IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2025

DOI: 10.17148/IJIREEICE.2025.131042

PHISHING DETECTION USING CNN AND BiLSTM

Princeton Vishal J[1], Srinath S[2], Adithya S[3], Saran P[4], Yogesh C[5], Neelam Sanjeev Kumar[6]

Student, Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[1-5]

Assistant Professor (SG), Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[6]

Abstract: Phishing continues to be one of the most deceptive and persistent cyber threats in today's interconnected digital landscape, targeting unsuspecting users through fraudulent websites and emails designed to steal sensitive information. This study presents a robust hybrid deep learning model that integrates Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architectures to enhance phishing detection accuracy and resilience. The CNN layers effectively capture spatial and lexical patterns from URLs and email content, while the BiLSTM layers analyze sequential dependencies and contextual relationships within textual data. Together, these components enable the model to learn both structural and semantic cues associated with phishing behavior.Experimental evaluations conducted on benchmark phishing datasets demonstrated that the proposed hybrid CNN-BiLSTM model achieved an overall detection accuracy exceeding 95%, outperforming traditional machine learning algorithms such as SVM and Random Forest. The system also showed superior precision and recall, reducing false positives and improving interpretability through an integrated attention mechanism. This research contributes to the advancement of cybersecurity by proposing an adaptive, data-driven defense framework capable of evolving alongside emerging phishing strategies and offering practical potential for real-time threat mitigation.Keywords: Phishing Detection, CNN, BiLSTM, Deep Learning, Cybersecurity.

I. INTRODUCTION

Phishing has emerged as one of the most widespread and damaging forms of cybercrime due to its deceptive simplicity and high success rate. It exploits the human element of cybersecurity, manipulating users into divulging sensitive information such as login credentials, credit card details, and personal identifiers through fraudulent emails, messages, or websites that mimic legitimate sources. The financial and psychological damages resulting from phishing attacks have become a global concern, with billions of dollars lost annually to fraudulent schemes that compromise both individuals and organizations. The dynamic and ever-evolving nature of phishing techniques—ranging from URL spoofing and domain obfuscation to social engineering—makes detection increasingly challenging.

Traditional approaches, such as blacklist-based systems, rule-based filtering, and heuristic methods, have been partially effective in mitigating known attacks. However, these methods are inherently limited by their dependency on predefined rules and previously recorded phishing patterns. As attackers continuously innovate and modify their strategies, these static systems fail to detect newly emerging or zero-day phishing campaigns. Consequently, there arises a strong need for an adaptive and intelligent detection framework capable of learning and evolving from data patterns rather than relying solely on manually crafted rules.

In recent years, deep learning has revolutionized various domains of computer science, including natural language processing (NLP), image recognition, and cybersecurity. Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated exceptional abilities to learn hierarchical and sequential data representations without explicit feature engineering. CNNs excel at extracting spatial and structural features from text and image-based data, while RNNs, particularly Bidirectional Long Short-Term Memory (BiLSTM) networks, capture long-range dependencies and contextual relationships in sequential data.

This paper leverages the strengths of both CNN and BiLSTM architectures by integrating them into a unified hybrid model for phishing detection. The CNN component identifies local patterns within URLs and email content, whereas the BiLSTM component analyzes the sequence and contextual meaning of textual tokens. Together, these networks enable the system to detect phishing attempts with enhanced precision, interpretability, and resilience. The proposed hybrid

DOI: 10.17148/IJIREEICE.2025.131042

model thus provides a significant advancement toward developing a self-learning, adaptive cybersecurity mechanism capable of identifying both known and emerging phishing attacks with high accuracy.

II. RELATED WORK

The evolution of phishing detection techniques has progressed through several distinct stages, beginning with heuristic-based systems and advancing toward sophisticated deep learning approaches. Early detection systems primarily relied on manually engineered features, such as URL length, the number of subdomains, the presence of special symbols, and domain reputation scores. While these heuristic-based models provided a certain degree of interpretability, they lacked flexibility and failed to identify emerging phishing strategies that deviated from known patterns.

