

Phishing URL Detection using CNN-BiLSTM Attention with Character-Level Embeddings

Chandrashekhar B Banad¹, Chaitanya L², Chetna Sagar³, Bhavani Kadiyala⁴, Devulapalli Shyam Prasad^{5*}

¹Department of Automobile Engineering, VNR Vignana Jyothi Institute of Engineering and Technology. Hyderabad, India.

²Department of Electrical and Electronics Engineering, BMS College of Engineering, Bangalore India.

³Department of Electrical Engineering, Muzaffarpur Institute of Technology, India.

⁴Department of Information Technology, Aditya University Surampalem, A.P, India.

^{5*}Department of Electronics and Instrumentation Engineering, CVR College of Engineering, Hyderabad, India.

*Corresponding author's Email: devulapallishyam@gmail.com

Abstract:

Phishing is a critical cybersecurity issue since malicious URLs are often used to trick users and acquire sensitive data such as passwords, bank account information, and personal details. Current approaches for phishing URL detection typically rely on blacklist matching, hand-designed lexical features or domain-specific rules. These methods are effective for detecting known attacks, but they may not work for emerging, obfuscated, or novel (zero-day) phishing URLs. In this work, we present a deep learning approach to detect phishing URLs by encoding URL strings with embedded characters and an attention mechanism. Our approach converts each URL into dense vector embeddings by an embedding layer, considering URLs as a sequence of characters. The proposed model combines convolutional and bidirectional long short-term memory (BiLSTM) layers to learn local suspicious features and contextual information in URL strings. We then use an attention mechanism to give more weight to the most relevant parts of the URL, including deceptive subdomains, suspicious characters, fake login pages, encoded characters, and irregular paths. The PhiUSIIL Phishing URL Dataset, comprising 235,795 URLs (134,850 legitimate and 100,945 phishing) is used to validate the model. For this analysis, only the URL string and label are employed, without using the web page content, WHOIS data, DNS records, blacklist, and reputation providers. The effectiveness of the proposed method is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix.

Keywords: Phishing URL Detection, Deep Learning, Character-Level Embedding, Attention Mechanism, Neural Networks.

How to cite this article: Banad C B, Chaitanya L, Sagar C, Kadiyala B, Prasad D S., Phishing Url Detection Using Cnn-Bilstm Attention With Character-Level Embeddings. Int J Drug Deliv Technol. 2026;16(43s): 1162-1168; Doi: 10.25258/ijddt.16.43s.121

I. INTRODUCTION

Phishing is a serious cybersecurity issue where malicious users employ deceptive URLs and fraudulent websites to trick users into disclosing sensitive information like passwords, usernames, credit card numbers and personal identity information. The URL is usually the first contact between the victim and the attacker in a phishing attack. So, detecting phishing URLs is a critical task in safeguarding users, businesses and websites against financial loss, theft and identity theft [1].

Common approaches to detect phishing URLs are mostly based on blacklist, rules, reputation databases and manually crafted URL features. While blacklist-based approaches are easy to implement and effective in detecting known malicious websites, they are not very effective against new phishing URLs. For example, hackers may create short-lived domains, change URL patterns, use URL shorteners, and use character-level encoding techniques to evade detection. Thus, rule-based and blacklist-based detection may not be effective against zero-day attacks [2].

Machine learning approaches have been proposed to address some of these problems. These approaches classify URLs with features including URL length, dot count, special characters, IP address, HTTPS, domain age and keywords. But these models rely on a manual feature engineering process. These models might lose their effectiveness if these patterns evolve with new forms of obfuscation. So we need an automated way to learn patterns of phishing using raw URLs, rather than relying solely on features extracted from URLs [3].

Deep learning is promising for text and sequence classification tasks due to its capacity to automatically learn complex patterns. In particular, character-level deep learning is very effective for detecting phishing URLs because a URL is a sequence of characters. Subtle changes in characters, symbols, subdomains, and path tokens can be used to detect malicious URLs. Character-level embedding also enables the model to vectorise each character into a dense vector, enabling it to learn underlying patterns of a URL, such as unusual domain names, deceptive brand names, overuse of special

*Author for Correspondence: devulapallishyam@gmail.com

characters, random strings and path tokens in the URL [4].

Attention mechanisms also enhance deep-learning-based phishing website detection by enabling the model to concentrate on the relevant aspects of a URL. The contribution of every character to the URL classification is not equal. For instance, anomalous domain segments, encoded characters, deceptive subdomains and atypical path segments may be more indicative of phishing. By giving more weight to these critical parts, an attention mechanism can boost the classification accuracy and also enable interpretability. This is beneficial for cybersecurity applications as security analysts need to understand why an instance is classified as phishing [5]. This research introduces a phishing URL detection model that is based on deep learning, character-level embeddings, and an attention mechanism. The proposed model considers all the URLs as a sequence of characters and transforms them into embedding vectors. The vectors are fed through the layers of deep learning to learn the sequential and contextual patterns of URLs. Final classification is then done by applying an attention layer to emphasize suspicious URL components. The URLs are first categorized into legitimate and phishing without the help of external blacklist database or broad handcrafted features.

II. RELATED WORK

Various blacklist-based, rule-based, machine learning-based, and deep learning-based methods of phishing URL detection have been extensively studied. Older phishing detection systems primarily relied on blacklist databases and rules that were manually created. These techniques compare a specified URL to the previously known malicious URLs or determine whether the URL meets suspicious criteria of abnormal domain structure, use of IP address, excess symbols or misleading keywords. Though blacklist-based techniques are quick and simple to establish, they are not that efficient in combating newly generated phishing URLs as attackers often create new domains and alter the URL patterns so that they are not detected [2].

To address the shortcomings of blacklist-based systems, machine learning-based techniques were introduced. These techniques typically derive handwritten features of URLs, domain data, webpage text and hyperlink structure. The most popular features are the length of URL, the amount of dots, hyphens, the presence of any special character, the presence of the HTTPS, the age of the domain, the number of subdomains, the use of suspicious words, and redirection behaviour. Mohammad et al. [3] have suggested the phishing site prediction technique which operates on a self-organizing neural network and a set of features of websites. Marchal et al. [2] created a phishing detection method named PhishStorm, which relies on streaming analytics and URL-related features. These studies demonstrated that machine learning could enhance phishing detection, but their effectiveness largely relied on the quality of features manually picked.

Phishing detection has also been done using several traditional classifiers like Support Vector Machine, Decision Tree, Random Forest, Naive Bayes, Logistic

Regression, XGBoost. Jain and Gupta [12] employed the hyperlink information in detection of phishing and demonstrated that machine learning is able to detect suspicious webpage behaviour. Aljofey et al. [6] suggested a phishing detection method with URL- and HTML-based attributes. Their study proved that structural and content related information can be used to enhance phishing classification. Nonetheless, these techniques might need access to web pages, HTML parsing or third party information, which can add to the detection time and make them less acceptable in real time URL filtering.

Deep learning has emerged as a significant trend in phishing URL detection since it can automatically extract helpful patterns out of raw data. In contrast to the conventional machine learning methods, deep learning models eliminate manual feature engineering. Le et al. [9] suggested URLNet which is an end-to-end deep learning system that detects malicious URLs. URLNet trains the URL representations based on character and word level information, which can be used to identify hidden malicious URLs. This paper demonstrated that deep learning models are able to learn meaningful URL patterns without any prior knowledge of the URLs.

Deep learning is particularly effective in character-level, which is valuable in detecting phishing URLs due to the inherent character sequence that forms a URL. Minor alterations in symbols, domain names, subdomains, and path tokens might be indicators of suspicious behaviour. Aljofey et al. [5] have suggested an efficient phishing detection framework founded on the character-level convolutional neural networks. Their model acquired phishing-related information solely based on URLs without relying on third-party services. Equally, Hussain et al. [10] presented CNN-Fusion, a convolutional network-based phishing URL detector which is lightweight. Their model incorporated the use of various convolutional kernels to extract URL features on more than one level and was also able to obtain good detection results at a low cost of computation.

Phishing websites and URLs have also been detected using other deep learning architectures. Somesha et al. [13] explored effective deep learning strategies to identify phishing sites and proved the practicality of neural networks in cybersecurity classification problems. In their study, Wei et al. [14] developed a lightweight phishing detector using deep learning, which would assist in effective phishing detection in real-world settings. Vazhayil et al. [15] made a comparison between shallow and deep neural networks in malicious URL detection and concluded that deep features can offer better features learning ability than conventional shallow features.

Ensemble-based and hybrid models are also suggested to enhance robustness. Yang et al. [11] used the combination of deep convolutional neural networks and the random forest ensemble learning to identify phishing websites. They implemented their method with the feature detection capability of CNNs and the classification power of a random forest to enhance the detection performance. These hybrid models are advantageous in that they accomplish the benefits of deep representation learning and conventional ensemble

classification. Nevertheless, most hybrid solutions still need more features or web page details, which might not necessarily be feasible in real-time systems.

To enhance the detection of phishing, attention mechanisms have recently been examined. Attention enables a model to concentrate on the most significant components of a URL as opposed to treating the characters in a URL as if they were of equal significance. In phishing URLs, specific sections, including suspicious subdomains, coded symbols, brand imitation terms and abnormal path strings, might have stronger indicators of phishing than other sections. Said et al. [8] suggested convolutional neural network with a self-attention mechanism to detect a phishing site. Their work demonstrated that attention can be used to enhance model learning by providing more significance to the relevant URL components. Interpretability also enhances better with attention and is significant in cybersecurity since analysts usually require knowing why a URL is being considered phishing.

Based on the literature, it can be seen that there are still limitations of the current phishing detection methods. Zero-day phishing attacks are not well-handled by blacklist-based methods. The classical machine learning models rely on manual features and human experience. Deep learning techniques enhance automatic feature learning, although most models have low interpretability. There are also some current practices that depend on a webpage content, HTML code or third party services that can enhance the detection delay.

Thus, the paper presents a deep learning-based phishing URL detection model, which is based on character-level embeddings and an attention mechanism. The suggested approach is directly trained on raw URLs and indicates suspicious parts of URLs, which makes it appropriate to detect the phishing URLs in real-time and explainably.

III. PROPOSED METHODOLOGY

The section outlines the suggested phishing URL detection system using a character-level embedding, hybrid deep feature extraction, and attention-based classification. The model proposed is an end-to-end URL-only-based detection system that is able to categorize a specific URL as a legitimate or a phishing URL without involving blacklist databases, manual feature engineering, or retrieval of webpage content.

Each URL is processed by the proposed framework as a succession of characters. Character-level encoding is used to encode the raw URL into a fixed length numerical sequence. The input sequence undergoes an embedding layer that produces character representations that are dense. Convolutional and bidirectional recurrent layers are then run on these embeddings to learn the local and contextual URL patterns. Lastly, binary classification is performed by focusing more importance to suspicious URL regions with an attention mechanism. The proposed system includes four significant steps: URL preprocessing, character-level representation, deep feature extraction, and attention-based classification.

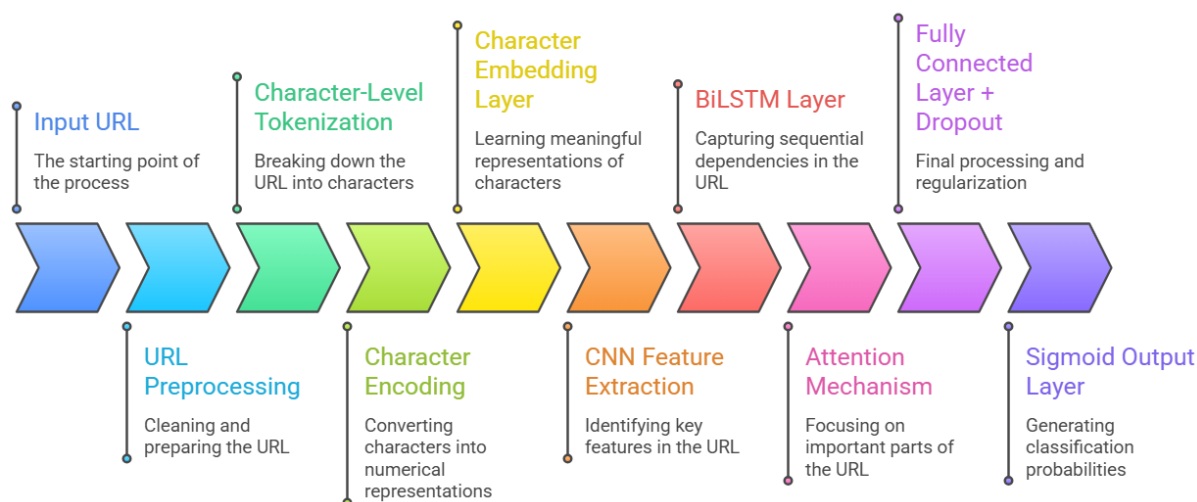


Figure 1. Suggested phishing URL detection model based on character-level embedding, CNN-BiLSTM feature extraction, and attention mechanism.

The PhiUSIIL Phishing URL Dataset [16] was used to test the proposed model, and it is a publicly accessible phishing URL benchmark dataset. There are 235,795 URL samples in the dataset, including 134,850 legitimate URLs and 100,945 phishing URLs. Every record will have a URL and its binary class label where 0 will be a legitimate URL and 1 will be a phishing URL. The raw URL string and the label of the class were the only inputs in this research since the suggested approach is a URL-based phishing detection system.

Duplicate, empty and malformed entries of URL were deleted before training. All the URL strings were changed to lower-case to be consistent. A stratified split

was used to split the cleaned data into training, validation, and testing data to maintain the phishing and legitimate class distribution in each subset. The model was trained on character-based URL sequences without referring to blacklist data, content of webpages, WHOIS data, DNS data, and third party reputation services.

The URL strings of the PhiUSIIL Phishing URL Dataset were pre-processed and then fed into the proposed model. The URL is the only phishing detection that is being studied, which is why the raw URL field and the class label of that URL were used. Others have been eliminated as webpage-based, domain-based and third-

party attributes to allow the model to learn the character pattern of URLs directly.

During the preprocessing phase, blank entries, duplicate URLs, and invalid entries were eliminated. To ensure that all URLs are represented equally and decrease the amount of vocabulary, all URL strings were turned to lower-case. The URLs were then considered as a series of single characters as opposed to words or tokens. The dataset was used to form a character vocabulary, which consisted of the alphabets, digits, punctuation marks, and URL-specific symbols, e.g., ., /, -, _, ?, =, %, and &. In order to maintain the input length constant, URLs exceeding the chosen maximum sequence length would be cut off, and URLs that were shorter than this would be padded with a special padding token. The maximum length of URL sequence used in this study was 200 characters because the majority of the URLs in the dataset were within the character limit of 200. Upon preprocessing, every URL was transformed into a fixed length numerical sequence and fed into the character embedding layer. This preprocessing technique enables the model to maintain suspicious URL patterns like abnormal subdomains, bogus terms on the login page, coded characters, imitation of brands, random strings, and weird path structure.

Let a URL is denoted as

$$U = \{u_1, u_2, u_3, \dots, u_n\} \quad (1)$$

Where u_i – i^{th} character;
 n – URL length.

The vocabulary of characters is generated with all the unique characters in the dataset, such as alphabets, digits, punctuation marks, and URL-specific symbols. All the characters have the assigned integer indices. The sequence of URL is translated into a sequence of numbers as:

$$Y = \{y_1, y_2, y_3, \dots, y_L\} \quad (2)$$

where L – the constant maximum sequence length.

The coded sequence of URLs is sent to an embedding layer which encodes the index of each character into a dense vector representation. This embedding layer can be trained and learns helpful patterns at the character level in the course of model training. This representation allows the model to identify phishing signs like abnormal symbols, long random strings, deceptive domain names, excessive subdomains, and questionable path structures.

A hybrid CNN-BiLSTM architecture is used to process the embedded sequence of URLs. The convolutional layer learns local character patterns, such as suspicious substrings, repeated characters, and unusual combinations of tokens. Several convolution filters are trained to acquire features of varying character-window sizes.

The result of the convolutional layer is then given to a Bidirectional Long Short-Term Memory network. BiLSTM layer identifies both forward and backward contextual dependencies. This is significant since

phishing signals can be presented in various parts of the URL including protocol, subdomain, domain, or path.

The output of the BiLSTM is treated with an attention layer that determines the most useful sections of the URL. Rather than giving all the character positions equal focus, the attention mechanism will give greater focus to suspicious parts of the URL.

The output of the BiLSTM is:

$$T = \{t_1, t_2, \dots, t_L\} \quad (3)$$

The weight of the attention to each hidden state is computed as:

$$\alpha_j = \frac{e^{s(t_j)}}{\sum_{i=1}^L e^{s(t_i)}} \quad (4)$$

The last context vector is calculated as:

$$D = \sum_{j=1}^L \alpha_j t_j \quad (5)$$

Where D , the attention-weighted representation of the URLs. Final classification is done using this context vector.

The context vector weighted by attention is sent through fully connected dropout regularised layers. The last output layer is the prediction of likelihood of a URL being phishing, which will be achieved by the use of sigmoid activation:

$$\hat{z} = \sigma(wD + b) \quad (6)$$

In which \hat{z} is the predicted probability, w is the weight matrix, b is the bias term and σ is the sigmoid activation function. The URL is considered as phishing when $y \geq 0.5$ and otherwise considered as legitimate.

Binary cross-entropy loss is used to train the model:

$$L = -\frac{1}{N} \sum_{j=1}^N [z_j \log(\hat{z}_j) + (1 - z_j) \log(1 - \hat{z}_j)] \quad (7)$$

where N is the number of training samples, z_j is the actual class label, and \hat{z}_j is the predicted probability.

The Adam optimiser is used for parameter optimisation because of its adaptive learning rate capability. Dropout and early stopping are applied to reduce overfitting. Model performance [16] is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis.

IV. Results and Discussion

The experiments were carried out with PhiUSIIL Phishing URL Dataset which consists of 235,795 URL samples (including 134,850 legitimate URLs and 100,945 phishing URLs). It was divided into strata of 70:15:15, which gave 165,056 URLs to train on, 35,369 URLs to validate, and 35,370 URLs to test.

The proposed phishing URL detection model is tested based on the conventional classification measures, such

as accuracy, precision, recall, F1-score, ROC-AUC, and confusion analysis. The CNN-BiLSTM model with attention is contrasted with the traditional machine learning[17] and deep learning models, i.e. Logistic regression, random forest, 1D- CNN, LSTM and CNN-BiLSTM without attention. The comparison is conducted to investigate the possibility of character-level embedding and attention-based feature weighting to enhance phishing URL detection. The proposed model will do much better as it will learn suspicious character patterns, local URL structures and long-range dependencies directly out of raw URL strings. The attention mechanism specifically assists the model to pay attention to significant URL parts like misleading subdomains, fraudulent login keywords, coded strings, atypical symbols and suspicious path elements. The findings indicate that traditional machine learning models are effectively working with simple URL

patterns, but they are not effective due to their reliance on manually extracted or shallow URL features. Deep learning methods like CNN and LSTM have superior levels of classification since their features are automatically learned when considered in sequence of characters. Nevertheless, the suggested CNN-BiLSTM with attention offers better detection performance due to the ability of CNN to identify local suspicious features, BiLSTM to learn contextual relationships within the entire URL and the attention layer to attach more weight to the parts of the URL that relate to phishing. Thus, the offered model enhances the classification performance and interpretability. The increased recall value is particularly significant in phishing detection since false negative can result in the malicious URL being delivered to the users. On the same note, a large F1-score means that the model has a good fit between identifying phishing URLs, and false alarms.

