International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Enhancing Phishing Detection A Novel Hybrid Deep Learning Framework For Cybercrime Forensics

[1]N Saibabu,[2]Chepuri Venkatesh

[1]M. Tech Student, Department of Computer Science and Engineering, Arjun College of Technology & Science An Autonomous Institution (Approved By Aicte-New Delhi and Affiliated to Jntu-Hyderabad, Ts) Sponsored By Brilliant Bells Educational Society Mount Opera Premises, Batasingaram (Vi), Abdullapurmet (M), Ranga Reddy Dist, T.S.501512

[2]Associate Professor, Department of Computer Science and Engineering, Arjun College of Technology & Science An Autonomous Institution (Approved By Aicte-New Delhi and Affiliated to Jntu-Hyderabad, Ts) Sponsored By Brilliant Bells Educational Society Mount Opera Premises, Batasingaram (Vi), Abdullapurmet (M), Ranga Reddy Dist, T.S.501512

ABSTRACT

Phishing attacks have emerged as a significant threat in the digital realm, exploiting both social engineering and technological vulnerabilities to deceive users into divulging sensitive information. Traditional detection methods often fall short in identifying sophisticated phishing attempts, necessitating advanced solutions. This paper introduces a novel hybrid deep learning framework that synergizes Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), and Multi-Layer Perceptron (MLP) algorithms to enhance phishing detection capabilities. The system architecture comprises two primary modules: Admin and User. The Admin module facilitates the training and deployment of machine learning models, while the User module allows for the monitoring of blocked URLs. By integrating these components, the framework aims to bolster cybercrime forensics and provide a robust defense against phishing attacks.

Keywords: Phishing Detection, Deep Learning, Cybercrime Forensics, Hybrid Framework.

I. INTRODUCTION

In today's digital era, the internet serves as a powerful platform for communication, business, and information exchange. However, the proliferation of online services has also brought forth various security challenges, with phishing being one of the most prevalent and damaging forms of cybercrime. Phishing attacks manipulate users into disclosing confidential data such as login credentials, credit card numbers, and personal information, leading to financial and reputational damage.

Traditional phishing detection techniques rely heavily on blacklisting or heuristic rules, which are insufficient in detecting novel or zero-day phishing threats. These approaches are reactive and often lag behind the rapidly evolving phishing landscape. This necessitates a more robust and adaptive mechanism for threat detection.

Machine learning (ML) and deep learning (DL) have shown immense promise in enhancing cybersecurity measures. By learning from data, these algorithms can detect patterns and anomalies that may not be

apparent to human analysts. Among the many ML techniques, Support Vector Machines, LightGBM, and Multi-Layer Perceptron stand out for their classification accuracy and speed.

This project proposes a hybrid framework combining SVM, LightGBM, and MLP for phishing detection. The framework is divided into two modules: Admin and User. The Admin module is responsible for training the models and testing URLs, while the User module allows viewing of blocked URLs.

The hybrid approach enables the system to benefit from the strengths of each algorithm. SVM offers high accuracy with smaller datasets, LightGBM provides efficiency and scalability, and MLP enables deep feature learning. Together, they form a resilient detection mechanism.

By implementing this architecture, we aim to improve the early detection of phishing threats and support cybersecurity professionals in forensics investigations. This system is particularly useful in real-time URL analysis and educational institutions, banking sectors, and enterprises seeking to protect their digital infrastructure.

II. LITERATURE SURVEY

A comprehensive literature survey was conducted to understand the current state of phishing detection using machine learning and deep learning algorithms. Ten significant research works and publications were reviewed, analyzing their methodologies, advantages, and limitations.

Detecting of URL Based Phishing Attack Using Machine Learning

AUTHORS: Ms. Sophiya Shikalgar, Mrs. Swati Narwane (2019)

ABSTRACT:

Phishing has become a major security concern on the internet, where attackers impersonate trusted websites to deceive users into divulging sensitive information. This paper presents an approach to detect phishing attacks based on analyzing the URLs using machine learning algorithms. The study focuses on extracting features from the URLs and training a model that can distinguish between legitimate and phishing URLs. Various machine learning algorithms are evaluated for their performance, including Random Forest, SVM, and Decision Trees. The authors emphasize the importance of selecting relevant features and using labeled datasets to enhance model accuracy. Results demonstrate that URL-based detection using ML can be an effective preventive mechanism against phishing attacks. However, the study also acknowledges the limitations of relying solely on URL features and suggests integrating other features like webpage content in future work.

Support Vector Machine Based Malware and Phishing Website Detection

AUTHORS: Rashmi Karnik, Dr. Gayathri M Bhandari

ABSTRACT:

This research addresses the growing challenge of detecting malware and phishing websites by leveraging the capabilities of Support Vector Machine (SVM) algorithms. The authors propose an efficient classification model that learns from URL patterns and domain-level data to distinguish between malicious and legitimate websites. The study introduces a structured dataset comprising various URL features such as length, special characters, and presence of IP addresses. The SVM model is trained and tested on this dataset, showing promising accuracy and low false-positive rates. Moreover, the paper discusses comparative results with other machine learning techniques, highlighting the superiority of SVM in identifying complex phishing patterns. The results suggest that SVM can effectively generalize from historical data to detect emerging threats. This model can be integrated into real-time web filters or browser extensions to enhance user protection.

Phishing Websites Detection Using Machine Learning

AUTHORS: Arun Kulkarni, Leonard L. Brown, III2

ABSTRACT:

The paper explores the potential of machine learning algorithms in combating phishing attacks by classifying suspicious websites based on extracted features. The authors create a dataset from known phishing and legitimate websites and apply several classification algorithms to detect malicious sites. These include Decision Tree, Naive Bayes, and Random Forest classifiers. The approach primarily involves feature engineering from URL characteristics, webpage metadata, and DNS information. The model is trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Results indicate that machine learning methods can automate the detection of phishing with significant accuracy, reducing reliance on manually maintained blacklists. Furthermore, the paper highlights the importance of updating training data regularly to adapt to evolving phishing tactics.

Phishing Websites Detection Using Machine Learning

AUTHORS: R. Kiruthiga, D. Akila

ABSTRACT:

This study presents an intelligent system for detecting phishing websites using a machine learning approach. The authors focus on improving detection accuracy through the use of extensive feature extraction and data preprocessing techniques. A range of algorithms is considered, including Decision Trees, Logistic Regression, and K-Nearest Neighbors, which are tested on a dataset comprising real-time URL data. The model evaluates features such as URL length, the use of HTTPS, subdomain count, and redirection chains. The results highlight the Decision Tree algorithm as the most effective, providing high classification accuracy and interpretability. The research also includes a discussion on the importance of feature selection, balancing datasets, and managing overfitting. The paper concludes by suggesting the deployment of such models in browser security tools or firewalls.

Hybrid Machine Learning: A Tool to Detect Phishing Attacks in Communication Networks

AUTHORS: Ademola Philip Abidoye, Boniface Kabaso

ABSTRACT:

This paper introduces a hybrid machine learning framework to improve phishing detection in communication networks. Unlike single-algorithm models, the proposed system combines multiple learning methods to increase classification accuracy and reduce false alarms. The authors employ ensemble techniques using Random Forests and Boosted Trees, alongside traditional classifiers like SVM and Naive Bayes. The data used for training includes URLs, email content, and website metadata. The hybrid approach is designed to capture both shallow patterns and deep relationships in the data, enabling it to detect sophisticated phishing attempts. Experimental results show that the hybrid model outperforms individual classifiers in both accuracy and robustness. The authors argue that such systems are better suited for real-world deployment, where attackers constantly adapt to evade detection.

Towards Lightweight URL-Based Phishing Detection

AUTHORS: Andrei Butnaru, Alexios Mylonas, Nikolaos Pitropakis

ABSTRACT:

This paper addresses the computational limitations of existing phishing detection models and proposes a lightweight solution based on URL features. The focus is on designing a system that can be deployed on low-resource devices without compromising detection accuracy. The authors extract

minimal but highly informative features from URLs and apply lightweight classifiers like Logistic Regression and Decision Trees. The study evaluates the trade-off between model complexity and detection performance, showing that even lightweight models can achieve competitive accuracy. Additionally, the paper discusses the deployment potential in mobile browsers and IoT environments. This approach supports real-time detection with minimal latency and is ideal for end-users with constrained computational resources.

Detecting Phishing Websites Using Machine Learning Technique

AUTHORS: Ashit Kumar Dutta (2021)
ABSTRACT:
The paper investigates phishing detection using supervised machine learning methods applied to a comprehensive dataset of phishing and legitimate URLs. The study emphasizes the importance of choosing relevant features from URLs, domain metadata, and SSL certificates. Various classifiers including Decision Tree, SVM, and Random Forest are tested for effectiveness. The model with the highest accuracy is then selected for deployment. Results show that using domain-based and lexical features enhances the model's predictive capability. The author also explores how phishing patterns evolve over time and how the model's accuracy changes accordingly. A dynamic training strategy is proposed to update the model regularly, allowing it to stay effective against new phishing tactics.

Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical Study

AUTHORS: Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Takeru Yokoi, Hamido Fujita
ABSTRACT:
This study explores the use of deep learning algorithms for phishing webpage classification. The authors design and train neural networks using large-scale datasets that include both phishing and legitimate websites. Features include not only URLs but also HTML content, JavaScript code, and metadata. The proposed deep learning model, particularly using CNNs and LSTMs, is shown to be more resilient to obfuscated phishing techniques. Unlike traditional ML methods, these models automatically learn feature representations without manual feature extraction. The paper discusses performance metrics, training efficiency, and real-time deployment challenges. The results reveal that deep learning models achieve higher accuracy and lower false detection rates compared to classical machine learning methods, making them suitable for scalable, cloud-based security applications.

Phishing Website Detection Using Diverse Machine Learning Algorithms

AUTHORS: Ammara Zamir, Hikmat Ullah Khan, Tassawar Iqbal
ABSTRACT:
The authors present a comparative study of different machine learning algorithms applied to phishing website detection. The dataset comprises features like URL structure, IP-based indicators, and security protocols. Algorithms such as SVM, KNN, Decision Tree, and AdaBoost are tested and analyzed based on performance metrics including accuracy, recall, and F1-score. The study shows that no single algorithm is best in all aspects and emphasizes the need to balance detection accuracy with computational efficiency. The authors suggest that adaptive systems using multiple classifiers can improve the robustness of phishing detection tools. This work contributes to understanding the strengths and weaknesses of various ML models in cybersecurity.

Phishing Detection Using Machine Learning Techniques

AUTHORS: Valid Shahrivari, Mohammad

Mahdi Darabi, Mohammad Izadi

ABSTRACT:
In this research, a machine learning-based framework is developed to detect phishing attacks by analyzing URLs and site behavior patterns. The authors construct a dataset with real-world examples of phishing and non-phishing websites, focusing on lexical, domain-based, and traffic-based features. Multiple algorithms including Random Forest, XGBoost, and Neural Networks are applied to identify which method yields the highest performance. The study also evaluates the model's ability to detect zero-day attacks. Results indicate that ensemble models and boosting methods outperform individual classifiers. The paper concludes with recommendations for implementing these models in commercial web browsers and email gateways to provide real-time protection against phishing threats.

III. EXISTING SYSTEM

Existing phishing detection systems are largely based on rule-based approaches, blacklisting, and content inspection. These systems struggle to detect new phishing techniques due to their static nature. Some use single machine learning models, which lack the robustness needed for complex attack detection.

IV. PROPOSED SYSTEM

The proposed system introduces a hybrid deep learning framework integrating SVM, LightGBM, and MLP models for phishing detection. It features an Admin module that manages model execution and URL testing, and a User module that views blocked URLs. This architecture combines the strengths of three algorithms, increasing detection precision and adaptability. The admin can train and test URLs in real-time, while users are kept informed about potential threats through a simple interface. This hybrid framework is designed to be scalable and adaptable, ensuring resilience against the evolving nature of phishing techniques.

V. SYSTEM ARCHITECTURE

The diagram illustrates the workflow of a Phishing Detection Hybrid Deep Learning Framework, showing the interaction between the user, administrator, and the machine learning models used for phishing URL detection. The framework is designed to provide secure access, efficient model execution, and accurate classification of suspicious websites.

