THOMAS ADEWUMI UNIVERSITY, OKO, KWARA STATE
Science | Technology | Medicine

Thomas Adewumi University
Journal of Innovation,
Science and Technology (TAU-JIST)

ISSN: 3043-503X

RESEARCH ARTICLE

DEVELOPMENT OF HYBRIDIZED CNN-BIGRU FRAMEWORK FOR DETECTION OF WEBSITE PHISHING ATTACKS

*Raji Abdullahi Egigogo[1]

[1]Al-Qalam University Katsina, (Fcapt), Kano State Nigeria.

Corresponding Author: rajiaegigogo@auk.edu.ng

ARTICLE DETAILS

Article History:
Received 02 July 2024
Accepted 05 October 2024
Available online 10 December 2024

ABSTRACT

Phishing attacks continue to pose a significant threat to cybersecurity by deceiving individuals into revealing sensitive information. Traditional detection methods often fail to counteract the sophistication of modern phishing tactics. This study introduces a hybrid framework that combines CNN and BiGRU to improve phishing detection. The CNN component extracts spatial features from website data, while the BiGRU component analyzes temporal sequences to identify phishing patterns. The framework was evaluated using a publicly available dataset that achieved a remarkable accuracy of 99.96%, precision of 99.92%, recall of 100%, and an F1-score of 99.92%. These results demonstrate significant improvement over existing methods, highlighting the framework's effectiveness and reliability in real-world cybersecurity applications.

KEYWORDS

Phishing detection, Website, Convolutional Neural Networks (CNN), Bidirectional Gated Recurrent Units (BiGRU), Hybrid framework, Deep learning

1. Introduction

Phishing attacks represent a serious and ever-evolving danger in the digital landscape. These schemes deceive individuals into sharing confidential details, such as passwords, credit card information, and usernames, by masquerading as real and reliable sources (Abdulrahman et al., 2019; Ivanov et al., 2021). Web-based phishing, in particular, involves creating fraudulent websites that imitate those of reputable organizations, tricking users into entering their confidential information. The increasing sophistication of these deceptive practices necessitates advanced detection mechanisms to protect users from phishing scams (Redi & Ernasari, 2023). Traditional detection methods rely on static features and predefined rules and often struggle to identify newly created phishing sites. As phishing schemes linger to grow, there is a pressing need for more robust and intelligent detection systems (Rivki et al., 2024).

The advent of deep learning (DL) has significantly advanced the development of more effective phishing detection methods. Their ability to learn hierarchical representations of data provides substantial improvements over traditional machine-learning techniques (Dong et al., 2021; LaraBenítez et al., 2021). These models can autonomously extract features from raw data, significantly reducing the dependence on manual feature engineering and the detection system to adapt to new and evolving phishing tactics (Adebowale et al., 2023).

In the perspective of phishing detection, the proposed framework leverages CNN and BiGRU to form a robust hybrid model. CNNs are

Quick Response Code | **Access this article online**

Website: https://journals.tau.edu.ng/index.php/tau-jist

DOI: https://doi.org/10.5281/zenodo.15003367

Cite the Article: *Development of Hybridized CNN-BiGRU Framework for Detection of Website Phishing Attacks*

renowned for their ability to extract spatial features from data automatically. They identify patterns and structures within images and text, making them ideal for analyzing website elements such as URLs, HTML content, and metadata (Kulkarni, 2023; Wei et al., 2020). By utilizing CNNs, the framework can effectively capture the intricate details that differentiate phishing websites from genuine ones.

Complementing the CNNs, BiGRUs are a Recurrent Neural Network (RNN) type that excels in processing sequential data. They are particularly adept at recognizing temporal dependencies and patterns within sequences. BiGRUs, with their bidirectional processing capability, can examine data in both forward and reverse directions, offering a thorough comprehension of the temporal relationships within the input data (Khandelwal et al., 2020). This bidirectional analysis is crucial for detecting phishing websites, as the order and context of elements on a web page can reveal malicious intent.

Integrating CNNs and BiGRUs in a hybridized framework enhances the detection system's ability to identify phishing websites. The CNN component extracts detailed spatial features from the website data, while the BiGRU component analyzes the temporal sequences of these features to detect patterns indicative of phishing. This combined approach enables the system to capture both spatial and temporal characteristics of phishing websites, improving detection accuracy and robustness.

1.2 Motivation

Conventional phishing detection systems, which frequently depend on static features or simple heuristics, fall short in the face of increasingly sophisticated and adaptive phishing strategies. These conventional methods struggle to identify newly created phishing sites and adapt to evolving tactics. This necessitates a more dynamic and comprehensive approach to enhance detection capabilities. The advent of deep learning has significantly advanced phishing detection, offering models that can learn hierarchical representations of data and adapt to new phishing methods. This study aims to create a hybrid solid framework that utilizes the advantages of CNN and BiGRU to enhance the performance and improved reliability of the detection of phishing websites.

2. Related Works

The literature presents numerous machine learning and deep learning frameworks for detecting phishing attacks. Subba (2023) developed a security framework utilizing a diverse stacking ensemble approach, incorporating three base classifiers and a meta-classifier. The model processes 44 extracted features from URLs and web pages, combining the results for the final prediction. The framework demonstrated high accuracy (99% for binary, 98% for multiclass) on benchmark datasets. Tenis & Santhosh (2023) presented a real-time phishing detection system using a deep learning approach, including whitelisting and blacklisting mechanisms. The adaptive RNN (a– RNN) model showed a superior accuracy of 99.18% across different datasets.

Alsharaiah et al. (2023) proposed a novel framework integrating random forest classifiers with kmeans clustering (RM-KmC) to improve feature correlation detection. Tested on a 5,000-sample dataset, the model achieved an accuracy of 98.64% with solid precision and recall metrics. Tang & Mahmoud (2022) introduced a browser plug-in-based deep learning framework for realtime phishing detection, achieving 99.18% accuracy with the RNN-GRU model through a blend of whitelist and blacklist filtering.

Liu et al. (2021) proposed a multistage detection model using the CASE framework, which exhibited high efficiency and performance and low false alarms in extensive evaluations. Kumar & Subba (2021) introduced a lightweight framework for phishing detection, analyzing URLs to extract key features, resulting in high precision and minimal false positives. Similarly, Zeng et al. (2020). it introduced PhishBench 2.0, a robust benchmarking platform for phishing detection systems with extensive features, classifiers, and metrics. It is set to be released on GitHub for community use.

Rendall et al., (2020) worked on a multi-layered detection framework that classifies phishing domains multiple times, achieving performance on par with leading detection systems. Sadique et al. (2020) presented a real-time phishing URL detection framework that achieved 87% accuracy, suggesting incremental learning techniques to improve detection effectiveness.

Hr et al. (2020) presented a browser-embedded anti-phishing system using a rule-extraction method paired with Random Forest Classification, reaching 99.36% accuracy for real-time phishing detection. Saravanan & Subramanian (2020) developed a framework for phishing detection that effectively extracts and selects features from websites, enhancing classification accuracy and outperforming current methods in experimental evaluations.

Elnagar & Thomas (2019) introduced a cognitive detection framework combining BLSTM-RNN and CNN models, incorporating image recognition to enhance the identification of phishing websites. (Rao & Pais 2019) discussed a machine learning-based framework employing heuristic features from URLs and source code, achieving 99.31% accuracy of the Random Forest algorithm for phishing detection.

Cuzzocrea et al. (2019) suggested a decision tree-based machine learning framework identifying and evaluating phishing assaults, exhibiting good performance in experimental evaluations. A PhishMon framework based on machine learning was created utilizing fifteen unique features, achieving 95.4% accuracy in detecting phishing sites with a low false positive rate of 1.3%. (Niakanlahiji et al., 2018). Yi et al. (2018) examined a Deep Belief Network (DBN) based framework for phishing detection, achieving a 90% true positive rate and minimizing false positives to 0.6%. Park et al. (2017) developed Phishing-Detective. This framework uses web scraping and data mining to detect phishing websites through heuristic analysis, though its performance may be affected by changing phishing strategies.

Table 1 Summary of the literature review

Author & Year	Methodology/ Algorithm	Results	Strengths	Limitations
Subba (2023)	Heterogeneous stacking ensemble; 3 base classifiers, 1 metaclassifier (FCNN)	99% accuracy (binary), 98% accuracy (multiclass)	High accuracy, comprehensive feature extraction	Increased computational complexity
Tenis & Santhosh (2023)	adaptive Recurrent Neural Networks (a–RNN)	99.18% accuracy	Reduced false positives, high detection accuracy	Complexity in realtime implementation
Alsharaiah et al. (2023)	Random Forest integrated with k-means clustering(RM-KmC)	98.64% accuracy, high precision and recall	Enhanced feature correlation detection	It may require significant computational resources
Tang & Mahmoud (2022)	Deep learning with whitelist/blacklist filtering; RNN-GRU model	99.18% accuracy	Effective real-time detection	Reliance on blacklisting could miss new phishing sites
Liu et al. (2021)	Multistage detection model; CASE framework	High efficiency, low false alarms	Short execution times	It may not handle all phishing attack types
Kumar & Subba (2021)	Lightweight machine learning; URL feature extraction	High precision, low false positive rate	Efficient with minimal resources	May miss more sophisticated phishing attacks
Sadique et al. (2020)	Real-time phishing URL detection	87% accuracy	Real-time capability suggests incremental learning	Moderate accuracy; further improvement is needed
Zeng et al. (2020)	PhishBench 2.0: benchmarking framework for phishing detection	Offers over 250 features, 12 classifiers, 17 metrics	Comprehensive feature and classifier set	High complexity for users unfamiliar with the framework
HR et al. (2020)	Browser-embedded system; rule-extraction; Random Forest	99.36% accuracy in real-time detection	Real-time phishing detection	Potential browser compatibility issues
Rendall et al. (2020)	Multi-layered detection; supervised machine learning	Comparable to state-of-the-art systems	Multi-tiered classification improves accuracy	Added complexity in classification processes
Saravanan & Subramanian (2020)	Feature selection and extraction; phishing detection module	Outperformed existing classifiers in experimental tests	Efficient feature selection enhances detection accuracy	Potential generalization issues with unseen d
Elnagar et al. (2019)	BLSTM-RNN and CNN with image recognition	Enhanced phishing detection	Cognitive approach with dual models	Computationally intensive
Rao & Pais (2019)	Feature-based machine learning; Random Forest	99.31% accuracy	Effective heuristic feature extraction	Dependence on third-party services
Cuzzocrea et al. (2019)	Decision tree-based machine learning	High accuracy in detecting phishing attacks	Simple and effective machine learning approach	May struggle with very dynamic phishing methods
Yi et al. (2018)	Deep Belief Networks (DBN)	90% true positive rate, 0.6% false positive rate	High accuracy in identifying phishing sites	Limited testing environment
Niakanlahiji et al. (2018)	PhishMon: machine learning with 15 novel features	95.4% accuracy, 1.3% false positive rate	Low false positives, novel feature set	Potentially complex implementation
Park et al. (2017)	Web scraping and data mining; heuristic analysis	Effective in detecting phishing websites	Dynamic approach	May be impacted by evolving phishing tactics

3. Proposed Framework

3.1 Overview

The proposed hybrid framework integrates CNN and BiGRU networks to leverage their complementary strengths. CNNs are employed for feature extraction, and BiGRUs are used for sequence modelling, resulting in a robust detection system. The framework consists of four (4) states, namely input, preprocessing, deep learning and detection stage, each having distinct features tailored towards the same purpose. The following is a thorough explanation of the stages depicted in Figure 1.

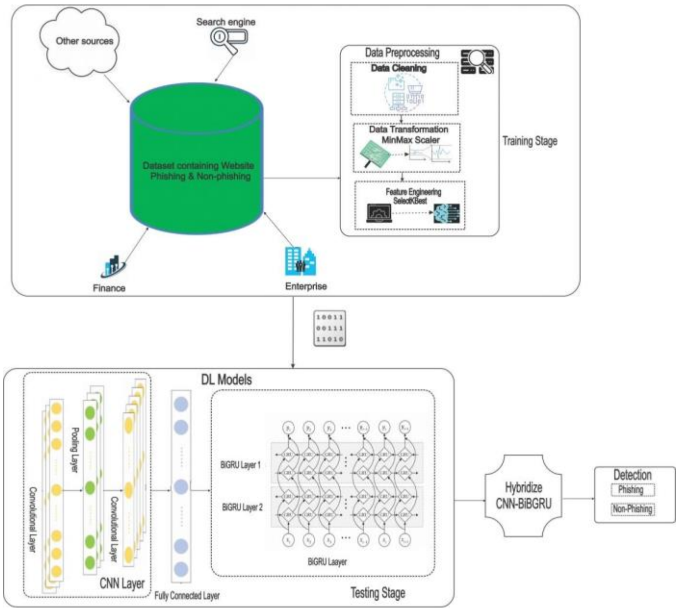


Figure 1: Hybridized CNN-BiGRU Framework for the detection of website phishing attack

I. In the first stage of this framework, the data input comprises phishing websites from different angles, such as enterprise, finance, search engine, and other sources, which will be fetched in the convolutional layer. This study employed datasets collected from IEEE Data port because they are publicly available and contain information on phishing attacks reported by users, the largest data science community in the world, offering strong tools and resources to support researchers in achieving their data science objectives. A community of security experts verifies it, and it has been widely used by various authors in their research (Liu *et al.,* 2019; Wang *et al.,* 2019; Do *et al.,* 2021; Srinivasan & P, 2023; Sajidha 2023).

II. Data Preprocessing Stage: During the data preprocessing stage, three crucial phases were carried out: data cleaning, transformation, and feature engineering. Data cleaning, an indispensable aspect of preprocessing, involved identifying and rectifying inconsistencies, errors, and irrelevant data within the dataset to enhance its quality and prepare it for utilization in deep learning models. