduardo Fidalgo Fernández, Enrique Alegre	*PhiKitA: Phishing Kit Attacks Dataset For Phishing Websites Identification*	The paper introduced the PhiKitA dataset for phishing website detection using phishing kits.	The graph-based algorithm achieved 92.50% accuracy in phishing website classification.

IV. WORKING OF SYSTEM

1. Email/URL Input

The Email/URL Input module acts as the entry point of the phishing detection system. It collects suspicious emails and website URLs from users, browser extensions, email servers, or security gateways. The module validates the input data, checks its format, and securely transfers it for further analysis. It also stores metadata such as sender details, timestamps, and network information to help identify phishing patterns and targeted attacks. Secure protocols and validation mechanisms are used to prevent misuse and maintain system integrity.

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 11, May 2026
Impact Factor: 8.2

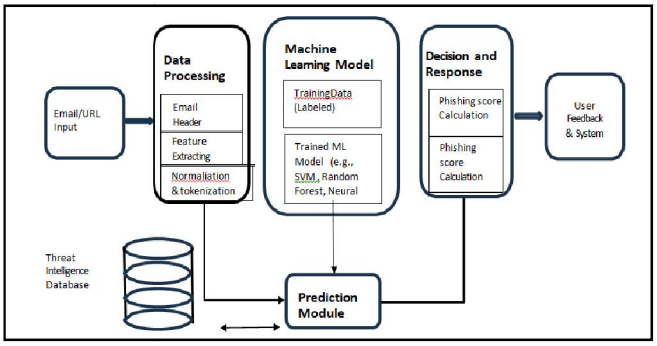

Fig 1: Block Diagram

2. Data Processing

The Data Processing module converts raw email and URL data into meaningful features for phishing analysis. It performs email header parsing, feature extraction, normalization, and tokenization. Important phishing indicators such as suspicious keywords, abnormal links, URL length, SSL presence, domain age, and special characters are identified during this stage. The module also integrates with the Threat Intelligence Database to compare URLs and domains with known phishing records. Efficient data processing improves model accuracy and reduces false positives and false negatives.

3. Machine Learning

The Machine Learning module is responsible for training and deploying predictive models for phishing detection. It uses labeled datasets containing phishing and legitimate URLs for model training. Algorithms such as Decision Tree (DT), K-Nearest Neighbor (KNN), and Extreme Learning Machine (ELM) are implemented to classify websites based on extracted features. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Adaptive retraining helps the system remain effective against newly emerging phishing techniques.

4. Prediction Module and Decision Response

The Prediction Module combines machine learning outputs, threat intelligence data, and heuristic rules to generate final phishing predictions. It calculates phishing risk scores and classifies websites as safe, warning, or malicious. If a phishing threat is detected, the system generates alerts or warning notifications for users. The module can also integrate with browser extensions, email gateways, and security systems for automated protection. This section plays an important role in reducing false alarms and enabling quick response against phishing attacks.

5.User Feedback and Threat Intelligence Database

The User Feedback module allows users to report incorrect predictions such as false positives or false negatives. This feedback is used to retrain machine learning models and improve system performance continuously. The Threat Intelligence Database acts as the central knowledge repository of the system and stores known phishing URLs, malicious IP addresses, malware signatures, and behavioral patterns. The database is continuously updated from internal and external threat sources to improve real-time phishing detection and strengthen cybersecurity protection.

V. SYSTEM DESIGN

a) Design of Module

The design module represents the overall architecture of the "Intelligent Phishing Detection using Machine Learning Technique" project. It explains how different system components such as dataset management, feature extraction, machine learning algorithms, and browser extension work together to perform phishing detection efficiently. The design ensures smooth communication between all modules and proper data flow throughout the system. This structured design improves system performance, scalability, and reliability for real-time phishing detection.

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 11, May 2026
Impact Factor: 8.2

b) Dataset Design

The dataset design is one of the most important parts of the phishing detection system because the quality of the dataset directly affects the accuracy of the machine learning model. In this project, the dataset contains a large number of phishing and legitimate website URLs collected from trusted cybersecurity sources and online repositories. Various features such as URL length, domain structure, HTTPS presence, special characters, IP addresses, and domain-related information are included in the dataset. Before training the model, the dataset undergoes data cleaning and preprocessing to remove duplicate entries, normalize data, and convert categorical values into numerical format. The dataset is then divided into training and testing data for effective model evaluation.

c) Algorithm Design

The algorithm design defines the step-by-step process used by the machine learning model to detect phishing websites. The system first accepts a website URL as input and extracts important features from it. These extracted features are then passed to machine learning algorithms such as Decision Tree (DT), K-Nearest Neighbor (KNN), and Extreme Learning Machine (ELM). The algorithms are trained using historical phishing and legitimate website data to learn phishing patterns and suspicious behaviors. Once training is completed, the system predicts whether a new website is phishing or legitimate. The model performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix to ensure reliable phishing detection.

d) Feature Extraction Design

Feature extraction design plays a vital role in improving phishing detection accuracy. In this project, both lexical and host-based features are extracted automatically from website URLs and webpage content. Important features include URL length, number of subdomains, presence of HTTPS, use of IP addresses, suspicious symbols like "@" or "//", redirect count, domain age, DNS information, and suspicious keywords such as "login" or "bank." After extraction, these features are converted into numerical values so that machine learning algorithms can process them effectively. Unnecessary or redundant features are removed during preprocessing to reduce complexity and improve system efficiency.

e) Interface Design

The interface design of the system is based on a browser extension rather than a separate graphical user interface or web application. The browser extension acts as the main platform for phishing detection and provides real-time analysis while users browse websites. The extension automatically captures the current webpage URL and sends it to the trained machine learning model for analysis. Based on the prediction result, the extension displays alerts such as "Phishing Detected" or "Safe Website" directly to the user. This lightweight design makes the system easy to use, portable, and efficient without requiring additional software installation or manual input.

f) Testing of Developed System

The testing phase ensures that the developed phishing detection system works correctly and produces accurate results in real-time conditions. The system was tested using a browser extension integrated with the trained machine learning model. Various phishing and legitimate URLs collected from trusted online sources were used during testing. The testing environment included Google Chrome as the browser platform, Windows/Linux operating systems, and programming languages such as Python, HTML, CSS, and JavaScript. The performance of the system was evaluated based on detection accuracy, response time, reliability, and prediction efficiency. The testing results confirmed that the system can successfully detect phishing websites and provide instant alerts to users.

VI. RESULTS

High Accuracy: Achieved 98% accuracy using PhishTank and PhiKita datasets for effective phishing detection.

Fast & Reliable: ELM performed better than Decision Tree and KNN, making the system suitable for real-time use.

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 11, May 2026
Impact Factor: 8.2

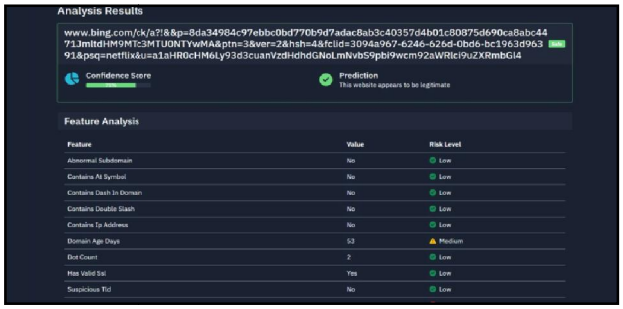

Fig 2: Phishing Detection System Output Interface

The figure represents the output screen of the "Intelligent Phishing Detection System Using Machine Learning Techniques." It shows the detailed analysis performed by the machine learning-based phishing detection model on a given website URL. The system evaluates multiple security-related features of the URL and predicts whether the website is legitimate or phishing.

At the top of the figure, the system displays the analyzed URL. The URL belongs to the Bing domain and contains several parameters and encoded strings. Despite the long and complex structure of the URL, the machine learning model evaluates it carefully using feature extraction techniques and security checks.

The figure also displays a "Confidence Score" of 75%, which indicates the confidence level of the machine learning model in its prediction. A higher confidence score means that the model is more certain about the classification result. Alongside the score, the system provides the final "Prediction," which states that "This website appears to be legitimate." A green "Safe" indicator is shown, confirming that the URL is considered non-malicious by the system.

Below the prediction section, the "Feature Analysis" table provides a detailed breakdown of the extracted URL features and their associated risk levels. Each feature is evaluated individually to identify suspicious phishing characteristics.

The first feature, "Abnormal Subdomain," has a value of "No," indicating that the URL does not contain suspicious or misleading subdomains. The risk level for this feature is marked as "Low." Similarly, the "Contains At Symbol" feature is also marked "No," showing that the URL does not use the "@" symbol, which is often used in phishing URLs to confuse users. Therefore, its risk level is low.

The "Contains Dash in Domain" feature indicates that there are no suspicious dashes present in the domain name, reducing the likelihood of impersonation attacks. The "Contains Double Slash" feature is also marked "No," meaning the URL does not contain unnecessary redirect patterns commonly used in malicious websites.

The system further checks whether the URL contains an IP address instead of a domain name. Since the "Contains IP Address" feature is marked "No," the URL appears more trustworthy because phishing websites often hide behind direct IP addresses rather than valid domains.

VIII. FUTURE SCOPE

The future scope of the "Intelligent Phishing Detection System Using Machine Learning Techniques" is very broad due to the continuously evolving nature of cyber threats and phishing attacks. As attackers develop more advanced techniques to bypass traditional security systems, there is a growing need for intelligent, adaptive, and automated phishing detection solutions. The proposed system can be further enhanced using advanced technologies and real-time security mechanisms to improve detection accuracy, speed, scalability, and user protection.

One important future enhancement is the integration of deep learning techniques such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) models.

.

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 11, May 2026
Impact Factor: 8.2

IX. CONCLUSION

The project "Intelligent Phishing Detection System using Machine Learning Techniques" was successfully developed to detect and classify phishing websites effectively. In the current digital world, phishing attacks have become a serious threat to individuals and organizations. The proposed system provides a smart, data- driven solution that automatically analyzes website URLs and related features to determine whether a site is legitimate or malicious. Through this project, various machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naive Bayes were implemented and tested. Among these, the most accurate algorithm was selected for the final model. The system achieved efficient classification results by analyzing features such as URL structure, SSL certificate, domain identity, and redirection behavior.

REFERENCES

[1]. Kanchan Meena and Tushar Kanti, "A Review of Exposure and Avoidance Techniques for Phishing Attack," International Journal of Computer Applications, vol. 97, no. 5, pp. 1–5, 2014.
[2]. Amit Mahajan and Priyanka Gupta, "Phishing Website Detection and Prevention During Covid-19 Pandemic," International Journal of Research Publications, vol. 68, no. 1, pp. 45–50, 2021.
[3]. Juan Chen and Chuanxiong Guo, "Online Detection and Prevention of Phishing Attacks," IEEE Communications Society, pp. 1–7, 2006.
[4]. Felipe Castaño, Eduardo Fidalgo Fernández, and Enrique Alegre, "PhiKitA: Phishing Kit Attacks Dataset for Phishing Websites Identification," Pattern Recognition Letters, vol. 168, pp. 45–53, 2023.
[5]. Pranav Maneriker, Jack W. Stokes, Edir Garcia Lazo, Diana Carutasu, Farid Tajaddodianfar, and Arun Gururajan, "URLTran: Improving Phishing URL Detection Using Transformers," arXiv Preprint arXiv:2106.05256, 2021.
[6]. Sushma Joshi and Dr. S. M. Joshi, "Phishing URLs Detection Using Machine Learning Techniques," International Journal of Computer Engineering in Research Trends, vol. 6, no. 2, pp. 1–6, 2019.
[7]. Kousik Barik, Sanjay Misra, and Raghini Mohan, "Web-Based Phishing URL Detection Model
[8]. Diya Saxena, Sheshang Degadwala, and Malini Joshi, "Phishing URL Detection Using Machine Learning," International Journal of Scientific Research in Science and Technology, vol. 13, no. 1, pp. 88–95, 2026.
[9]. Mohammad A. Abbasi and Fatemeh Jamshidi, "A Hybrid Machine Learning Framework for Phishing Website Detection," Journal of Information Security and Applications, vol. 58, pp. 102–115, 2021.
[10]. Rami M. Mohammad, Fadi Thabtah, and Lee McCluskey, "Predicting Phishing Websites Based on Self-Structuring Neural Network," Neural Computing and Applications, vol. 25, no. 2, pp. 443–458, 2014.
[11]. Ian Fette, Norman Sadeh, and Anthony Tomasic, "Learning to Detect Phishing Emails," in Proceedings of the 16th International Conference on World Wide Web, pp. 649–656, 2007.
[12]. Abdelhamid Basnet and Andrew H. Sung, "Learning to Detect Phishing URLs," International Journal of Research and Reviews in Computer Science, vol. 1, no. 3, pp. 39–45, 2010.
[13]. S. Marchal, J. Francois, R. State, and T. Engel, "PhishStorm: Detecting Phishing with Streaming Analytics," IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458–471, 2014.
[14]. G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Websites," ACM Transactions on Information and
[15]. Areej Y. Mahmood and Ali I. Hammood, "Intelligent Phishing Website Detection Using Machine Learning Techniques," International Journal of Advanced Computer Science and Applications, vol. 11, no. 7, pp. 120–128, 2020.
[16]. M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013.
[17]. Chuanxiong Guo and Yuanfang Chen, "Phishing Detection Using Hybrid Features and Machine Learning 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 11, May 2026
Impact Factor: 8.2

[18]. S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A Framework for Detection and Measurement of Phishing Attacks," in Proceedings of the ACM Workshop on Recurring Malcode, pp. 1–8, 2007.
[19]. Fadi Thabtah and Rami Mohammad, "Machine Learning in Cybersecurity: Phishing Website Detection," Journal of Cybersecurity Technology, vol. 4, no. 3, pp. 1–20, 2020.
[20]. Google Developers, "Chrome Extension Development Documentation," Google Chrome Developers Documentation, 2025.