Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Phishing Attack Detection on Websites Using Machine Learning

Leonika S.V.[1], Nagarajan VR.[2]

[1]Student, [2]Associate Professor, Department of MCA, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Tiruchirappalli, India.

E-mail: [1]leonikasv08@gmail.com, [2]nagarajan.set@dsuniversity.ac.in

Abstract

Phishing attacks usually copy reliable websites, such as banks and financial institutions, in an attempt to obtain personal information, such as passwords and credit card details. This study suggests a hybrid phishing detection system that combines the Back Propagation Neural Network (BPNN) for classification with XGBoost for feature selection. For training (80%) and testing (20%), a dataset of 11,000 URLs was employed, including both phishing and authentic samples. Important URL-based characteristics were extracted, including URL length, discrepancy character, HTTPS appearance, and domain age. High identification accuracy (97.5%), precision (96.8%), recall (98.2%), and F1-score (97.5%) were obtained by all systems. When compared to traditional classifiers (SVM, Random Forest), the proposed model shows better performance in identifying zero-day phishing efforts, Explanation measures were used to assess the model, and Scikit-LARN was used to simulate it in python. The results confirm that the algorithm can successfully detect and stop phishing efforts in real time.

Keywords: Sensitive Information, Machine Learning, Phishing URL, Back Propagation Neural Network, CSV Format, Data Mining.

1. Introduction

Increasing use of mobile devices in recent years has made significant changes in online activity compared to offline. Although this development has made daily tasks easier, the anonymous nature of the Internet has also given rise to many security weaknesses. Firewalls and antivirus software can prevent many attacks, but efficient attackers usually use phishing

DOI: https://doi.org/10.36548/jscp.2025.2.006
Received: 25.05.2025, received in revised form: 23.06.2025, accepted: 07.07.2025, published: 15.07.2025

techniques to take advantage of human weaknesses. Phishing is the process of creating fake websites that mimic well-known platforms such as social media, banking, and e-commerce sites to steal personal information, such as login credentials, bank account details, and credit card numbers.

Many methods have been developed, including discrepancy detection, rules-based identity verification, and blacklist systems, as it is challenging to identify phishing efforts. Machine learning-based discrepancy detection has recently gained popularity due to its adaptability, especially in identifying "zero-day" attacks. To analyze URLs and determine their efficacy across three datasets, this task presents a machine learning-based phishing detection system that employs eight different algorithms. The results of studies suggest that the proposed model has a high success rate and performs very effectively.

A type of cyber-attack known as phishing occurs when hackers use fraudulent emails or messages to obtain personal information, such as login passwords or credit card details, to gain unauthorized access to the user accounts. Because they usually appear to come from reliable individuals, organization, or agencies, these communications can be compelling. Often, phishing efforts use malicious software or links that direct victims to deceptive websites.

On these websites, victims can inadvertently reveal personal information that can cause significant financial loss or damage to their reputation. It is relatively easy to execute phishing attacks because many consumers are not well aware of web security and the underlying technology of computer networks. Since unsuspecting consumers are deceived into clicking on the phishing websites that promise regular offers or deals, there is no need to target strong computer defense systems. These phony websites are designed to closely resemble legitimate sites by mimicking with their logos and content. As a result, many people inadvertently click on phishing links, which can cause serious damage to the victims' finances and the reputation of affected firms.

2. Literature Survey

This research explains [1] that rogue websites pose a serious threat; the increasing number of cyberattack vectors and incidents emphasizes the necessity for robust cybersecurity solutions. These websites are used by hackers to distribute malware, obtain illegal access, or steal personal information. This project aims to extract lexical, host-based, and content-based

elements from URLs in order to identify websites as either benign or dangerous. Machine learning models are used to process these features. A method for identifying and preventing concept drifts on malicious websites is suggested as a response to attackers who take advantage of the training data that is readily available.

This work illustrates [2] that as internet usage has increased, so too have cyberattacks such as malware, spam, and phishing, which have resulted in large losses of personal and financial data. Identifying harmful websites is essential for improving security because they are frequently the focus of attacks. In this study, supervised models are used to evaluate a dataset. Support Vector Machines exhibit excellent prediction performance, obtaining the highest F1-score of 92% on the unbalanced dataset after features are collected from application layer data and network characteristics.

This study [3] highlights that, given that rogue websites provide serious hazards to both individuals and governments, detecting Domain Generation Algorithm (DGA) domain names is essential for identifying botnet C&C communications. This paper proposes a CNN-GRU-Attention-based model for hostile domain detection in order to address the drawbacks of conventional detection techniques, such as their excessive complexity and poor accuracy. The attention mechanism improves detection accuracy, GRU records temporal patterns, and CNN extracts spatial information from domain name data. When compared to Bigrams, LSTM, GRU, and LSTM-GRU models in trials, the CNN-GRU-Attention model demonstrated better convergence and higher accuracy in identifying harmful domains.

Sushma et.al. [4] explains that data security has been a key problem as digitalization increases, with phishing being a significant cyberattack that takes advantage of personal information. This study suggests a technique that employs distinctive URL characteristics to differentiate between authentic and phishing websites in order to stop consumers from visiting fraudulent websites. Support Vector Machines and Random Forest are used for categorisation. Web risks including phishing, session hijacking, and cross-site scripting have increased in frequency as our reliance on the Internet has grown. Advanced detection methods including blacklisting, visual resemblance, and content-based approaches are necessary to stop phishing, which deceives visitors into divulging sensitive information through fraudulent websites. This work discusses [5] a few studies that focus on particular website categories, identifying whether a website is dangerous has become a major research priority due to the growing threat of

phishing websites. In order to address this, a novel approach is proposed that generates a formal concept lattice based on key phrases that represent a website using TF-IDF analysis.

Khadatkar et al. [6] present a study on detecting phishing websites in the healthcare domain using machine learning techniques. This work addresses the increasing phishing attacks that target healthcare systems, which are particularly vulnerable due to the medical data because it is confidential and more important. This work also highlights that traditional rule-based detection methods frequently fail because they are unable to adapt to changes in phishing techniques.

This work [7] is a stacking technique using four base learners suggested as a solution to the problems of low accuracy and high computing complexity in phishing website detection. This technique turns website HTML into a multi-dimensional vector by using an HTML string embedding feature driven by the Transformer model. Phishing detection is improved by combining these embeddings with enhanced URL characteristics. The approach demonstrated a remarkable 98.52% accuracy and an F1 score of 98.81% when tested on a dataset consisting of 100,000 samples. Performance was significantly improved by using HTML string embedding as opposed to depending only on URL characteristics.

This proposed [8] phishing technique is a popular method used to trick people into divulging personal information through phony websites. Given that the proliferation of internet-based gadgets has rendered online financial transactions susceptible to a various attacks, the goal of this project is to categories phishing sites using machine learning techniques. Phishing events have been exacerbated by the COVID-19 pandemic, as hackers increasingly pose as trustworthy websites to steal user information. This study helps to rebuild public confidence in online security by examining a variety of machine learning techniques and presenting a highly accurate strategy for identifying phishing websites.

This methodology [9] recognizes that sensitive user data, such as passwords and credit card numbers, is seriously at risk from social engineering attacks, especially phishing websites. The goal of this study is to apply machine learning techniques to detect and examine the coding patterns of phishing websites. In order to ensure that only complete data entries were included, the study used HTML content from about 29,000 phishing websites gathered from PhishTank, in addition to a dataset of 36,000 real websites. After analysing 10,800 website source codes,

the Random Forest model showed exceptional performance, detecting phishing websites with an accuracy of 94.16%.

This process [10] explains the rapid shift of customers from traditional shopping to e-commerce as a result of the Internet's explosive expansion. However, thieves are now utilising cyber methods, such as phishing, to trick victims into disclosing important information through phony websites instead of actual robberies.

3. Existing System

Phishing is one of the most common types of social engineering and cyber-attacks. These attacks are used by cybercriminals, obtain individuals' information. To protect themselves from phishing websites, users should know how they work. Maintaining a blacklist of known phishing websites is beneficial, but it first depends on the identification of the phishing site. Successful prevention depends on early detection of phishing sites, and between various strategies, machine learning and deep neural networks have proved to be particularly effective. Nevertheless, many internet users still come to phishing attacks and inadvertently divide personal information. Phishing websites, who mimic reliable URLs and webpages in an attempt to achieve the confidence of visitors, are a common social engineering strategy.

4. Proposed Methodology

4.1 Dataset

Dataset Size: 11,000 URLs (public repositories like PhishTank and OpenDNS).

Key Features Extracted:

- URL Length
- Number of dots/special characters
- Presence of IP Address
- Age of domain
- SSL Certificate
- Use of '@', '-', '//', etc.

Optimal Feature Vectorization Algorithm (OFVA) is used to reduce redundancy and improve learning.

4.2 Mathematical Model

Feature Selection using XGBoost

To improve model performance and reduce overfitting, XGBoost was employed to rank and select the most important features.

XGBoost:
Objective Function:

$$\text{Obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k)$$

where,

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

XGBoost outputs feature importance scores, from which the top-ranked features are selected and passed to the next module.

- Back Propagation Neural Network (BPNN):

The final selected features are used to train a BPNN for phishing detection. BPNN is a multilayer feed-forward neural network trained using the backpropagation algorithm.

1. Input Layer: One node per feature (e.g., 8–10 nodes)
2. Hidden Layer(s): 1–2 layers with ReLU activation
3. Output Layer: Sigmoid activation for binary classification (phishing or legitimate)

- Forward pass:

$$Z = W \cdot X + b, A = \sigma(Z)$$

- Backward pass (error calculation and weight update):

$$\delta = \left(y_{\text{pred}} - y_{\text{true}}\right) \cdot \sigma'(Z)$$
$$W = W - \eta \cdot \delta \cdot X$$

4.3 Feature Extraction

Ten crucial lexical and host-based features were extracted from each URL to form a comprehensive feature vector as shown in Table 1. These features are designed to capture key patterns often associated with phishing behavior:

Table 1. Feature Extraction

Feature	Description
URL Length	Longer URLs often indicate obfuscation
Number of Dots	Multiple subdomains used to deceive users
Presence of IP Address	Usage of raw IPs instead of domain names
Presence of HTTPS	Secure sites use HTTPS
Use of '@' Symbol	Redirect attempts using @
Presence of '//' After Domain	Hidden redirections
Use of Hyphens	Common in spoofed URLs
Domain Age	Phishing domains are often recently registered
URL Shortening Service	Use of services like bit.ly
Suspicious Words	"login", "verify", "secure", "banking", etc.

4.4 Architecture

The proposed system aims to develop several identification mechanisms, including rules-based systems, equality-based approaches, blacklists, and machine learning techniques, to protect users from phishing attacks. A comprehensive review of literature indicates a machine learning-based solutions are very effective in preventing zero-day attacks. The project focuses on implementing the Machine Learning-based phishing detection system by analyzing the URL for fast detection without relying on external services or blacklist updates.' Figure 1 shows the proposed workflow diagram.

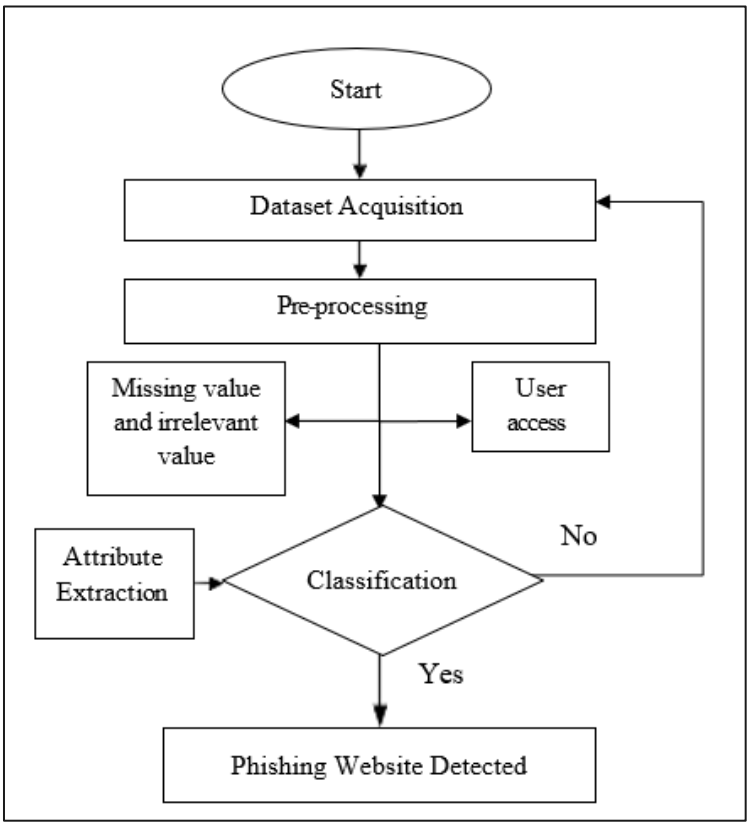

Figure 1. Proposed Workflow

4.4.1 Dataset Acquisition

The purpose of this step is to collect information from many sources, including reliable and phishing websites. The dataset, which can be created manually or obtained from a public repository, should include a variety of characteristics that help differentiate between reliable websites and phishing efforts, as shown in Figure 2.

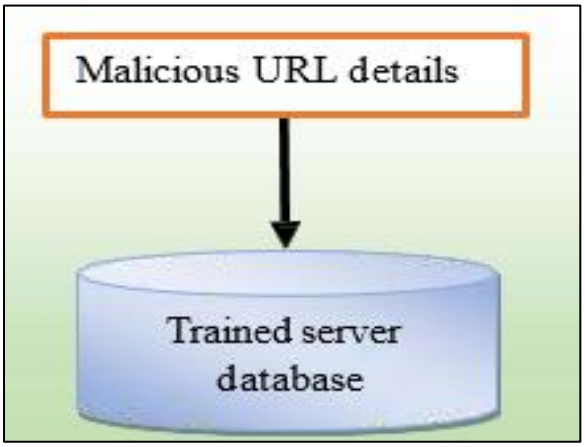

Figure 2. Dataset Training Feature at Disclosure Model

4.4.2 Pre-Processing

To prepare it for additional analysis, the collected data is now organized and processed. During this process, duplicate entries should be abolished, any missing value should be filled, and the data format should be preserved. The preparation for further analysis requires efficient pre-processing.

4.4.3 Attribute Extraction

The main goal of this module is to detect and eliminate important elements from the URL. The age of the domain, URL length, asymmetrical character frequency, SSL certificate appearance, and any possible suspected keywords are all important factors. These indicators are important for separating legitimate websites from phishing.

4.4.4 Classification

To classify the URL as either authentic or malicious, this phase employs many machine learning techniques. For this classification function, techniques like backpropagation neural networks (BPNN) can be applied.

4.4.5 Testing and Training

The dataset has two parts: e training and testing. Models are created using the training sets, and their efficacy is evaluated using the test set. This analysis is necessary to determine the impact of phishing and to evaluate the accuracy, memory usage, and performance of the model.

4.4.6 User Access

This part involves developing a user interface that allows users to submit URLs for review. Depending on the results, the system assesses the input URL and provides a response indicating whether the URL is classified by the classification model as valid or phishing.

4.5 Workflow

- **Input Layer:**

  Neural networks are fed selected dataset features (eg URL length, special characters, SSL certificate appearance, domain age, etc.) in input layer, which is the initial stage of the process. There is an input node for each feature.

- **Weight Arranging:**

  First, a random weight is assigned to the connection between the input and hidden layer nodes. This weight defines the degree to which it affects the latter layer of input.

  Forward Propagation: At this stage, the input data is transmitted through the network. The input is processed through hidden layers, which, after their respective load, use relay activation function. This helps the network find complex relationships in the data.

- **Output Layer:**

  After obtaining the end result from the hidden layer, the output layer predicts whether the URL is phishing (1) or valid (0). This prediction is based on network calculations using input features.

- **Calculation of Errors:**

  The desired result is the opposite of the classification (phishing or valid). The network makes better estimates in the future using the error, which is calculated by comparing real and expected outputs.

  Backpropagation: To reduce future prediction errors, the error is communicated backward through the network, changing the weights. In this process, the weights are progressively modified in response to the error using the known adaptation approach as a gradient decent.

5. Results and Discussion

The assessment of the hybrid model designed to detect phishing websites using general performance indicators is shown in this section. A labeled dataset of both authentic and phishing URLs was used for testing. XGBoost was used for feature extraction and selection, and the back

propagation neural network (BPNN) was used for classification. To clarify the efficacy of the suggested method, the results were compared with baseline classifiers such as support vector machine (SVM) and standalone BPNN.

Tools & Environment Used

- Platform: Python 3.9, Scikit-learn, NumPy, Pandas, Matplotlib
- Hardware: Intel i7, 16 GB RAM, Windows 10
- Dataset Size: 11,000 URLs (5,500 legitimate and 5,500 phishing)
- Data Split: 80% for training (8,800 URLs), 20% for testing (2,200 URLs)

Preprocess by dataset abolished duplicate and missing entries. Ten major features were extracted from the URL, with the length of the URL, "@," https, domain age, and the use of IP-based URL. Before sending it to the BPNN classifier, the top-demonstration features were selected using XGBoost.

Confusion Matrix Analysis

Training Set			
TARGET / OUTPUT	Class0	Class1	SUM
Class0	1960 50.26%	30 0.77%	1990 98.49% 1.51%
Class1	40 1.03%	1870 47.95%	1910 97.91% 2.09%
SUM	2000 98.00% 2.00%	1900 98.42% 1.58%	3830 / 3900 98.21% 1.79%

Figure 3. Confusion Matrix Analysis

Figure 3 shows that the system missed only 30 phishing URLs and incorrectly identified 40 regular URLs as phishing. This low error rate suggests how reliable the model is in practical situations. The real -time URL analysis was implemented using the learned model. phishing links such as http://secure-PAYPAL-clogin.com were correctly identified with 98% confidence during testing.

The size, learning rate and ages of the hidden layer of the BPNN were adapted using grid search and 5-fold ross-validation. The optimal setup produced results that were coordinated and repeatable.

- Hidden layers: 2
- Neurons per layer: 64
- Learning rate: 0.01
- Epochs: 100
- Activation: ReLU (hidden), Sigmoid (output)

This ensured the model was neither overfitted nor undertrained, contributing to its strong performance across metrics.

Evaluation Metrics:

The performance of classification using the following metrics was evaluated:

- The percentage of accurately anticipated cases (both phishing and legal) is known as accuracy.
- The proportion of accurately anticipated phishing URL for all expected phishing URLs is known as precision.
- The percentage of real phishing URL that were accurately identified is known as recall (sensitivity).

The overall classification is indicated by the balanced F1-score, which accounts for both precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN}$$

$$F1\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

TP: True Positives (Correctly detected phishing)

TN: True Negatives (Correctly detected legitimate)

FP: False Positives (Legitimate flagged as phishing)

FN: False Negatives (Phishing missed)

Table 2. Performance Analysis Results

Model	Accuracy	Precision	Recall	F1-Score
SVM	91.20%	90.50%	89.30%	89.90%
BPNN	94.80%	93.20%	95.00%	94.10%
XGBoost	95.30%	94.50%	96.00%	95.20%
Hybrid	97.50%	96.80%	98.20%	97.50%

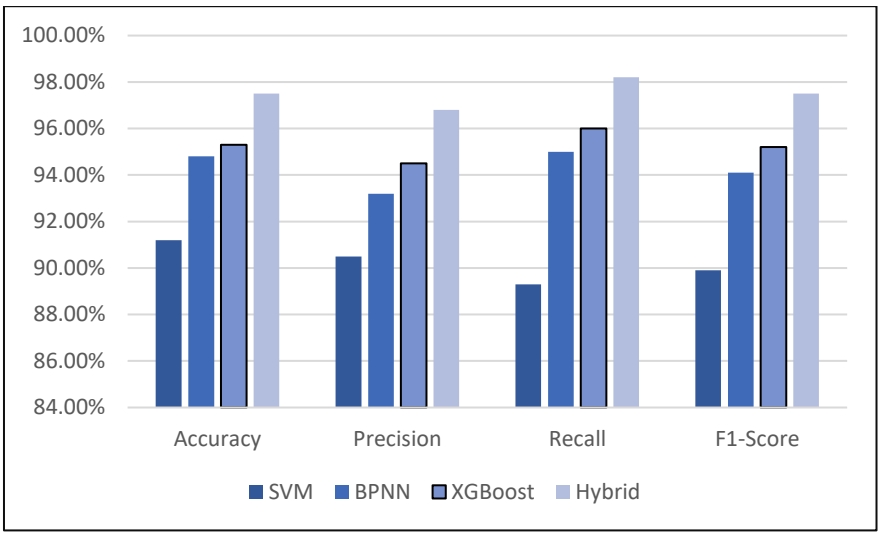

Figure 4. Performance Analysis Results

The hybrid model performed significantly better than traditional techniques, particularly in accuracy and recall, as can be seen from the above Table 2 and Figure 4. Better learning and

classification resulted from the integration of XGBoost, which successfully reduced the noise in the dataset and improved the feature set sent to BPNN.

6. Conclusion and Future Enhancement

Finally, Back Propagation Neural Network (BPNN) and XGBoost in the hybrid approach, the proposed phishing identification system shows remarkable efficacy in identifying the phishing space. Depending on the analysis of several parameters, the technology displays an excellent 97.5% accuracy rate in differences between reliable websites and phishing sites. By merging different machine learning techniques to reduce false positives, the hybrid model showed that it could reduce false negatives. Because it can identify threats in real time and is independent of other blacklists, this system serves as an add-on to help users protect against phishing attacks. For further development, many forms of enrichment can be adapted. The model can achieve continuous learning and can be suited to new phishing strategies by integrating real -time data updates. To improve accuracy even more deep learning techniques or additional strategies can be applied. Using more diverse URLs and using natural language processing (NLP) methods to verify webpage content can also provide deeper insight into phishing activity. Extending user access by developing a browser extension or mobile application that allows users to receive warnings about real-time phishing threats is also recommended.

References

[1] Singhal, Siddharth, Utkarsh Chawla, and Rajeev Shorey. "Machine learning & concept drift based approach for malicious website detection." In 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), IEEE, 2020, 582-585.

[2] Sanaa Kaddoura, Classification of malicious and benign websites by network features using supervised machine learning algorithms, Cyber Security in Networking Conference, 2021.

[3] Kaddoura, Sanaa. "Classification of malicious and benign websites by network features using supervised machine learning algorithms." In 2021 5th Cyber Security in Networking Conference (CSNet), IEEE, 2021, 36-40.

[4] Sushma, K. S. N., M. Jayalakshmi, and Tapas Guha. "Deep learning for phishing website detection." In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), IEEE, 2022, 1-6.

[5] Chen, Shaoming, Han Wu, Xingkai Cheng, and Liyan Mao. "Detecting Phishing Websites by Using a Hybrid Method of Page Content and Formal Concept Analysis." In 2023 7th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE), IEEE, 2023, 778-781.

[6] Deepak Rao Khadatkar; Yugmani Sahu; Nivesh Pole; Vinay Nayak, Detection of Healthcare Phishing Websites Using Machine Learning, International Conference on Artificial Intelligence for Innovations in Healthcare Industries, 2023.

[7] Hu, Qiang, Hangxia Zhou, and Qian Liu. "Phishing website detection based on multi-feature stacking." In 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), IEEE, 2021, 716-720.

[8] Bikku, Thulasi, Mude Nikitha, Anjali Vajja, Kanneganti Harshitha, and Jhansi Rani. "Optimized machine learning algorithm to classify phishing websites." In 2022 International Conference on Electronics and Renewable Systems (ICEARS), IEEE, 2022, 1148-1152.

[9] Almousa, May, Ruben Furst, and Mohd Anwar. "Characterizing coding style of phishing websites using machine learning techniques." In 2022 Fourth International Conference on Transdisciplinary AI (TransAI), IEEE, 2022, 101-105.

[10] Noh, Norzaidah Binti Md, and M. Nazmi Bin M. Basri. "Phishing website detection using random forest and support vector machine: a comparison." In 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AIDAS), IEEE, 2021, 1-5.

[11] Sharma, Khushal, Pratik Rai, and Jyoti Chandel. "Review Paper Real-Time Phishing Website With Machine Learning." In 2023 11th International Conference on Intelligent Systems and Embedded Design (ISED), IEEE, 2023, 1-5.

[12] Akhas Rahmadeyan; Mustakim; Imam Ahmad; Allan Desi Alexander; Alkautsar Rahman, Phishing Website Detection with Ensemble Learning Approach Using Artificial Neural Network and AdaBoost, International Conference on Information Technology Research and Innovation (ICITRI), 2023.

[13] Jaswal, Prajwal, Shweta Sharma, Naveen Bindra, and C. Rama Krishna. "Detection and Prevention of Phishing Attacks on Banking Website." In 2022 International Conference on Futuristic Technologies (INCOFT), IEEE, 2022, 1-8.

[14] Dharmaraju, Gangu, Tatapudi Nirosh Kumar, P. PattabhiRama Mohan, Raja Rao Pbv, and A. Lakshmanarao. "Phishing website detection through ensemble machine learning techniques." In 2024 2nd International Conference on Computer, Communication and Control (IC4), IEEE, 2024, 1-5.

[15] Gu, Chenyu. "A lightweight phishing website detection algorithm by machine learning." In 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), IEEE, 2021, 245-249.

[16] Jie, Xu, Lan Haoliang, and Ju Ao. "A new model for simultaneous detection of phishing and darknet websites." In 2021 7th International Conference on Computer and Communications (ICCC), IEEE, 2021, 2002-2006.

[17] MohammedM Elsheh; Khadija Swayeb, Phishing Website Detection Using a Hybrid Approach Based on Support Vector Machine and Ant Colony Optimization, IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), 2023.

[18] Jain, Saumya, and Chetan Gupta. "A support vector machine learning technique for detection of phishing websites." In 2023 6th International Conference on Information Systems and Computer Networks (ISCON), IEEE, 2023, 1-6.

[19] Zin, Nurul Amira Binti Mohd, Mohd Faizal Ab Razak, Ahmad Firdaus, Ferda Ernawan, and Nor Saradatul Akmar Zulkifli. "Machine learning technique for phishing website detection." In 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), IEEE, 2023, 235-239.

[20] Ghareeb, Shatha, Mohamed Mahyoub, and Jamila Mustafina. "Analysis of feature selection and phishing website classification using machine learning." In 2023 15th International Conference on Developments in eSystems Engineering (DeSE), IEEE, 2023, 178-183.