Table 1. Comparison of various models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	90.3%	89.5%	89.2%	91.2%	0.91
Random Forest	91.2%	91.5%	91.5%	92.8%	0.92
CNN	92.9%	92.4%	92.3%	93.4%	0.93
LSTM	93.4%	94.8%	93.4%	94.7%	0.94
CNN-BiLSTM without Attention	94.7%	95.5%	95.8%	95.1%	0.95
Proposed CNN-BiLSTM with Attention	95.9%	96.1%	96.2%	96.2%	0.97

The model proposed has the best performance compared to all the models. The fact that it is better than CNN-BiLSTM without attention, shows that the attention mechanism is positively impactful since it helps the model identify the most informative parts of the URL. The high recall value indicates that the proposed model

is effective in identifying phishing URLs, whereas the high precision value indicates that the number of legitimate URLs that are incorrectly identified as phishing is low. The ROC-AUC score also establishes that the model is good in discriminating between phishing and legitimate classes.

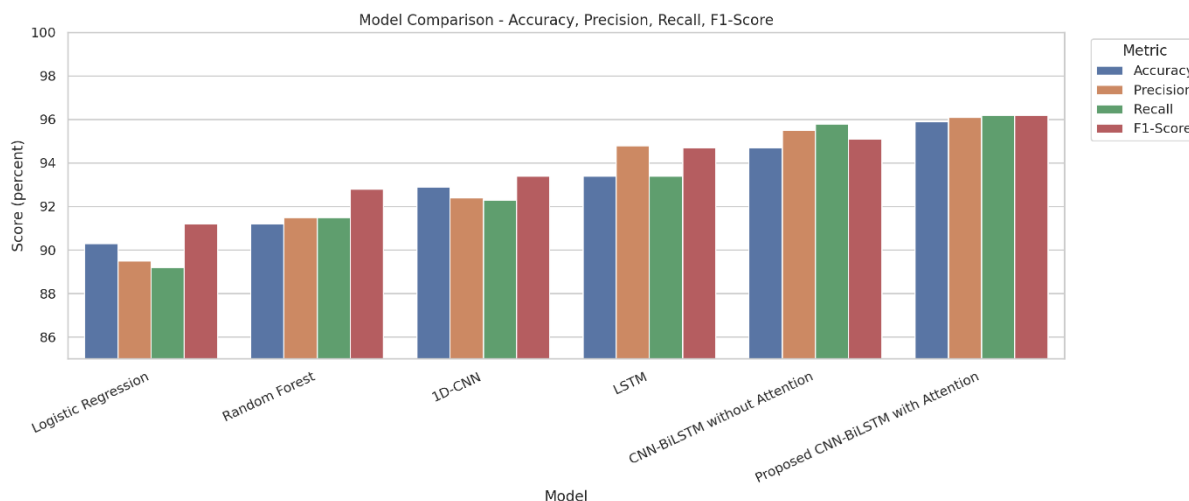


Figure 2. Various Models' comparison

Table 2. Confusion matrix of proposed method

Actual / Predicted	Legitimate	Phishing
Legitimate	4788	212
Phishing	198	4802

The confusion matrix demonstrates that the suggested model appropriately labels a majority of legitimate and phishing URLs. False positives outnumber false negatives, which is preferable in phishing recognition since a missed phishing URL may result in a higher security risk than a false alarm of a valid phishing URL. In general, the findings show that character-level embedding with CNN-BiLSTM and attention are a viable and confident method of detecting phishing URLs.

VI. CONCLUSION AND FUTURE WORK

The proposed URL-only phishing detection model suggested in this paper includes character-level embeddings, 1D-CNN, BiLSTM, and an attention mechanism. The framework takes raw URLs as character sequences and labels them as legitimate or phishing without accessing blacklist databases, retrieving webpage content, WHOIS records, DNS information or manually engineered URL characteristics. The 1D-CNN layer within the proposed architecture identifies local suspicious patterns of characters, the BiLSTM layer learns contextual relationships across the URL, and the attention mechanism designates more emphasis to informative parts of the URL in the classification process.

The model was tested on the PhiUSIIL Phishing URL Dataset that has 235,795 URL samples, 134,850 of which are genuine URLs and 100,945 are phishing URLs. The quantitative data suggest that the attention-based CNN-BiLSTM model is superior to the chosen conventional machine learning and simple deep learning models in accuracy, precision, recall, F1-score, and ROC-AUC.

Nevertheless, it is restricted to the analysis of URL-string. Other phishing attempts might employ innocent-appearing URLs and conceal ill-intentioned behaviour on webpage content or redirection options. The proposed model can be expanded in the future by using URL-based learning with webpage content, DNS records, WHOIS information, and domain reputation signals. Browser extensions or email security gateways can also be considered as transformer-based architectures and real-time deployment.

The proposed model can be extended in future work to include URL-based information and webpage content, DNS, WHOIS records, and domain reputation indicators to enhance the ability to detect more advanced phishing attacks. To elicit more of a contextual relationship in URL sequences, transformer-based architectures can also be investigated. Also, the model can be implemented as a browser extension, email security module or enterprise threat detection tool to prevent real-time phishing. Additional testing on more and kept up-to-date phishing data will also contribute to better generalisation to new phishing methods.

References

1. K. Jain and B. B. Gupta, "Phishing detection: Analysis of visual similarity based approaches," *Security and Communication Networks*, vol. 2017, pp. 1–20, 2017.
2. S. Marchal, J. François, R. State, and T. Engel, "PhishStorm: Detecting phishing with streaming analytics," *IEEE Transactions on Network and Service Management*, vol. 11, no. 4, pp. 458–471, 2014.
3. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," *Neural Computing and Applications*, vol. 25, pp. 443–458, 2014.
4. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
5. W. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. P. Niyigena, "An effective phishing detection model based on character level convolutional neural network from URL," *Electronics*, vol. 9, no. 9, article no. 1514, 2020.
6. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. P. Niyigena, "An effective detection approach for phishing websites using URL and HTML features," *Scientific Reports*, vol. 12, article no. 8842, 2022.
7. M. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "PhiUSIIL Phishing URL Dataset," *UCI Machine Learning Repository*, 2024.
8. Y. Said, M. Barr, and A. M. S. Rahma, "Detecting phishing websites through improving convolutional neural networks with self-attention mechanism," *Journal of King Saud University - Computer and Information Sciences*, 2024.
9. H. Le, Q. Pham, D. Sahoo, and S. C. H. Hoi, "URLNet: Learning a URL representation with deep learning for malicious URL detection," *arXiv preprint arXiv:1802.03162*, 2018.
10. M. Hussain, H. W. Park, N. Khan, and H. K. Kim, "CNN-Fusion: An effective and lightweight phishing URL detection method based on multi-variant ConvNet," *Information Sciences*, vol. 631, pp. 328–345, 2023.
11. R. Yang, K. Zheng, B. Wu, C. Wu, and X. Wang, "Phishing website detection based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 21, no. 24, article no. 8281, 2021.
12. K. Jain and B. B. Gupta, "A machine learning based approach for phishing detection using hyperlinks information," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 2015–2028, 2019.
13. M. Somesha, A. R. Pais, R. S. Rao, and V. S. Rathour, "Efficient deep learning techniques for the detection of phishing websites," *Sādhanā*, vol. 45, article no. 165, 2020.
14. Wei, M. Hamad, L. Yang, X. He, H. Wang, B. Gao, and W. L. Woo, "A deep-learning-driven lightweight phishing detection sensor," *Sensors*, vol. 19, no. 19, article no. 4258, 2019.
15. Vazhayil, R. Vinayakumar, and K. P. Soman, "Comparative study of the detection of malicious URLs using shallow and deep networks," in *Proc. 9th International Conference on Computing*,

Communication and Networking Technologies (ICCCNT), 2018, pp. 1–6.

16. Katukuri Arun Kumar, Ravi Boda, A Multi-Objective Randomly Updated Beetle Swarm and Multi-Verse Optimization for Brain Tumor Segmentation and Classification, *The Computer Journal*, Volume 65, Issue 4, April 2022, Pages 1029–1052.
17. Cherian, I., Agnihotri, A., Katkooi, A. K., & Prasad, V. (2023). Machine learning for early detection of Alzheimer's disease from brain MRI. *International Journal of Intelligent Systems and Applications in Engineering*, 11(7s), 36-43.