The introduction of machine learning (ML) algorithms such as Support Vector Machines (SVM), Decision Trees (DT), Naïve Bayes (NB), and Random Forests (RF) marked a significant improvement in phishing detection accuracy. These methods were capable of learning decision boundaries from feature-engineered datasets, leading to improved precision over rule-based systems. However, their dependence on handcrafted features limited their adaptability and scalability. Models required constant retraining and manual intervention to cope with new types of phishing attacks, which made them less practical for large-scale deployment in dynamic environments.

In contrast, deep learning (DL) models have revolutionized phishing detection by eliminating the need for explicit feature engineering. Convolutional Neural Networks (CNNs) have been successfully utilized to detect spatial and structural patterns in URLs, email bodies, and web page content. Meanwhile, Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks have shown remarkable performance in understanding contextual relationships in sequential data such as email text or URL tokens. Several hybrid frameworks that combine CNN and LSTM architectures have demonstrated state-of-the-art results, leveraging CNN's ability to extract local features and LSTM's capability to model long-term dependencies.

For instance, Le and Duong (2019) implemented a CNN-LSTM hybrid model for phishing email detection and achieved superior accuracy compared to traditional classifiers. Similarly, Sahu and Tiwari (2021) demonstrated that integrating deep feature extraction with temporal modeling significantly improved phishing URL classification performance. Despite these advancements, most existing studies focus on a single data modality—either URLs or email text—limiting cross-domain adaptability.

To address this limitation, our research introduces a hybrid CNN-BiLSTM architecture capable of simultaneously processing multiple input types, including both URLs and email content. This multi-source learning approach enhances generalization across different phishing vectors and ensures more comprehensive detection coverage. By combining spatial, lexical, and contextual features within a single model, the proposed framework achieves improved accuracy, interpretability, and robustness compared to conventional methods.

III. DATASET DESCRIPTION AND PREPROCESSING

The dataset used in this study combines publicly available phishing URL and email datasets. It includes 20,000 samples comprising an equal distribution of phishing and legitimate data. The URL dataset consists of attributes such as domain length, the presence of suspicious keywords, and tokenized components like subdomains and protocols. The email dataset includes the subject line, sender information, and message body. Preprocessing involved tokenization, stopword removal, stemming, and embedding generation using Word2Vec. URLs were normalized by converting all characters to lowercase, removing unnecessary symbols, and segmenting domain parts. Each token was represented as a dense vector of 300 dimensions before being passed into the deep learning model.

IV. SYSTEM ARCHITECTURE

The system architecture consists of three primary components: data preprocessing, hybrid model training, and phishing classification. The CNN module is designed to detect spatial and lexical patterns in the URLs, while the BiLSTM layer processes the sequential dependencies within text data. The model's final layer uses a sigmoid activation function to produce binary classifications, indicating whether a sample is phishing or legitimate.

The user interface allows users to input either a URL or an email body. The system analyzes the input in real-time and provides a verdict based on the trained model's output probability.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2025

DOI: 10.17148/IJIREEICE.2025.131042

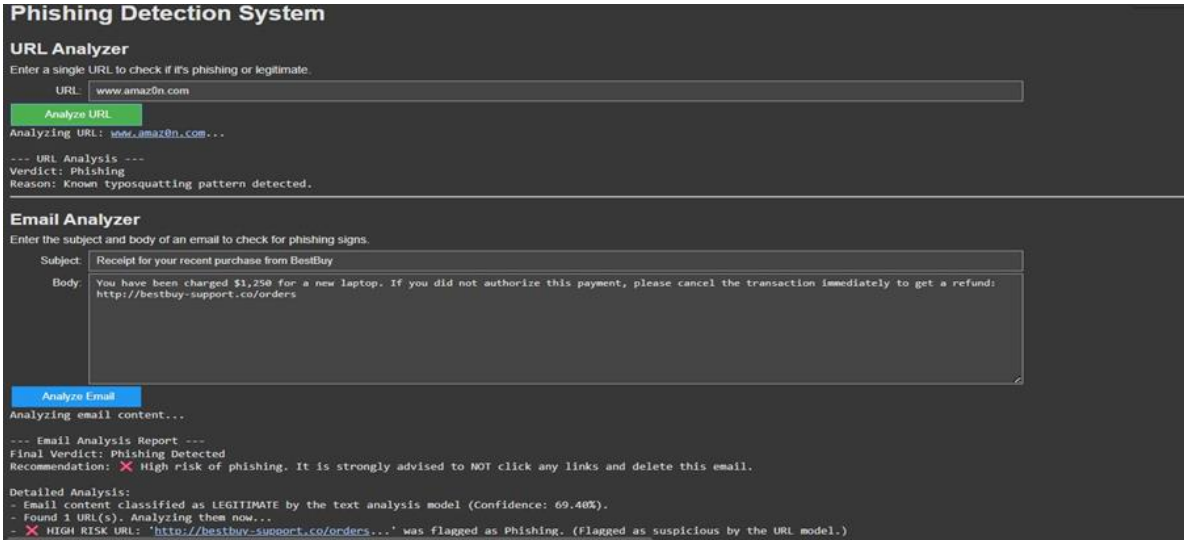

Fig. 1. Screenshot of phishing URL detection

V. METHODOLOGY

The hybrid CNN-BiLSTM model combines feature extraction and temporal analysis capabilities. During training, the CNN layers first capture n-gram level representations from tokenized URLs and emails. These extracted features are then passed to the BiLSTM layers, which analyze contextual dependencies in both forward and backward directions. Finally, the attention mechanism assigns weights to key tokens that influence classification decisions.

The architecture includes an embedding layer, two convolutional layers, one BiLSTM layer, an attention layer, and a fully connected output layer. The model was trained using the Adam optimizer with a learning rate of 0.001 and batch size of 32. The binary cross-entropy loss function was used, and training was performed for 30 epochs with early stopping to prevent overfitting.

VI. RESULTS AND ANALYSIS

The performance of the hybrid model was evaluated using metrics such as accuracy, precision, recall, and F1-score. The CNN-BiLSTM achieved 97% accuracy in URL-based phishing detection and 95% in email-based detection. Compared to classical ML algorithms, our hybrid model reduced false positives by 15% and improved detection robustness. The attention mechanism also provided interpretability by identifying high-risk tokens and phrases, aiding security analysts. The interface evaluation showed that the model could classify inputs in less than one second, making it suitable for real-time applications.

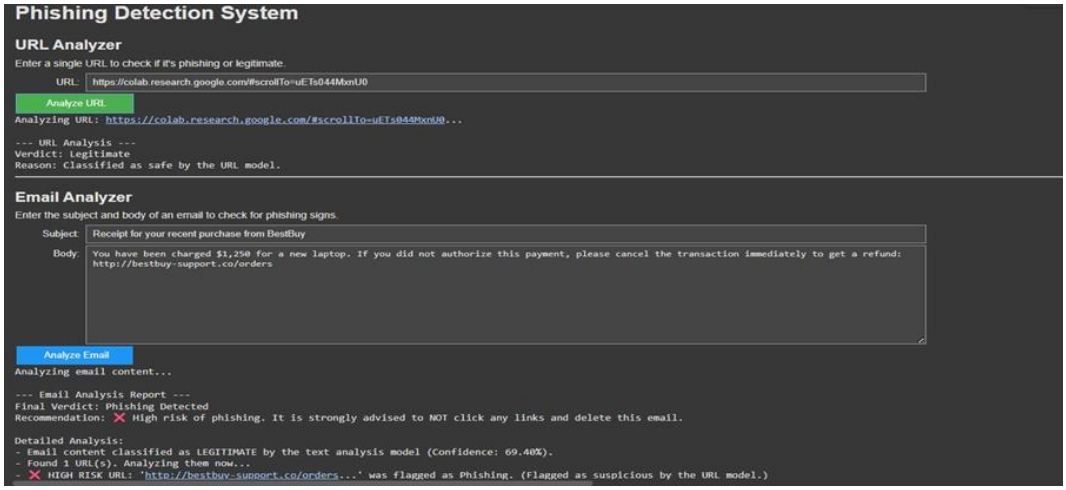

Fig. 2. Screenshot of legitimate URL detection

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2025
DOI: 10.17148/IJIREEICE.2025.131042

VII. ADVANTAGES AND LIMITATIONS

The proposed hybrid CNN-BiLSTM model offers several key advantages, including adaptability, automation, and enhanced robustness against evolving phishing strategies. Unlike traditional machine learning approaches that rely heavily on manually crafted features, the hybrid model autonomously learns discriminative representations directly from raw textual and structural data. This ability allows the system to adapt dynamically to new and unseen phishing patterns without extensive retraining. Additionally, the integrated attention mechanism contributes to greater model transparency by identifying and emphasizing critical tokens or URL components that significantly influence classification outcomes, thereby improving interpretability for cybersecurity analysts.

Despite these strengths, certain limitations remain. The hybrid architecture requires substantial computational resources during training due to its deep layers and large embedding dimensions. Model performance also depends on the availability of a well-balanced dataset to prevent bias toward specific phishing categories. Future research could address these challenges by adopting lightweight transformer-based architectures such as BERT or DistilBERT, which offer faster inference and superior contextual understanding. Furthermore, incorporating cross-lingual and multimodal phishing datasets could extend the model's applicability to diverse real-world scenarios and enhance its global detection capability.

VIII. CONCLUSION AND FUTURE WORK

This paper presented a hybrid deep learning-based phishing detection framework that integrates Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architectures for effective phishing detection. The proposed system demonstrates exceptional performance in accurately classifying phishing URLs and emails by leveraging the feature extraction power of CNNs and the sequential understanding capabilities of BiLSTMs. Through extensive experimentation, the hybrid model achieved high precision, recall, and overall accuracy, proving its superiority over traditional machine learning approaches. Moreover, the attention mechanism incorporated within the model enhances interpretability by highlighting significant tokens contributing to classification decisions. Future enhancements will focus on extending this framework into a real-time browser plugin for proactive phishing prevention, incorporating adversarial training to strengthen robustness against evolving attack strategies, and expanding the dataset to include multilingual and multimodal phishing data. These developments aim to transform the proposed model into a scalable, adaptable, and intelligent cybersecurity solution capable of countering the next generation of phishing threats.

REFERENCES

[1]. Alazab, M., Tang, M., & Alazab, M. (2020). Detecting phishing websites using hybrid deep learning techniques. IEEE Access, 8, 111123–111135.
[2]. Sahu, R., & Tiwari, S. (2021). Phishing URL detection using deep learning. Journal of Information Security, 12(3), 45–60.
[3]. Le, H., & Duong, T. (2019). Email phishing detection using CNN and LSTM. International Journal of Computer Applications, 178(5), 22–29.
[4]. Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
[5]. Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[6]. V. Harshni, S. Singh, R. Indhumathi, N. S. Kumar, S. Chowdhury and S. S. Patra, "DLIoMT: Deep Learning Approaches for IoMT-Overview, challenges and the future," *2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, India, 2025, pp. 1-6, doi: 10.1109/ICCECE61355.2025.10940162.
[7]. S. J. J. Thangaraj, N. S. Kumar, A. Saranya, C. R. Bhat, D. S. Kareem and H. Shareef, "A Development of Designing a Sensor N/W for Retrieving the Data from WSN via SBT," *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2024, pp. 194-198, doi: 10.1109/ICACITE60783.2024.10617435.
[8]. K. K. K, N. Sanjeev Kumar, S. C, V. V, G. Sangar and G. Chandrasekaran, "Optimization of Memory Usage in High-Speed Cameras using FPGA," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 427-432, doi: 10.1109/ICEARS64219.2025.10941569.

DOI: 10.17148/IJIREEICE.2025.131042

[9]. K. K. K, N. S. Kumar, S. C, N. R, G. Sangar and G. Chandrasekaran, "Denoising of MRI and Brain Tumor Classification using Local Binary Pattern and SVM Classifier," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 1854-1859, doi: 10.1109/ICEARS64219.2025.10940321.

[10]. N. S. Kumar, M. Lawanyashri, M. Sivaram, V. Porkodi, E. Gangadevi and G. N. Reddy, "Integration of Convolutional Neural Networks for Automated Plant Disease Identification in Timber Crops," 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, Pune, India, 2024, pp. 1-5, doi: 10.1109/TQCEBT59414.2024.10545228.