iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.70029

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

AI Driven Phishing Detection Model

Aishwarya D[1], Divyadharshini B[2], Jeslin J[3], Dr. GV. Shrichandran[4]
[1,2,3]*Student,* [4]*Assistant Professor, SRM Institute of Science and Technology, Chennai*

Abstract: ***Phishing attacks are a significant cybersecurity threat as they trick people into revealing personal information through fake websites. This project introduces an integrated CNN-LSTM model to detect phishing URLs. It uses Convolutional Neural Networks (CNNs) to look for local patterns and Long Short-Term Memory (LSTM) networks to analyze the order of information in URLs. To further clarify, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are implemented, giving insights into how the model predicts. The trained model is served as a FastAPI/Flask web service, enabling real-time URL analysis. A browser extension is also created to communicate with the API, facilitating on-the-fly phishing detection as users surf the web. The system offers predictions as well as explanations, enhancing user trust and security awareness. By integrating deep learning, explainability, and web deployment, this project provides a real-world and scalable cybersecurity solution, with potential for further improvements using Graph Neural Networks (GNNs), Reinforcement Learning, or Meta-Learning., or Math in Paper Title or Abstract.***

I. INTRODUCTION

With growing use of the internet by individuals and organizations for communications, transactions, and information exchange, cybersecurity attacks in the form of phishing have gained momentum. Phishing is a deception technique where the attackers design fake websites to get sensitive information like usernames, passwords, and financial information. Rule-based phishing detection techniques are typically outdated to counter the changing methods employed by cybercriminals. To address this problem, machine learning and deep learning methods have been extensively explored for phishing classification. In this project, we introduce a hybrid CNN-LSTM model to classify phishing and authentic URLs. CNNs (Convolutional Neural Networks) contribute to local pattern extraction in URLs, whereas LSTMs (Long Short-Term Memory networks) extract sequential dependencies, the model thus becomes stronger against adversarial attacks. In addition, explainability of the model is another essential component of AI-based cybersecurity. In order to offer improved explanations for phishing detection, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are used in explaining the model's decision. The users learn why the URL is identified as phishing, thus relying on the system. For deployment to the real world, the learned model is used within a FastAPI/Flask web application, enabling real-time URL checking. A browser extension is further implemented to talk to the API, giving end-users real-time phishing warnings upon surfing the internet.

II. EXISTING SYSTEM

Existing phishing detection systems primarily rely on traditional methods, such as blacklists and whitelists, which involve comparing URLs to a predefined list of known phishing sites. While these systems are quick and easy to implement, they are ineffective against novel phishing URLs and require constant updates, making them unsuitable for real-time detection. Heuristic-based methods, which analyse URL characteristics like domain name patterns and the presence of suspicious characters, have also been employed. However, these approaches lack flexibility and often fail to detect sophisticated phishing attempts that do not follow traditional patterns. Machine learning approaches, including Support Vector Machines (SVM), Decision Trees, and Random Forests, have been used to analyse features of URLs such as structure, keywords, and domain age. Although these models provide better accuracy, they suffer from overfitting and often require manual feature engineering, limiting their ability to generalize to new data. Deep learning methods, particularly Convolutional Neural Networks (CNNs), have shown promise by automatically learning URL features, but challenges remain in capturing sequential dependencies and context, which are critical for phishing detection. Furthermore, the lack of explainability in these models has raised concerns, as users may not trust decisions made by "black box" systems. Techniques like SHAP and LIME, which aim to provide explanations for model predictions, have been explored but are not yet fully integrated into phishing detection systems. Lastly, most existing systems do not offer real-time detection or browser integration, limiting their applicability in dynamic environments where phishing sites can emerge rapidly.

III. PROPOSED SYSTEM

The proposed system introduces a hybrid approach that combines Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to improve phishing URL detection. This hybrid CNN-LSTM model is designed to capture both local patterns within the URL structure (using CNN) and sequential dependencies (using LSTM), making it more robust and accurate than traditional methods. The system also integrates SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) to provide model explainability, allowing users to understand why a particular URL was flagged as phishing. This transparency is essential for trust and wider adoption, especially in real-world security applications.