Initially, the user accesses the system through the Login module. After successful authentication, the user's request is forwarded to the administrator. The administrator also logs into the system using secure credentials, ensuring that only authorized personnel can manage and operate the phishing detection framework. The administrator has the ability to monitor the system, initiate model execution, and log out after completing the required tasks.

Once authenticated, the administrator enters the Phishing Detection Hybrid Deep Learning Framework, where multiple machine learning and deep learning algorithms are executed sequentially. The process begins with the Support Vector Machine (SVM) model, which performs the initial classification of website features by identifying patterns that distinguish legitimate URLs from phishing URLs. SVM provides a strong baseline classification using its capability to create optimal decision boundaries.

After the SVM stage, the extracted features and intermediate outputs are processed using the Light Gradient Boosting Machine (LightGBM). LightGBM improves prediction performance by building efficient gradient-boosted decision trees. It handles large datasets effectively, reduces computational time, and enhances the overall detection accuracy by capturing complex relationships among phishing indicators.

The processed data is then passed to the Multi-Layer Perceptron (MLP), a deep neural network capable of learning nonlinear feature representations. The MLP further refines the classification by identifying hidden patterns that traditional machine learning algorithms may not capture. Combining SVM, LightGBM, and MLP creates a hybrid framework that improves robustness and minimizes false positives and false negatives.

Once model execution is complete, the framework proceeds to the Test URL module. Here, the user or administrator submits a website URL for analysis. The trained hybrid model evaluates the URL based on learned phishing characteristics and predicts whether the website is legitimate or phishing. The prediction result is displayed to the user, enabling quick and informed decision-making before accessing potentially harmful websites.

Finally, after testing and monitoring are completed, both the administrator and the framework provide a Logout option to securely terminate the session. This ensures system security by preventing unauthorized access and maintaining the confidentiality of phishing detection operations. Overall, the framework integrates secure authentication with hybrid machine learning and deep learning techniques to deliver an efficient, reliable, and accurate phishing detection system.

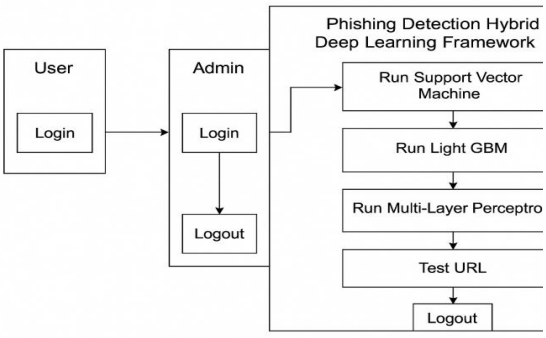

Fig 5.1: System Architecture

VI. IMPLEMENTATION

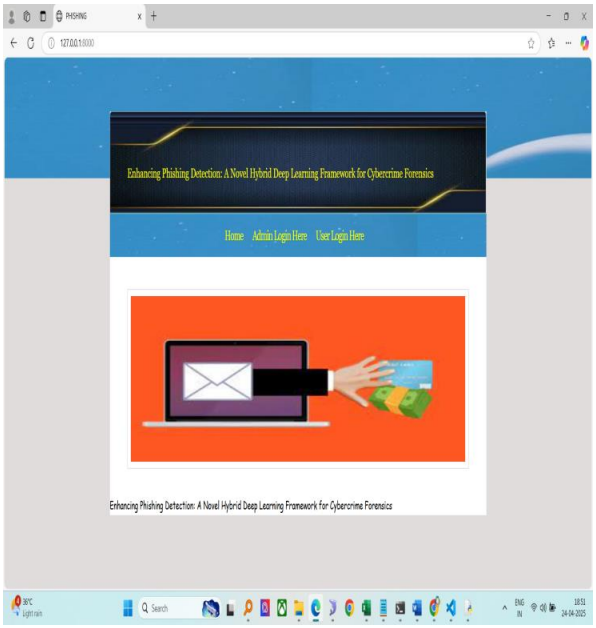

Fig 6.1: Home Page

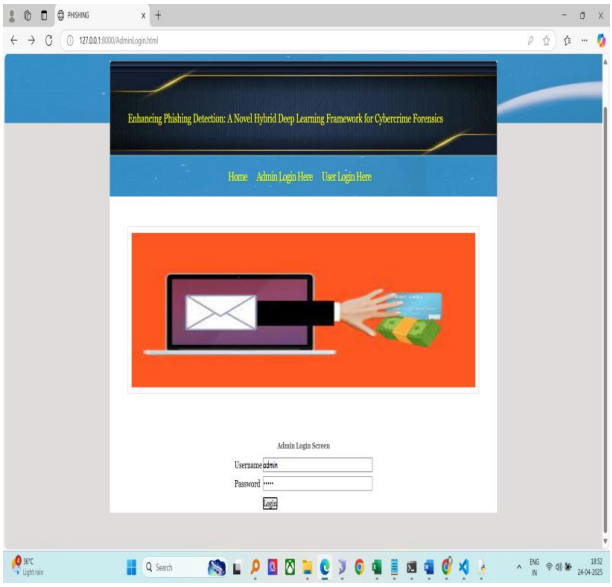

Fig 6.2: Admin Login

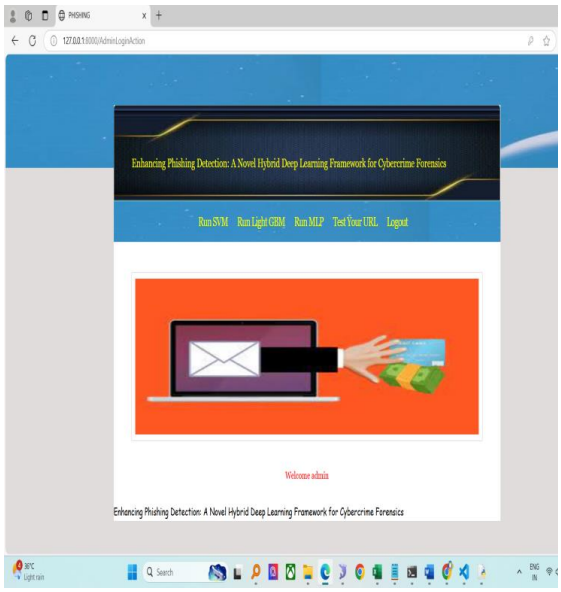

Fig 6.3: Admin Home

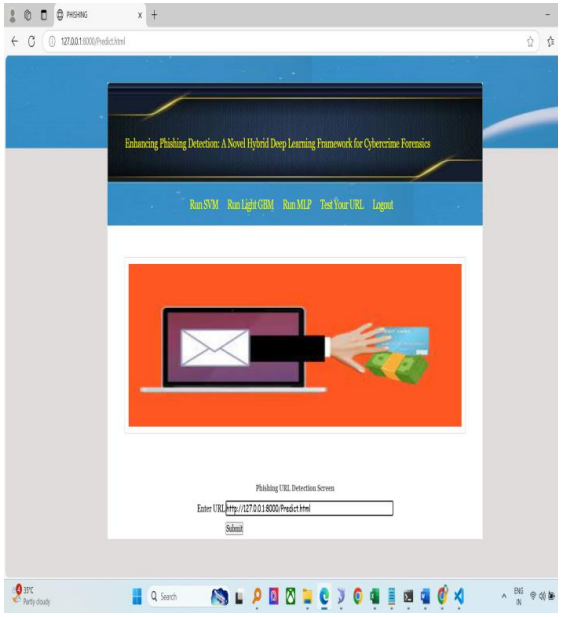

Fig 6.4: Prediction Input

Fig 6.5: Result Page

VII. CONCLUSION

This project presents a novel hybrid deep learning framework that enhances phishing detection using a combination of SVM, LightGBM, and MLP algorithms. By dividing responsibilities across Admin and User modules, it ensures efficient model management and user interaction. The system provides real-time phishing detection and supports cybercrime forensic efforts, making it a valuable tool in combating phishing threats.

VIII. FUTURE SCOPE

Future developments may include integrating Natural Language Processing (NLP) for detecting phishing content in emails and social media messages. Additionally, the framework can be expanded to mobile platforms and integrated with browser extensions for real-time phishing prevention. Federated learning approaches could also be adopted for privacy-preserving threat detection across multiple organizations.

IX. REFERENCES

[1] F. S. Alsubaei, A. A. Almazroi, and N. Ayub, "Enhancing Phishing Detection: A

Novel Hybrid Deep Learning Framework for Cybercrime Forensics," IEEE Access, vol. 12, pp. 8373–8389, 2024.
DOI: 10.1109/ACCESS.2024.3351946.

[2] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," IEEE Access, vol. 10, pp. 36429–36463, 2022.
DOI: 10.1109/ACCESS.2022.3151903.

[3] O. K. Sahingoz, E. Buber, and E. Kugu, "DEPHIDES: Deep Learning Based Phishing Detection System," IEEE Access, vol. 12, pp. 8052–8070, 2024.
DOI: 10.1109/ACCESS.2024.3352629.

[4] R. W. Purwanto, A. Pal, A. Blair, and S. Jha, "PhishSim: Aiding Phishing Website Detection With a Feature-Free Tool," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 1497–1512, 2022.
DOI: 10.1109/TIFS.2022.3164212.

[5] S. Al-Ahmadi, A. Alotaibi, and O. Alsaleh, "Phishing Detection With Generative Adversarial Networks," IEEE Access, vol. 10, pp. 42459–42468, 2022.
DOI: 10.1109/ACCESS.2022.3168235.

[6] Z. Dou, I. Khalil, A. Khreishah, A. Al-Fuqaha, and M. Guizani, "Systematization of Knowledge (SoK): A Systematic Review of Software-Based Web Phishing Detection," IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2797–2819, 2017.
DOI: 10.1109/COMST.2017.2752087.

[7] A. El Aassal, S. Baki, A. Das, and R. M. Verma, "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs," IEEE Access, vol. 8, pp. 22170–22192, 2020.
DOI: 10.1109/ACCESS.2020.2969780.

[8] S. Asiri, Y. Xiao, S. Alzahrani, and T. Li, "PhishingRTDS: A Real-Time Detection System for Phishing Attacks Using a Deep Learning Model," Computers & Security, vol. 141, Art. no. 103843, 2024.
DOI: 10.1016/j.cose.2024.103843.

[9] A. J. S. Albahadili, A. Akbas, and J. Rahebi, "Detection of Phishing URLs With Deep Learning Based on GAN-CNN-LSTM Network and Swarm Intelligence Algorithms," Signal, Image and Video Processing, vol. 18, no. 6, 2024.
DOI: 10.1007/s11760-024-03204-2.

[10] M. A. Alohali, N. Alasmari, M. Maashi, A. M. Nouri, M. Rizwanullah, I. Yaseen, A. E. Osman, and A. A. Alneil, "Metaheuristics With Deep Learning Driven Phishing Detection for Sustainable and Secure Environment," Sustainable Energy Technologies and Assessments, vol. 56, Art. no. 103114, 2023.
DOI: 10.1016/j.seta.2023.103114.

[11] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or Not Phishing? A Survey on the Detection of Phishing Websites," IEEE Access, vol. 11, pp. 18499–18519, 2023.
DOI: 10.1109/ACCESS.2023.3247135.

[12] W. Li, S. Manickam, S. U. A. Laghari, and Y.-W. Chong, "Uncovering the Cloak: A Systematic Review of Techniques Used to Conceal Phishing Websites," IEEE Access, vol. 11, pp. 71925–71939, 2023.
DOI: 10.1109/ACCESS.2023.3293063.

[13] E. Zhu, Z. Chen, J. Cui, and H. Zhong, "MOE/RF: A Novel Phishing Detection Model Based on Revised Multiobjective Evolution Optimization Algorithm and Random Forest," IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 4461–4478, 2022.
DOI: 10.1109/TNSM.2022.3162885.

[14] K. Barik, S. Misra, and R. Mohan, "Web-Based Phishing URL Detection Model Using Deep Learning Optimization Techniques," International Journal of Data

Science and Analytics, 2025.
DOI: 10.1007/s41060-025-00728-9.

[15] G. S. Nayak, B. Muniyal, and M. C. Belavagi, "Enhancing Phishing Detection: A Machine Learning Approach With Feature Selection and Deep Learning Models," IEEE Access, vol. 13, pp. 33308–33320, 2025.
DOI: 10.1109/ACCESS.2025.3543738