www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2688-2692

e-ISSN : 2583-1062
Impact Factor : 7.001

URL PHISHING DETECTION USING DIFFERENT DEEP LEARNING ALGORITHMS

Mirza Azeem Baig[*1], Mohammed Kamran Shareef[*2], Syed Ammar Ali[*3], Unnati Khanapurkar[*4]

[*1,2,3]UG students, Dept. of CSE, MCET, Hyderabad, India.

[*4]Asst. Prof., Dept. of CSE, MCET, Hyderabad, India.

DOI: https://www.doi.org/10.58257/IJPREMS40039

ABSTRACT

Phishing is a common cyberattack that takes advantage of human trust by tricking users into revealing sensitive information using fraudulent URLs. Traditional detection approaches tend to be inadequate in detecting advanced phishing attacks, and require more intelligent and adaptive measures. This work explores the use of deep learning models—namely Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks—for phishing URL detection. A publicly available dataset is employed to train and test each model according to performance measures such as accuracy, precision, recall, and F1-score. The results show that deep learning models have high accuracy in detecting malicious URLs and can provide a scalable real-time defense solution against phishing attacks. This study adds to the research on AI-based cybersecurity solutions for countering online threats.

Keywords: CNN, RNN, LSTM, URL Phishing Detection, Deep Learning, Cybersecurity, Accuracy, Precision, Recall, F1-Score.

1. INTRODUCTION

Today, in an increasingly digital world, the internet is a part of daily life and has made communication, trade, and information sharing more efficient. However, the increase in dependency on internet-based services has exposed its users to cyber threats, one of which is the phishing attack, the most common and deceptive form of cybercrime. Phishing attacks utilize harmful URLs (Uniform Resource Locators) to deceive individuals into disclosing confidential information, including login details, financial information, or personal data. These attacks take advantage of human fallibility and trust, frequently circumventing conventional detection methods by employing advanced strategies such as URL obfuscation and dynamic content creation. As phishing techniques become increasingly advanced, there is a critical need for effective real-time detection methods.

Machine learning approaches, including deep neural networks, random forests, and support vector machines, have demonstrated promising results in classifying phishing attempts. To mitigate the risk, deep learning progress has allowed the development of complex systems that are extremely accurate in identifying phishing URLs. Unlike traditional rule-based or blacklist-based techniques, deep learning models can observe very complex patterns and characteristics drawn from URLs, thus proving to be highly effective when dealing with dynamic features of phishing threats. Phishing attacks pose a significant threat to online security, with attackers employing sophisticated tactics across email, social media, and instant messaging to deceive users.

This project is based on evaluation of the performance of three different deep learning architectures, CNN, RNN, and LSTM networks. The proposed architecture will test the ability of these different architectures to identify phishing URLs while using a dataset of legitimate and malicious URLs, with their corresponding relevant features used for training and evaluation. The merits and demerits of each algorithm are judged using critical performance indicators such as accuracy, precision, recall, and F1 score.

The subject of this project is to work on implementing various deep learning architectures like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks to identify phishing URLs. All of these architectures have their strengths in dealing with various data representations and patterns. The aim is to develop a system that can intelligently label a given URL as phishing or legitimate in real-time, thus reducing the threat of data theft and enhancing overall internet safety. To measure the performance of each model, metrics like accuracy, precision, recall, and F1-score are analyzed. These measures indicate how well each model performs in identifying phishing attempts correctly while keeping false positives and false negatives to a minimum, finally determining the most efficient model to deploy

The system is built using Python and TensorFlow so that it is scalable and compatible with cybersecurity frameworks. After the models are ready, they can then be readied for deployment on real-time systems. This would entail

transforming the trained model to a web-friendly format and embedding it on applications like browser extensions, email filters, or security gateways. A human-readable interface or API can also be created to provide real-time URL analysis, warning users to warn them before accessing potentially malicious links. RESTful API services are implemented with frameworks such as Flask to expose the model to other applications.

This research hopes to offer pragmatic insights to organizations striving to strengthen their defenses against phishing threats by identifying the most efficient algorithm for phishing detection. The findings emphasize the need for continued research to develop adaptive, scalable, and efficient solutions to combat evolving phishing threats. Furthermore, the integration of advanced AI-driven approaches has the potential to revolutionize the field by enabling proactive threat, and phishing detection and mitigation.

2. LITERATURE REVIEW

[1] The paper presents a hybrid security framework that combines AI-assisted phishing detection with generative AI-powered end-user education, achieving an accuracy of above 94%. The framework provides a comprehensive approach to phishing attack detection and end-user education, enhancing cybersecurity and providing effective defense strategies. The paper highlights the importance of customizable education that is adaptable to business models and assesses security posture. The framework uses a deep learning model to detect phishing attacks and provides a user-friendly interface for end-users to report suspicious emails and websites. The paper evaluates the performance of the framework using accuracy, precision, and recall metrics, and concludes by highlighting the importance of continued research and development in phishing attack detection and end-user education.

[2] The paper provides a comprehensive review of phishing attacks, prevention, and response strategies, aiming to provide a comprehensive understanding of phishing attacks and effective defense strategies. The paper highlights the importance of a comprehensive approach to phishing attack prevention and response, including education, awareness, and technological solutions. The paper discusses the importance of end-user education in preventing phishing attacks and the need for customizable education that is adaptable to business models. The paper also emphasizes the importance of incident response planning and continuous monitoring and updating of security systems to ensure they remain effective in detecting new and evolving phishing attacks. The paper concludes by highlighting the importance of continued research and development in phishing attack detection and prevention, and the need for businesses to invest in effective cybersecurity solutions.

[3] The paper presents a study on improving phishing website detection using feature selection and ensemble learning, achieving an accuracy of 96% and 98% on two different datasets. The paper highlights the importance of feature selection and ensemble learning in improving detection accuracy, robustness to noise and outliers, and handling high-dimensional data. The paper provides a comprehensive approach to phishing website detection, discussing the importance of machine learning-based systems and the use of feature selection and ensemble learning techniques. The paper evaluates the performance of the proposed system using accuracy, precision, and recall metrics, and concludes by highlighting the importance of continued research and development in phishing website detection. The paper also emphasizes the need for businesses to have a comprehensive approach to phishing attack detection, including feature selection, ensemble learning, and machine learning-based systems.

[4] The paper presents a deep neural network called WebPhish for phishing website detection, achieving an accuracy of 98.1% using a deep neural network with an embedding layer and convolutional layers. The paper highlights the importance of learning complex patterns in phishing websites and achieving high accuracy and provides a comprehensive approach to phishing website detection. The paper discusses the importance of deep learning-based systems in phishing website detection and the use of CNNs. The paper evaluates the performance of the proposed system using accuracy, precision, and recall metrics, and concludes by highlighting the importance of continued research and development in phishing website detection. The paper also emphasizes the need for businesses to have a comprehensive approach to phishing attack detection, including deep learning and machine learning-based systems.

[5] The paper provides a comprehensive review of artificial intelligence in cybersecurity, aiming to provide a comprehensive understanding of AI-powered defense strategies and enhancing cybersecurity. The paper discusses the importance of AI in cybersecurity, including phishing attack detection, malware detection, and incident response. The paper highlights the importance of a comprehensive approach to cybersecurity, including AI-powered defense strategies, machine learning-based systems, and deep learning-based systems.

[6] The paper provides a comprehensive overview of AI techniques, including machine learning, deep learning, and natural language processing, and discusses their applications in cybersecurity. The paper concludes by highlighting the importance of continued research and development in AI in cybersecurity, and the need for businesses to invest in effective cybersecurity solutions.

[7] This research article creates a phishing detection system based on machine learning (CNN, RNN, LSTM) and ensemble techniques (RF, XGBoost) on 11,055 websites. Feature selection determined the most important indicators such as SSL status and URL patterns. The Random Forest model obtained 99% accuracy, surpassing other methods, exhibiting better ability to identify phishing sites while being capable of adapting to new threats. The framework provides an efficient real-time cybersecurity solution.

[8] The paper evaluates the performance of phishing detection systems using accuracy, precision, and recall metrics, and concludes by highlighting the importance of continued research and development in deep learning techniques for phishing detection. The paper also emphasizes the need for businesses to have a comprehensive approach to phishing attack detection, including deep learning-based systems and machine learning-based systems.

[9] The paper presents a real-time phishing URL detection system based on lexical and host-based feature extraction using a statistical machine learning classifier. This system does not use high-dimensional feature spaces and thus achieves robust performance with low error rates in detecting phishing URLs. Thorough experiments demonstrate its suitability for real-world applications in terms of high accuracy and fast processing. The system is language-independent and scalable and can detect phishing sites without relying on web content. It significantly contributes toward security through its novel approach in the analysis of phishing URL characteristics toward real-time detection.

[10] This paper provides a sound, automated framework for detecting phishing URLs using online classifiers with learning adaptation towards the dynamic growth of the space of URLs. It relies on a mechanism of postponed feature collection: it first prefers cheap features such as lexical attributes for rapid classification. Selective sampling yields 87% accuracy, and real-time adaptability; it strongly focuses on scalable and efficient performance. This approach faces problems, such as an unbalanced dataset and computationally expensive procedures. This study fills the lacuna in existing methodologies through an all-inclusive, real-time mechanism for detecting harmful URLs that enhances online security.

Table -1: Taxonomy of surveyed methodologies

Authors	Title	Research focus	Remarks
Shougfta Mushtaq, Tabassum Javed, Mazliham Mohd Su'ud, 2024 [1]	Ensemble Learning-Powered URL Phishing Detection	Feature selection with ensemble classifiers / Up to 98%	High accuracy, minimal features, struggles with dynamic techniques.
FNU Jimmy, 2024 [2]	Phishing Attackers: Prevention and Response Strategies	Preventive measures and response frameworks / N/A	Comprehensive but lacks validation.
Akshaya Arun,Nasr Abosata, 2024 [3]	Next Generation of Phishing Attacks Using AI-Powered Browsers	Real-time ML-powered browser extension / 99.11%	Effective for zero-day attacks, limited scalability.
Zainab Alshingiti, Rabeah Alaqel, Jalal Al-Muhtadi, 2023 [4]	A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN	LSTM-CNN hybrid model / CNN: 99.2%, LSTM-CNN: 97.6%, LSTM: 96.8%	High accuracy, avoids manual feature engineering, needs large datasets.
Kavitha Dhanushkodi,S. Thejas, 2024 [5]	AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation	AI-driven security models for IoT, 5G / 95%+	Addresses zero- day threats; emphasizes scalability and explainability.
Chidimma Opara, Yingke Chen, Bo.Wei, 2023 [6]	Look Before You Leap: Detecting Phishing Web Pages	Deep neural networks with URL and HTML embeddings / 98.1%	Working with raw data requires high resources.

Table -2: Taxonomy of surveyed methodologies

Authors	Title	Research focus	Remarks
Ume Zara, Kashif Ayyub, Hikmat Ullah Khan, Ali Daud, Tariq Alsafi, Saima Gulzar Ahmad,2024 [7]	Phishing Website Detection Using Deep Learning Models	Developing an advanced phishing detection system using machine learning and ensemble methods with feature optimization techniques	Provides an effective real-time cybersecurity solution with adaptability to emerging phishing threats.
Maria Sameen, Kyunghyun Han, Seong Oun Hwang, 2020 [8]	PhishHaven: AI Phishing URLs Detection System	Ensemble learning with lexical analysis / 98%	Real-time detection is resource- intensive.
Jianyi Zhang, Yonghao Wang, 2012 [9]	A Real-time Automatic Detection of Phishing URLs	Statistical machine learning classifier with new feature extraction methods / Accuracy above 93%	High accuracy in real- time scenarios; avoids traditional feature extraction ; requires balanced datasets
Farhan Sadique, Raghav Kaul, Shahriar Badsha, Shamik Sengupta, 2020 [10]	An Automated Framework for Real- time Phishing URL Detection	Online learning and selective sampling with delayed feature collection / 87% accuracy	Suitable for dynamic environments; effective framework but requires more features and tuning.

3. CONCLUSION

In this study, we explored the implementation of deep learning algorithms—CNN, RNN, and LSTM—for detecting phishing URLs, a critical challenge in the field of cybersecurity. Through extensive data preprocessing, model training, and evaluation, each model demonstrated unique strengths in classifying malicious and legitimate URLs based on their structural patterns. Among the models tested, the RNN achieved the highest prediction accuracy and robustness in handling sequential URL data, while CNN and LSTM are the best algorithms since they capture sequential patterns and spatial features, respectively.The performance was validated using metrics such as accuracy, precision, recall, and F1-score, confirming the potential of deep learning approaches in real-time phishing detection. The use of an API for deployment further supports practical applications in web browsers, email filters, or mobile apps. Future work may focus on enhancing model generalization to detect zero-day phishing attacks and improving response time for real-world scalability.

4. REFERENCES

[1] S. Mushtaq, T. Javed, and M. Mazliham Mohd Su'ud, "Ensemble Learning- Powered URL Phishing Detection: A Performance Driven Approach," Journal of Informatics and Web Engineering, vol. 3, no. 2, pp. 135–142, Jun. 2024.

[2] FNU Jimmy, "Phishing attackers: prevention and response strategies," Journal of Artificial Intelligence General Science (JAIGS), vol. 2, no. 1, pp. 308–316, Feb. 2024.

[3] A. Arun and N. Abosata, "Next Generation of Phishing Attacks using AI-powered Browsers," IEEE Xplore.

[4] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM- CNN," Electronics, vol. 12, no. 232, 2023.

[5] K. Dhanushkodi and S. Thejas, "AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation," IEEE Access, vol. XX, 2024.

[6] C. Opara, Y. Chen, and Bo. Wei, "Look Before You Leap: Detecting Phishing Web Pages by Exploiting Raw URL And HTML Characteristics," Proceedings of IEEE ICCSNT, Sep. 2023.

[7] Ume Zara, Kashif Ayyub, Hikmat Ullah Khan, Ali Daud, Tariq Alsahfi, Saima Gulzar Ahmad, "Phishing Website Detection Using Deep Learning Models," IEEE Access, vol. 12, no. 2169-3536, 25 October 2024.

[8] Maria Sameen, Kyunghyun Han, Seong Oun Hwang, "PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System," IEEE Access, vol. 8, pp. 83425-83443, 2020.

[9] J. Zhang and Y. Wang, "A Real-time Automatic Detection of Phishing URLs," in Proc. 2012 International Conference on Communication, Signal Processing, and Networking (ICCSNT), 2012, pp. 1-5.

[10] Farhan Sadique, R. Kaul, S. Badsha, and S. Sengupta, "An Automated Framework for Real-time Phishing URL Detection," Dept. of Computer Science and Engineering, University of Nevada, Reno, NV, USA, 2020.