The model is deployed as a Fast API/Flask web service , allowing users to send URLs and receive real-time predictions along with explanations of the decision making process. This API is designed to be scalable and can be hosted on cloud platforms such as AWS,GCP, or Heroku for public access. Additionally, a browser extension is developed to interact with the deployed API for real-time phishing detection. The extension communicates with the backend, sending the current webpage URL to the API and receiving predictions along with explanations, which are then displayed as warnings if the site is determined to be phishing. This end-to-end system not only enhances the accuracy and robustness of phishing detection but also ensures that users are provided with transparent and actionable insights, all while offering real-time protection against phishing attacks.

IV. RESULT&DISCUSSION

The proposed phishing detection system demonstrates high efficiency in both prediction accuracy and real-time usability. By leveraging a hybrid CNN-LSTM architecture, the model captures intricate patterns and sequential dependencies in URLs, resulting in more accurate classification compared to traditional machine learning methods. The integration of SHAP and LIME further enhances the system by providing transparent and interpretable results, which is crucial in cybersecurity applications. The use of FastAPI ensures minimal latency in API responses, making the system suitable for real-time applications. Additionally, the browser extension provides immediate feedback to users without noticeable delay, thereby enabling proactive protection against phishing threats. Overall, the system effectively balances deep learning performance, explainability , and real-time response , making it a practical and efficient solution for web security.

Traditional phishing detection systems often rely on rule-based techniques or conventional machine learning algorithms like decision trees, SVMs, or logistic regression. While these methods can perform reasonably well, they typically require manual feature engineering and struggle to generalize against evolving phishing tactics. Moreover, many lack real-time capabilities and fail to provide interpretability, making it difficult for users to understand why a website is flagged as malicious. In contrast, the proposed system uses a hybrid CNN-LSTM deep learning model that automatically learns complex patterns from URLs without manual intervention. It outperforms traditional methods in accuracy and adaptability, especially against sophisticated phishing attacks. Furthermore, by integrating SHAP and LIME, the system offers explainable predictions, addressing a major shortcoming of black-box models. The addition of a browser extension connected to a lightweight Fast API backend enables real-time phishing detection, which is rarely found in existing solutions.Overall, the proposed system provides a more intelligent, transparent, and practical approach to phishing prevention.

