Journal of
Information Technology,
Cybersecurity, and Artificial Intelligence

(RESEARCH ARTICLE)

A Hybrid Approach with a Focus on Preprocessing Techniques for Detecting Phishing Websites

Abeer Salawi

AbeerSalawi87@gmail.com

Journal of Information Technology, Cybersecurity, and Artificial Intelligence, 2025, 2(2), 145–155

Article DOI: https://doi.org/10.70715/jitcai.2025.v2.i2.023

Abstract

Phishing is regarded as a significant cybersecurity problem in the digital era, utilizing the fabrication of fraudulent websites to deceive users and expropriate their sensitive information, including passwords and financial data. The growing dependence on the internet has led to a marked increase in the frequency of these attacks, resulting in considerable financial losses for individuals and businesses alike. This underscores the pressing necessity for efficient strategies to counteract such assaults.

This research aims to create a hybrid model for identifying phishing websites via URL analysis. The suggested model combines Convolutional Neural Networks (CNN) with Long Short-Term Memory networks (LSTM) and an Attention Mechanism to make predictions more accurate and uncover hidden patterns in the data. The model was trained on the "Url_Detection_Dataset" from the Kaggle platform, and its performance was assessed using precision, recall, and F1-score measures. The results showed that the hybrid model is better than traditional methods at telling apart real and harmful URLs, making it a useful tool in cybersecurity. The results provide a framework for subsequent research and promote the creation of more resilient, flexible, and effective solutions.

Keywords: Cybersecurity, Phishing Detection, Hybrid Approach, URL Analysis, Deep Learning

1. Introduction

In the contemporary digital age, the virtual realm has grown increasingly interconnected, resulting in a notable increase in cybersecurity concerns impacting both individuals and enterprises. Phishing is a prevalent cyber hazard utilized by cybercriminals to acquire personal data, adversely affecting cybersecurity and privacy and posing a significant risk to the stability of financial institutions. As cybercrime proliferates, phishing persists as a significant concern, wherein attackers fabricate deceptive websites to manipulate users into revealing critical information, including passwords and credit card numbers. [1][2]. The principal targets of these assaults comprise major enterprises, financial institutions, payment service providers, and military and government agencies, who frequently endure significant detriment regarding financial loss and reputation (APWG Security Report, 2017). Recent statistics indicate a significant global rise in phishing attacks. Cybersecurity records from (IC3.GOV ,2023) reveal that the FBI's Internet Crime Complaint Center recorded 880,418 cybercrime complaints, with estimated losses above $12.5 billion. The report does not delineate damages exclusively linked to phishing assaults, although it underscores phishing as one of the predominant forms of cybercrime. Furthermore, investment fraud emerged as the most financially detrimental kind of cybercrime in 2023, with losses escalating from $3.31 billion in 2022 to $4.57 billion in 2023—an increase of 38%. We anticipate that the escalating trend in cyberattacks will continue, necessitating enhanced cybersecurity protocols and heightened awareness of phishing threats. Furthermore, predictions from (cultureddata.net) indicate a surge in QR code-related fraud, where scammers trick victims into divulging their personal and sensitive payment details. This incident highlights the pressing necessity to improve public knowledge and vigilance. Research indicates that the efficacy of phishing assaults is attributed to the challenge of recognizing deceptive websites, since several consumers fail to adequately scrutinize URLs during online navigation. Consequently, identifying phishing websites has emerged as a vital

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

responsibility for cybersecurity experts who are engaged in creating sophisticated methods to safeguard consumers against these escalating attacks.

1.1. What is phishing?

Throughout the years, cybersecurity experts and researchers have offered numerous definitions of "phishing," examining it from diverse viewpoints. Phishing is ever developing, resulting in no singular, set definition; rather, definitions fluctuate based on context and use.

One study [4] characterizes phishing as “a deceptive process that entails constructing a counterfeit version of a legitimate webpage to mislead the user into revealing personal or financial information, including passwords.” Other academics [5] characterize it as "a variant of online identity theft intended to obtain sensitive information, including bank account details and credentials for financial services".

To have a thorough understanding of phishing, it is crucial to provide a complete definition that encompasses its various characteristics. This study offers a novel definition of phishing, characterizing it as an assault that merges persuasive techniques and emotional manipulation with technical expertise to mislead the target. The assailant depends on gaining the victim's confidence and persuading them, subsequently using this trust to deliver a threat or incite the victim to undertake a detrimental action, such as disclosing critical information or executing a perilous command. This definition provides a more lucid understanding of the assault method, rendering it beneficial for researchers and cybersecurity experts.

1.2. What is the mechanism of phishing?

Figure 1 delineates the typical progression of a phishing attempt, which generally unfolds in multiple stages [6]:

- **Data Acquisition:** The assailant initiates the process by aggregating information regarding the target's online conduct and digital patterns to ascertain the most efficacious method of manipulation.
- **Assault Strategy:** Following the accumulation of adequate intelligence, the assailant determines the suitable means of attack, be it via email, counterfeit websites, or SMS communications.
- **Preparation Phase:** During this phase, the attacker discerns vulnerable flaws to entice the victim, such as creating a counterfeit website that closely mimics a reputable source.
- **Execution of the Attack:** Upon finalizing preparations, the attacker implements the strategy and anticipates the victim's interaction, such as inputting personal data or downloading a harmful file.
- **Data Exploitation:** Ultimately, the perpetrator employs the acquired data—either by monetizing it or leveraging it in more fraudulent activities.

These phases encapsulate the essence of phishing operations, illustrating how perpetrators exploit trust and influence victims to fulfill their nefarious aims.

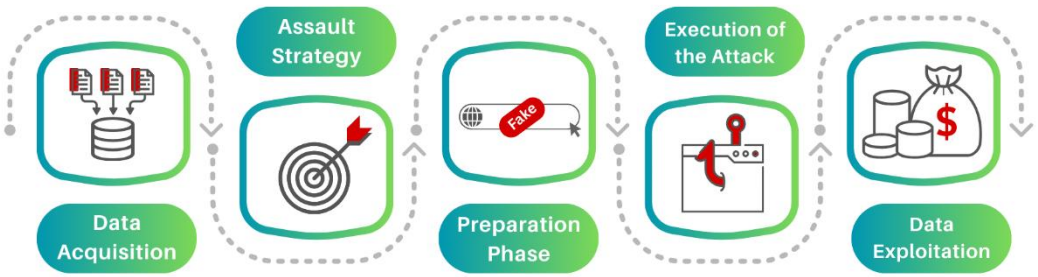

Figure 1: Phishing Process Flow and Phases

This research seeks to build an effective method for detecting and countering phishing by amalgamating various techniques to identify phishing websites through deep learning. A range of deep learning models will be methodically assessed and contrasted.

We have structured this research as follows:

Section 2 provides a literature review .Section 3 examines the architectural underpinnings of the suggested methodology .Section 4 delineates the execution and assessment of the suggested methodology .Section 5 presents conclusions and outlines directions for future research.

2. Literature Review

This section examines prior research and evaluates the efficacy of different models in combating phishing attempts. This analysis is crucial as it highlights significant concepts and pertinent academic research, thereby augmenting existing knowledge and improving strategies for mitigating phishing dangers. Additionally, it seeks to elucidate the various methodologies utilized in previous studies, thereby enhancing the suggested framework's effectiveness in addressing such concerns.

A study by [7] assessed the efficacy of deep learning algorithms in identifying phishing websites. The research examined multiple models, including Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and more deep learning architectures. It evaluated various methodologies, including URL analysis, webpage content examination, and user behavior surveillance, to determine their efficacy in detecting cyber risks. The authors advocated for the creation of hybrid models that combine deep learning with conventional AI methods to improve precision and efficacy in identifying complex phishing assaults. In another study [8], researchers developed a system that uses deep learning to detect phishing by employing three main models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a combined LSTM-CNN model. The research concentrated on categorizing web URLs as phishing or legitimate, with the objective of enhancing the precision and efficacy of phishing detection systems. The study emphasized problems like the necessity for extensive and balanced datasets, as well as the significance of enhancing model performance to minimize training and prediction duration. The authors highlighted the need for better hybrid models and advanced learning methods, like adaptive learning and federated learning, to tackle the changing nature of phishing attacks and improve cybersecurity protections. Study [9] introduced a model utilizing Artificial Neural Networks (ANN) for the detection of phishing websites. The model included 17 input features, comprising one hidden layer and an output layer with two neurons for binary classification. The experimental results indicated that the model attained an accuracy of 92.48% on both training and testing datasets, demonstrating its efficacy in differentiating between authentic and phishing websites. The authors found that better choosing features and improving the network design could boost phishing detection accuracy and suggested creating advanced methods to keep up with the constantly changing phishing threats. Research [10] validated that machine learning and deep learning methodologies are effective instruments for identifying fraudulent activity. Nonetheless, they encounter obstacles such as data imbalance and the complexity of interpreting deep models. The study advocated for enhancing model accuracy and interpretability via strategies like resampling and weight adjustment to rectify imbalanced data. The authors proposed utilizing modern methodologies such as reinforcement learning to improve the efficacy of fraud detection. Study [11] created a browser plugin to safeguard users from phishing attacks by employing deep learning to analyze dubious URLs. A sophisticated alert mechanism was established to identify phishing websites and notify users prior to their engagement with them. The research examined 651,191 URL samples categorized as either malicious or valid and evaluated various machine and deep learning models. It advised augmenting the dataset to bolster the model's flexibility to novel threats and raise the efficacy of phishing detection.

Previous research shows that phishing detection methods have improved using machine and deep learning, pointing out the need for combined models that mix deep learning features with traditional AI techniques to boost accuracy and efficiency. They also stress ongoing challenges, such as uneven data and the limited understanding of deep models, which require better performance through adaptive and federated learning methods. They also point out ongoing challenges, such as uneven data and the difficulty in understanding deep models, which require better performance through flexible and shared learning methods. This paper presents an enhanced hybrid model to efficiently combat phishing threats, emphasizing dataset growth and the model's adaptability to contemporary cyberattacks.

3. Proposed Methodology

This study primarily examines phishing attempts, which are increasingly complicated and persistent. With the rising frequency and complexity of these attacks, the demand for comprehensive anti-phishing solutions has reached unprecedented levels. To tackle this difficulty, we present a phishing detection methodology utilizing a hybrid CNN-LSTM-Attention model.

The suggested model uses the strong optimization abilities of the Adam method to improve the hyperparameters of the deep learning structure, aiming to achieve the best detection results [12]. The methodology includes critical preprocessing stages such as data cleaning, feature extraction, data balancing, and data augmentation through the introduction of random noise. We meticulously structure these procedures to optimize each strategy's efficacy while preserving the impact of others.

The model converts raw URL data into superior input features. This technique produces essential elements that encapsulate the vital information necessary to effectively differentiate between authentic and phishing URLs**. Figure 2** depicts the comprehensive architecture of the proposed model along with the preprocessing pipeline.

The model aims to make phishing detection more accurate and reliable by using convolutional layers (CNN) to find local patterns, recurrent layers (LSTM) to learn from sequences over time, and an attention mechanism to focus on important parts of the input. This hybrid design allows the system to accurately detect nuanced and intricate patterns in URL architectures that frequently signify phishing attempts.

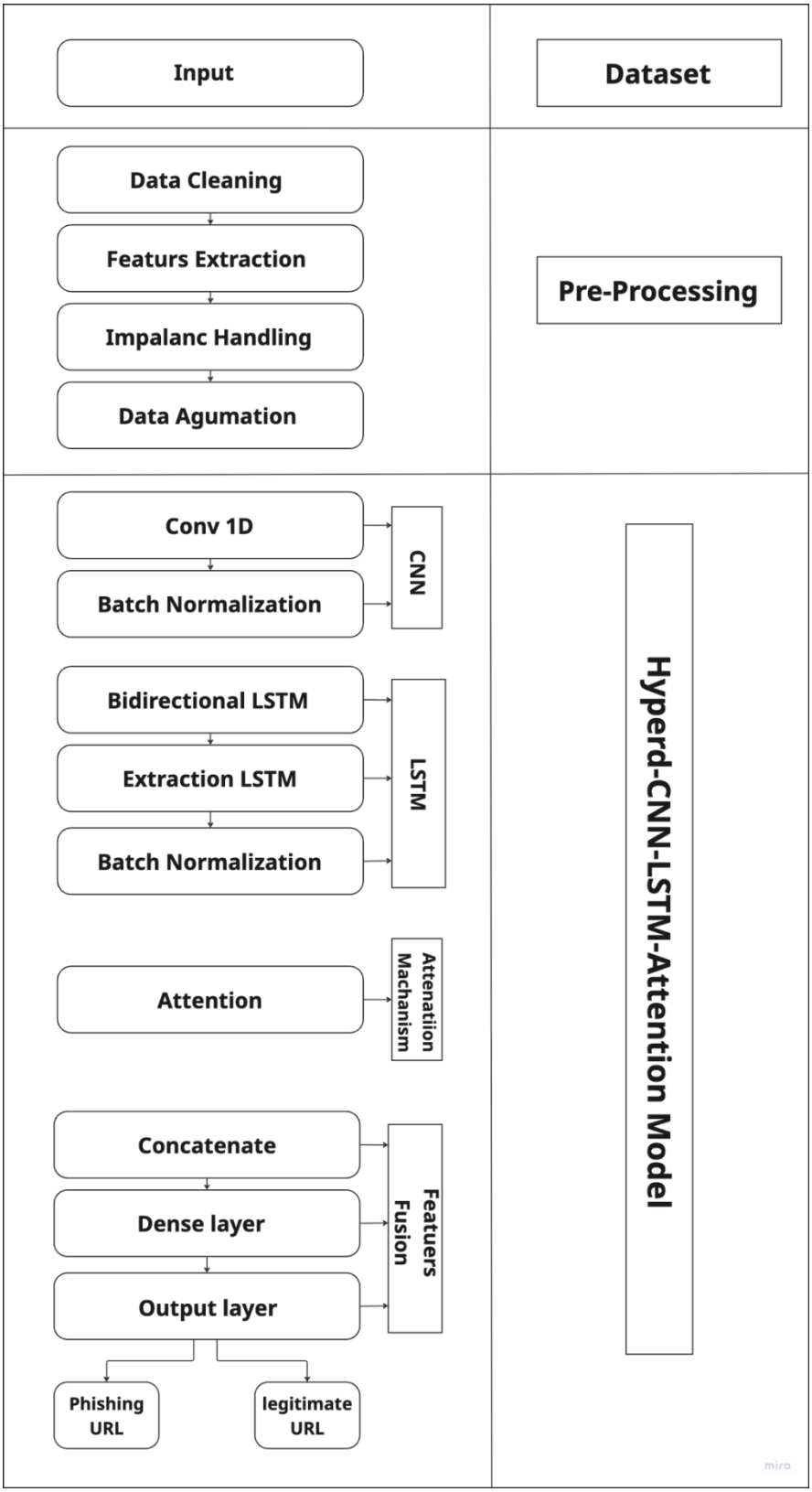

Figure 2: Hyperd-CNN-LSTM-Attention Model

3.1. Principal Improvements in the Proposed Methodology

This study's proposed methodology presents several significant improvements designed to boost the detection accuracy and generalization capacity of the phishing URL classification model:

3.1.1. Enhancements in Preprocessing:

The data pretreatment procedures—including data cleaning, feature extraction, data balancing, and augmentation—substantially improve the model's capacity to discern significant patterns. The Synthetic Minority Oversampling Technique (SMOTE) was utilized to equilibrate class distributions through the oversampling of the minority class, enhancing model performance in underrepresented categories and reducing bias toward the majority class [13]. Furthermore, random noise injection (data augmentation) was employed as a method to augment data diversity and mitigate overfitting, thereby improving the model's generalization of novel data [14]. These preprocessing strategies jointly enhance the training dataset and augment the model's resilience in detecting phishing URLs.

3.1.2. Feature Extraction Utilizing Convolutional Neural Networks:

The convolutional neural network (CNN) layer does localized feature extraction using convolutional operations, collecting spatial correlations within sequential data. The Conv1D layer was utilized to extract spatial patterns from raw URL data, and the BatchNormalization layer was incorporated to expedite training and stabilize learning. These elements render CNN an effective instrument for analyzing intricate time-series-like input configurations, such as URLs.

3.1.3. Temporal Pattern Recognition with Bi-LSTM:

A Bidirectional Long Short-Term Memory (Bi-LSTM) layer was used to understand the timing patterns in the data, allowing the model to look at sequences both forwards and backwards. An additional LSTM layer was added to improve the understanding of sequence patterns, and BatchNormalization was used again to make training more stable and reduce changes within the model.

3.1.4. The attention mechanism:

dynamically allocates weights to various segments of the input sequence, allowing the model to concentrate on the most pertinent aspects. This selective emphasis improves the model's capacity to discern complex links within the data. Additionally, using concatenation layers (Concatenate) helps combine features from both CNN and LSTM parts effectively, improving how well the model represents the features.

3.1.5. Optimization with Adam:

Due to architecture's dependence on CNN for local feature extraction, LSTM for sequence modeling, and attention for contextual weighting, the Adam optimizer is an appropriate selection. Adam adeptly manages the rapid and accurate parameter adjustments necessary for deep neural networks, optimizing both training velocity and convergence efficacy. It has demonstrated significant efficacy in phishing detection tasks, where adaptation to varied input patterns is essential [13].

The proposed hybrid model overcomes the constraints of previous methods that depended exclusively on CNN or LSTM by using comprehensive preprocessing techniques, CNN for local feature extraction, LSTM for temporal modeling, and attention for feature prioritizing. This architecture aims to augment the identification of phishing URLs and bolster web security by alleviating the threats posed by fake websites.

3.2. Evaluation Metrics

A validation dataset is employed to acquire an impartial assessment of the model's efficacy throughout the training phase. We utilize various performance evaluation metrics to thoroughly evaluate the classification efficacy of the proposed hybrid deep learning model (CNN-LSTM-Attention). These parameters are crucial for assessing the efficacy, accuracy, and resilience of the algorithm in identifying phishing URLs. The following are the principal metrics employed for performance assessment:

3.2.1. Accuracy

Accuracy is one of the most commonly used metrics in evaluating classification models. It measures the ratio of correctly predicted observations to the total number of predictions made. The formula for accuracy is given by:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

- **TP** (True Positives): Number of correctly predicted phishing URLs
- **TN** (True Negatives): Number of correctly predicted legitimate URLs
- **FP** (False Positives): Number of legitimate URLs incorrectly classified as phishing
- **FN** (False Negatives): Number of phishing URLs incorrectly classified as legitimate

A higher accuracy value generally indicates a better-performing model. However, in imbalanced datasets—such as phishing detection tasks—accuracy alone may not provide a complete picture, and other metrics such as precision, recall, and F1-score are often required for a more comprehensive evaluation.

3.2.2. Precision

Precision measures the proportion of positive identifications that were correct. In the context of phishing detection, it reflects how many of the URLs identified as phishing are truly malicious. The formula for precision is:

$$\text{Precision} = \frac{TP}{TP+FP}$$

A high precision value indicates a low rate of false positives, which is critical for minimizing the risk of falsely classifying legitimate websites as phishing attempts.

3.2.3. Recall (Sensitivity)

Recall, also known as sensitivity or true positive rate, quantifies the model's ability to correctly identify actual phishing URLs. It is defined as:

$$\text{Recall} = \frac{TP}{TP+FN}$$

A high recall means the model successfully detects most of the phishing URLs, reducing the risk of false negatives—cases where malicious sites go undetected.

3.2.4. F1-Score

The F1-score is the harmonic mean of precision and recall. It provides a single metric that balances both false positives and false negatives, especially useful in cases where the dataset is imbalanced:

$$\textbf{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

A higher F1-score indicates that the model maintains a good balance between precision and recall, which is essential for robust phishing detection systems.

Within the framework of the proposed hybrid model for phishing website classification, which integrates convolutional neural networks (CNN), long-short-term memory networks (LSTM), and the attention mechanism, precision, recall, and the F1-score are deemed more critical evaluation metrics than overall accuracy. This difference is chiefly attributable to the data imbalance between phishing and legitimate websites, which may result in deceptive outcomes when depending exclusively on overall accuracy.

Emphasizing precision reduces false positives, meaning good websites misidentified as phishing, thus eliminating unwanted notifications and maintaining user trust. Conversely, optimizing recall improves the model's capacity to identify a greater number of genuine phishing websites, which is crucial in cybersecurity applications.

The F1 score combines precision and recall into one number, making it a better way to measure how well a model works when there is an imbalance in classes. The F1 score is the most appropriate metric for assessing the efficacy of the suggested model, as it precisely indicates the algorithm's ability to identify phishing websites while minimizing false positives.

4. Experiments and Results

4.1. Dataset

In our research, we utilized a publicly accessible and current dataset that provides a thorough depiction of websites. This is essential because phishing websites typically have a short lifespan, existing for only a few days or even hours, as shown in previous studies [15]. Consequently, it is essential that phishing datasets are regularly updated and accessible online.

Consequently, we employed a current dataset for phishing detection sourced from Kaggle, referred to as the Url_Detection_Dataset. This dataset is meticulously maintained to facilitate studies in phishing detection and URL analysis. It comprises 822,010 URL samples, along with a collection of extracted attributes designed for the identification of phishing activity. The dataset categorizes URLs as either harmful or legitimate, rendering it extremely appropriate for training machine learning models to differentiate between phishing and secure websites. Moreover, it facilitates a diverse array of applications, encompassing phishing detection systems, malicious behavior analysis, and the enhancement of cybersecurity tools [16].

4.2. Analysis of Results

Our studies utilized a dataset including 822,010 URLs, which contained 427,028 phishing websites and 394,982 real websites. Subsequent to preparation procedures including data cleansing and feature extraction, the dataset was partitioned into 80% for training and 20% for testing.

The outcomes, detailed in **Table 1**, illustrate the efficacy of various deep learning models utilized for the classification task: Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and the suggested Hybrid CNN-LSTM-Attention model. The models were assessed using four primary performance metrics: accuracy, precision, recall, and F1-score. These metrics offer a thorough evaluation of each model's prediction efficacy and its equilibrium in accurately detecting positive and negative instances.

Models	Accuracy	Precision	Recall	F1-Score
ANN	94.86%	94.01%	95.79%	94.89%
CNN	94.75%	93.52%	96.13%	94.81%
RNN	92.15%	90.59%	94.01%	92.27%
LSTM	93.18%	91.56%	95.07%	93.28%
Hybrid-CNN-LSTM-Attention	94.16%	92.55%	96.01%	94.25%

Table 1: Classification Result

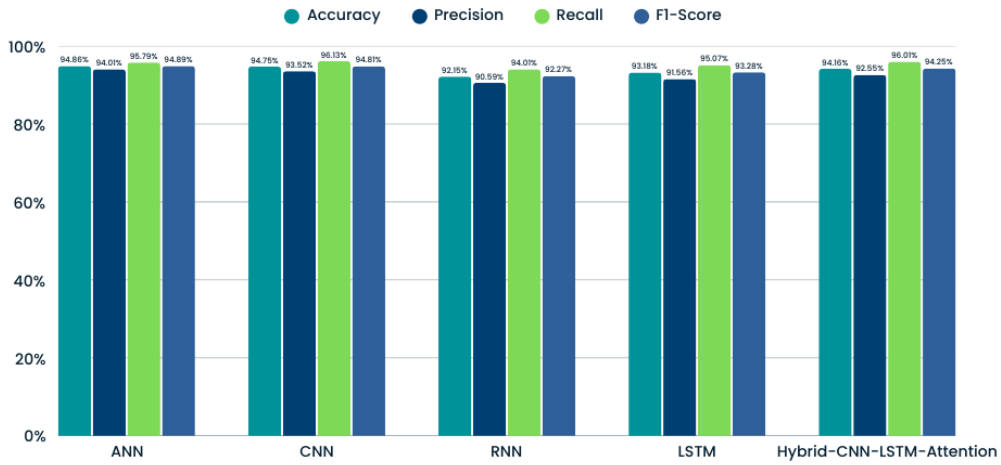

Figure 3: Comparison of models in detecting phishing URL

In terms of overall accuracy, defined as the ratio of correct predictions to the total number of cases, the ANN model attained the highest result at 94.86%, closely followed by the CNN model at 94.75%. The proposed hybrid model achieved an accuracy of 94.16%, marginally lower than the leading two models, although it still surpassed both RNN and LSTM. While not the pinnacle of accuracy, this performance remains within a high-performance spectrum, demonstrating the hybrid model's robust capacity to accurately categorize the majority of data. Regarding precision—the ratio of true positive predictions to all positive predictions—the ANN model achieved 94.01%, while the hybrid model attained 92.55% in second place. This result shows that the hybrid model effectively lowers false positives, making it particularly suitable for situations where it's crucial to avoid mistakenly labeling real situations as threats. Conversely, recall, which assesses the model's capacity to recognize true positive events, demonstrates a significant advantage for the hybrid model. It achieved a score of 96.01%, the second highest following CNN's 96.13%. The hybrid architecture exhibits significant sensitivity to phishing incidents, guaranteeing minimal oversight—an imperative characteristic in vital sectors like cybersecurity or healthcare, where neglecting positive cases can have severe repercussions. The ANN achieved the highest F1-score ranking at 94.89%, followed by the CNN at 94.81% and the hybrid model at 94.25%, in the evaluation that combines precision and recall into a single metric. Despite the hybrid model not attaining the highest score, the discrepancy of 0.0064 from the optimal model is statistically insignificant and does not undermine its scientific or practical significance. The balanced performance exhibited by the hybrid model across all evaluation criteria is more significant than the individual measures. The model uses a sophisticated design that includes convolutional layers (CNN) to pick up important features in Phishing URLs, LSTM units to understand time-related patterns, and an attention mechanism to highlight key information, allowing it to handle complicated and sequential data better than older models that rely on just one component. The results show that the hybrid model aims to do well in numerical performance while also creating a design that enhances prediction accuracy in different situations. Its ability to maintain dependable performance, particularly in challenging datasets, supports its use as a strong and flexible model for real-world phishing detection. Its ability to maintain consistent performance, particularly in challenging datasets that are noisy, unbalanced, or complex, supports its use as a strong and flexible model for real-world phishing detection applications. Such performance further reinforces its position as a fundamental contribution of the given research.

5. 5. Future Work

Real-Time Testing on Live Data: Although the current experiments utilize historical datasets, it is essential to assess the proposed model in real-time scenarios. Future endeavors should focus on implementing the model using live streaming URL data to evaluate its efficacy in dynamic settings. This evaluation encompasses the assessment of latency, throughput, and detection accuracy in a practical deployment context. Furthermore, incorporating the model into a real-time phishing detection system could yield significant information regarding its responsiveness and resilience against novel and developing threats.

Enhancement of Multi-Level Classification Systems: The existing binary classification method—differentiating between dangerous and benign URLs—can be improved by implementing a multi-level classification framework. This would enable the model to classify URLs into more precise threat categories, such as phishing, malware, and spam.

- Phishing
- Malware Dissemination
- Monetary Deception
- Unsuitable or Malevolent Content

This detailed classification makes it easier to understand model predictions and helps cybersecurity systems respond better by tailoring countermeasures to the specific type of threat identified. Future research may investigate methods for dataset enhancement and labeling to facilitate comprehensive categorization.

6. 6.Conclusion

The findings of this study illustrate the efficacy of the suggested hybrid model (Hybrid-CNN-LSTM-Attention) in phishing URL classification. The model employs a comprehensive approach to malicious URL detection by integrating spatial feature extraction through Convolutional Neural Networks (CNN), temporal dependency modeling via Long Short-Term Memory (LSTM) units, and selective focus utilizing the Attention Mechanism.

Despite not achieving the maximum score in any individual metric, the model continuously demonstrated balanced and robust performance across all principal evaluation indicators, including accuracy, precision, recall, and F1-score. This indicates its robust adaptability to imbalanced and intrinsically complicated data distributions, a prevalent trait in real-world cybersecurity contexts.

Furthermore, the implementation of preprocessing techniques—such as data balancing through the Synthetic Minority Over-sampling Technique (SMOTE) and data augmentation by incorporating random noise—substantially enhanced the model's generalization capacity and improved the classification of underrepresented classes, while also reducing bias.

The proposed method shows a clear improvement in performance and provides a flexible framework that sets the stage for future progress in smart phishing detection systems. The research results show that the model is an important addition to the field and help encourage more studies to make cyber defense systems better at dealing with new online threats.

7. 7. References

[1] Rao, R. S., Vaishnavi, T., & Pais, A. R. (2020). CatchPhish: detection of phishing websites by inspecting URLs. Journal of Ambient Intelligence and Humanized Computing, 11(2), 813-825.

[2] Jha, A. K., Muthalagu, R., & Pawar, P. M. (2023). Intelligent phishing website detection using machine learning. Multimedia Tools and Applications, 82(19), 29431-29456.

[3] Prasad, Y. B., & Dondeti, V. (2025). PDSMV3-DCRNN: A novel ensemble deep learning framework for enhancing phishing detection and URL extraction. Computers & Security, 148, 104123.

[4] Van der Merwe, A., Loock, M., & Dabrowski, M. (2005, January). Characteristics and responsibilities involved in a phishing attack. In Proceedings of the 4th international symposium on Information and communication technologies (pp. 249-254).

[5] Kirda, E., & Kruegel, C. (2005, July). Protecting users against phishing attacks with antiphish. In 29th Annual International Computer Software and Applications Conference (COMPSAC'05) (Vol. 1, pp. 517-524). IEEE.

[6] Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060.

[7] Do, N. Q., Selamat, A., Krejcar, O., Herrera-Viedma, E., & Fujita, H. (2022). Deep learning for phishing detection: Taxonomy, current challenges and future directions. Ieee Access, 10, 36429-36463.

[8] Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232.

[9] Zhu, E., Ju, Y., Chen, Z., Liu, F., & Fang, X. (2020). DTOF-ANN: an artificial neural network phishing detection model based on decision tree and optimal features. Applied Soft Computing, 95, 106505.

[10] Gandhar, A., Gupta, K., Pandey, A. K., & Raj, D. (2024). Fraud detection using machine learning and deep learning. SN Computer Science, 5(5), 453.

[11] Linh, D. M., Hung, H. D., Chau, H. M., Vu, Q. S., & Tran, T. N. (2024). Real-time phishing detection using deep learning methods by extensions. International Journal of Electrical and Computer Engineering (IJECE), 14(3), 3021-3035.

[12] Khatun, M. A., Yousuf, M. A., Ahmed, S., Uddin, M. Z., Alyami, S. A., Al-Ashhab, S., ... & Moni, M. A. (2022). Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. IEEE Journal of Translational Engineering in Health and Medicine, 10, 1-16.

[13] Sha, L., Raković, M., Das, A., Gašević, D., & Chen, G. (2022). Leveraging class balancing techniques to alleviate algorithmic bias for predictive tasks in education. IEEE Transactions on Learning Technologies, 15(4), 481-492.

[14] Poojary, R., Raina, R., & Mondal, A. K. (2021). Effect of data-augmentation on fine-tuned CNN model performance. algorithms, 5, 6.

[15] Lee, K., Lim, K., Kim, H., Kwon, Y., & Kim, D. (2025, April). 7 Days Later: Analyzing Phishing-Site Lifespan After Detected. In Proceedings of the ACM on Web Conference 2025 (pp. 945-956).

[16] URL-Detection Dataset, Kaggle. https://www.kaggle.com/datasets/vishvapatel09/url-detection-dataset