International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3278-3281
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0419

Deep Learning-Based Phishing URL Detection Using Long-Term Memory (LSTM)

Jesintha V[1], Gokulapriya R[2], Harini G[3], Elanchezhiyan E [4], Siva Ganesh M [5]
[1,2,3]Department of Computer Science and Engineering, Paavai Engineering College, Namakkal, Tamilnadu
[4]Assistant Professor, Department of Computer Science and Engineering, Paavai Engineering College, Namakkal, Tamilnadu
[5]Associate Professor, Department of Cyber Security, Fatima Michael College of Engineering and Technology, Madurai, Tamilnadu
Emails: *jesintha1993victor@gmail.com[1], gokulapriya03112004@gmail.com[2], hariniganapathi.2005@gmail.com[3], elanchezhiyanelangopec@paavai.edu.in[4], sivabma@gmail.com[5]*

Abstract
With the rapid growth of online services and digital transactions, phishing attacks have emerged as a serious Cybersecurity threat, targeting users by imitating legitimate websites to steal sensitive information. This project presents a machine learning-based phishing URL detection system that aims to identify and prevent access to fraudulent websites. The proposed approach utilizes a Random Forest classifier to distinguish between phishing and legitimate URLs based on a set of extracted websites and URL features. A labeled dataset containing both genuine and malicious URL, the system analyzes its features and classifies it as either safe or phishing in real time. The experimental results demonstrate that the proposed system effectively detects phishing URLs with high accuracy, thereby enhancing user security and reducing the risk of online fraud. This approach provides a reliable and automated solution for improving Cybersecurity in web-based environments.
Keywords: *Data Preprocessing; Malicious Website Detection; Phishing URL Detection; Random Forest Classifier; Web Security.*

1. Introduction
In recent years, the use of machine learning (ML) techniques in Cyber security has increased. One of the best ways to protect against zero-day attacks is to use machine learning to categorize IP traffic and distinguish malicious traffic for intrusion detection. New research is being carried outing measurable traffic characteristics and machine learning techniques. Phishing is an online thievery that steals an individual's identity and private data. It is a form of extortion in which the perpetrator gains complete access to the private information of another individual. By posing as a legitimate website, for instance, these phishing websites convince users to provide their account information, by utilizing HTTPS. That induces a user to rely on the fictitious website. They promise safety and privacy, but they get the user's identity information. The majority of money transfers take place online. Everything, from paying bills to transferring funds, is done through websites or apps. Consequently, identifying such a fake website is crucial. Phishing is a fraudulent method that obtains customer identification information and financial credentials through the use of technological and social cunning. Social media platforms use spoof emails from real businesses and organizations to give users access to bogus websites where they can reveal financial information like usernames and passwords. Hackers often use tools to intercept usernames and passwords from users' online accounts when they want to steal credentials from computers. Phishers can steal user information using a variety of channels, such as email, Uniform Resource Locators (URL), instant messages, forum postings, phone calls, and text messages. Phishing content impersonates legitimate content in structure and tempts users to access it in order to obtain their

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3278-3281
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0419

sensitive data. Phishing main goal is to obtain specific personal information for financial gain or identity theft. [1-5]

1.1.Types of Attacks

1.1.1. Cross-Site Scripting (XSS)

Cross-site scripting (XSS) is a type of web application attack that involves injecting malicious scripts into web pages that are viewed by other users. This is typically accomplished by injecting the script into a form input field or URL parameter that is then stored in the web application's database. When another user views the page that contains the malicious script, the script is executed in their browser, allowing the attacker to steal data or perform other malicious actions on the user's behalf.

1.1.2. Cross-Site Request Forgery (CSRF)

Cross-site request forgery (CSRF) is a type of web application attack that tricks a user into executing an unwanted action on a web application that they are already authenticated with. This is typically accomplished by sending a specially crafted link or script to the user, which then performs the unwanted action when clicked.

1.1.3. XML External Entity (XXE)

XML External Entity (XXE) is a type of web application attack that involves exploiting vulnerabilities in XML parsers used by a web application. This can allow an attacker to read sensitive data or execute unauthorized actions on the web application's server. XXE attacks typically involve injecting specially crafted XML payloads that exploit the XML parser's ability to read external entities. XXE attacks can be prevented by disabling external entity parsing or using secure XML parsers that properly sanitize input data.

1.1.4. Injection Attacks

Injection attacks involve inserting malicious code into a web application, typically in the form of input data such as SQL queries, commands, or scripts. Injection attacks are successful when an application fails to properly validate and sanitize input data. These attacks can be prevented by properly validating and sanitizing input data and using parameterized queries to access databases.

1.1.5. Fuzz Testing (Fuzzing)

Fuzz testing, also known as fuzzing is a technique used to discover vulnerabilities in a web application by sending it random or invalid input data. The goal of fuzz testing is to identify how the web application responds to different inputs and to find errors and crashes. [6-10]

1.2.Methods for Attacks

- **Web Application Firewalls (WAFs):** WAFs are designed to monitor and filter HTTP/HTTPS traffic between a web application and the internet. They analyze incoming traffic and can detect and block common web attacks like SQL injection, cross-site scripting (XSS), and more.
- **Log Analysis:** Analyzing web server logs and application logs can help identify suspicious activities or patterns indicative of an attack. Look for anomalies in user behavior, unusual HTTP requests, or error messages that might signal an ongoing attack.
- **Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS):** These systems monitor network or system activities for malicious activities or policy violations. They can detect and respond to attacks by analyzing traffic patterns, signatures, and behavior to identify potential threats.
- **Behavioral Analytics:** Monitoring the behavior of users and applications can help detect anomalies. This involves establishing a baseline of normal behavior and flagging any deviations from that norm, which might indicate an attack.
- **Machine Learning:** Detecting web application attacks often involves employing various machine learning (ML) methods that can analyze patterns, anomalies, and behaviors within web traffic data.

2. Methodology

The proposed system employs a deep learning–based approach for effective detection of phishing URLs. Initially, a dataset containing both legitimate and phishing URLs is collected and preprocessed. URL-based features such as length, token patterns, special characters, domain information, and sequential structures are extracted and encoded to make them suitable for deep learning analysis. A Long Short-Term Memory (LSTM) network is then trained using

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3278-3281
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0419

the labeled URL dataset. The LSTM model is chosen for its ability to capture sequential dependencies and patterns within URLs, enabling accurate differentiation between legitimate and phishing links. During training, the model learns common characteristics associated with fraudulent URLs and adapts to variations in phishing strategies.

2.1.Modules

2.1.1. Admin Module

The Admin Module serves as the backbone of the Phishing URL Detection System, allowing administrators to control, configure, and maintain the system efficiently. Administrators are responsible for managing user accounts, overseeing system activity, and ensuring that phishing detection models remain accurate and up-to-date. This centralized control ensures that the system operates reliably and securely, preventing unauthorized access and maintaining high system integrity. The module allows administrators to upload datasets containing labeled phishing and legitimate URLs. These datasets provide the information necessary to train machine learning models. Along with uploading, administrators perform preprocessing tasks such as cleaning missing values, normalizing URL features, extracting domain details, and converting raw data into a format suitable for training. Proper preprocessing ensures that the model receives accurate, structured data to learn patterns effectively Finally, the Model Build sub-module trains a Random Forest classifier using the preprocessed datasets. Random Forest, an ensemble of decision trees, is chosen for its high accuracy and ability to handle complex URL features without over fitting. Once trained, the model is deployed for real-time classification, enabling the system to analyze user-submitted URLs and detect phishing attempts efficiently.

2.1.2. User Module

The User Module provides a simple, intuitive interface for users to check URLs safely. It ensures that even non-technical users can benefit from phishing detection without requiring deep knowledge of Cybersecurity or machine learning. The module supports both individual users and organizational deployments, enhancing online safety across platforms. Users first register by providing basic information such as username, email, and password. The system validates this information and stores it securely, ensuring that user data is protected. After registration, users log in using their credentials to access the URL verification system, maintaining a secure and personalized environment for URL analysis. Once logged in, users input URLs they want to check. The system extracts URL features such as domain reputation, SSL certificate validity, presence of suspicious characters, URL length, and content behavior. The Result sub-module then analyzes the URL using the trained Random Forest model and provides real-time feedback, classifying the URL as safe or phishing. The report may also include risk indicators and reasons for the classification, helping users understand threats and make informed decisions while browsing.

3. Results

The model was able to reduce false predictions compared to other methods. The proposed model using Random Forest showed high accuracy in detecting phishing websites. The model helps improve online security by protecting users from phishing attacks. It can successfully identify malicious websites that look similar to real ones. Figure 1 shows System Architecture

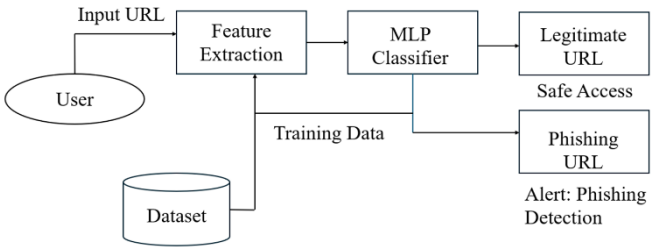

Figure 1 System Architecture

Conclusion

The phishing threat has emerged among the most common threats to internet users, organizations, and network operators. Using faked emails or false pages, the attackers acquires the client's confidential data in a phishing assault. Phishing scams, which include multiple hoaxes on the websites, are frequent entrance sites for online social engineering attacks. Phishing websites appear genuine, but they are

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3278-3281
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0419

difficult to spot because attackers mimic the form and function of legitimate websites. This project presents an intelligent model for an efficient phishing detection protocol. It utilizes a Random Forest after selecting the highest correlated features from the dataset hidden layers. The proposed approach's performance is evaluated using various evaluation metrics such as specificity, accuracy, and sensitivity. The exploratory experiments demonstrated that the proposed method outperforms existing machine learning and neural network classifiers for detecting malicious websites.

References

[1]. Remya, S., et al. "BGL-PhishNet: Phishing Website Detection Using Hybrid Model-BERT, GNN, and LightGBM." IEEE Access 13 (2025): 47552-47569.

[2]. Nayak, Ganesh S., Balachandra Muniyal, and Manjula C. Belavagi. "Enhancing phishing detection: a machine learning approach with feature selection and deep learning models." IEEE Access (2025)

[3]. Li, Wenhao, et al. "A state-of-the-art review on phishing website detection techniques." IEEE Access (2024)

[4]. Pillai, Manu J., et al. "Evasion attacks and defense mechanisms for machine learning-based web phishing classifiers." IEEE Access 12 (2023): 19375-19387.

[5]. Lee, Jehyun, et al. "Multimodal large language models for phishing webpage detection and identification." 2024 APWG Symposium on Electronic Crime Research (eCrime). IEEE Access (2024).

[6]. Putra, Fauzan Prasetyo Eka, et al. "Analysis of phishing attack trends, impacts and prevention methods: Literature study." Brilliance: Research of Artificial Intelligence 4.1 (2024): 413-421.

[7]. Ejaz, Asif, Adnan Noor Mian, and Sanaullah Manzoor. "Life-long phishing attack detection using continual learning." Scientific reports 13.1 (2023): 11488.

[8]. Butt, Umer Ahmed, et al. "Cloud-based email phishing attack using machine and deep learning algorithm." Complex & Intelligent Systems 9.3 (2023): 3043-3070.

[9]. Wang, Yanbin, et al. "A lightweight multi-view learning approach for phishing attack detection using transformer with mixture of experts." Applied Sciences 13.13 (2023): 7429.

[10]. Das, Sanchari, Christena Nippert-Eng, and L. Jean Camp. "Evaluating user susceptibility to phishing attacks." Information & Computer Security 30.1 (2022): 1-18.