V. CONCLUSION

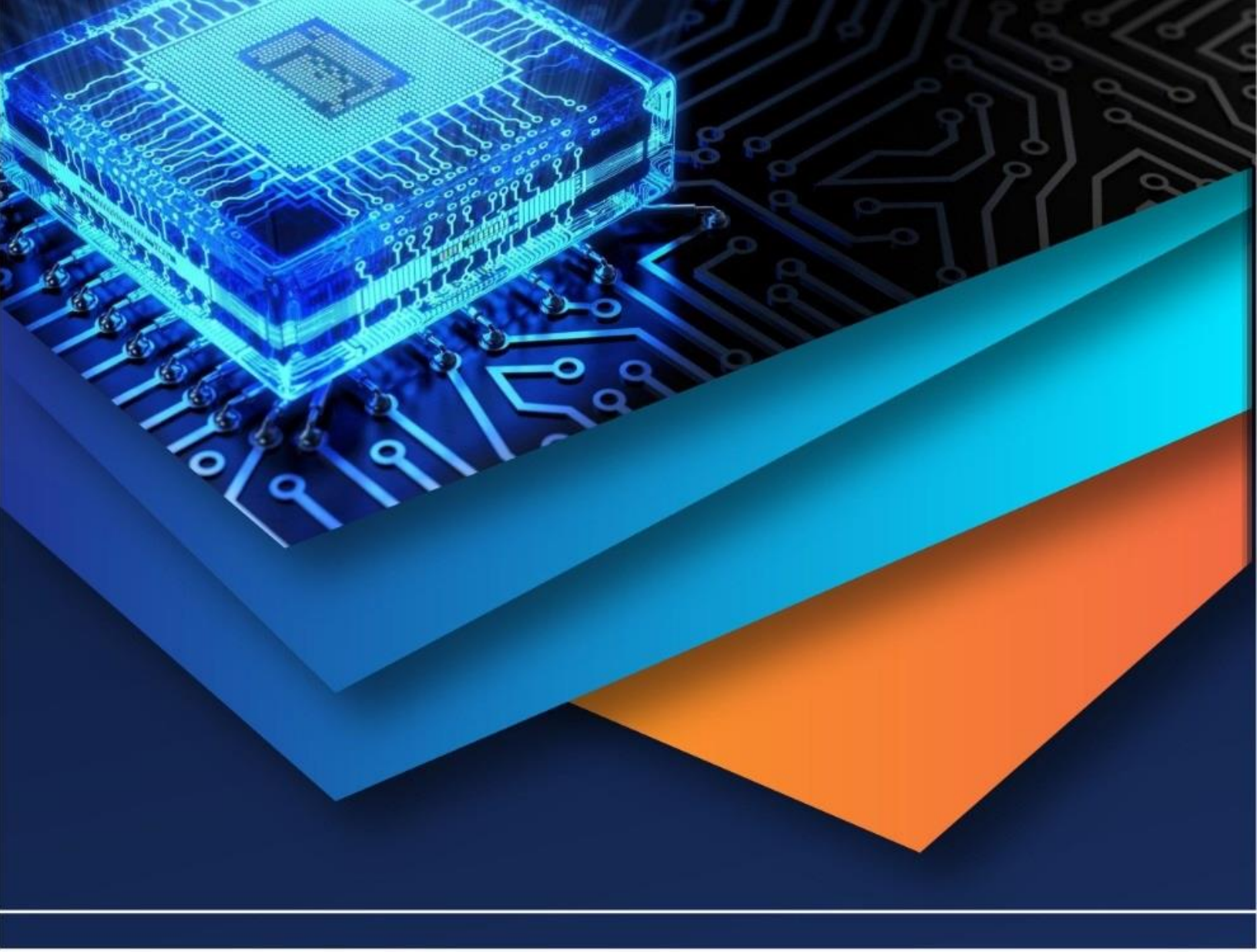
The proposed phishing detection system effectively combines deep learning, explainability, and real-time deployment to provide a comprehensive security solution. The hybrid CNN-LSTM model ensures high accuracy by capturing both local and sequential features of URLs. With SHAP and LIME, the system is more trustworthy for end users .Fast API integration and a
browser extension enable smooth real-time detection and user alerts. Compared to traditional methods, this system is more adaptive, user-friendly, and robust against modern phishing threats. It lays a strong foundation for future enhancements using advanced AI techniques.

VI. FUTURE ENHANCEMENTS

The proposed system opens several avenues for future enhancements. One potential improvement is the incorporation of Reinforcement Learning to allow the model to adapt dynamically to new phishing techniques based on user interactions and feedback. Another direction could involve integrating Graph Neural Networks (GNNs) to analyze relationships between domains, URLs, and hosting infrastructure for deeper threat analysis. Expanding the dataset to include multilingual phishing websites and more diverse attack types can also improve the model's generalizability. Additionally, implementing a cloud-based dashboard for centralized monitoring and user management could make the system suitable for enterprise use. Finally,. adding multi-platform support for the browser extension-including Firefox and Edge would increase accessibility and user adoption.

REFERENCES

[1] Altwaijry, Najwa, Isra Al-Turaiki, Reem Alotaibi, and Fatimah Alakeel. 2024.
[2] T. Niu and B. Wu, "Visual-based Phishing Website Recognition," 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2024, pp. 992-997, doi: 10.1109/IMCEC59810.2024.10575293.
[3] S. Asiri, Y . Xiao,S . Alzahrani, S . Li" A survey of Intelligent Detection Designs of HTML URL Phishing Attacks," in IEEE Access, vol 11, pp. 6421-6443, 2023
[4] J. V. Jawade and S. N. Ghosh, "Phishing Website Detection Using Fast.ai library," 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2021, pp. 1-5.
[5] F. Tajaddodianfar, J. W. Stokes and A. Gururajan, "Texception: A Character/Word-Level Deep Learning Model for Phishing URL Detection," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 2857-2861.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY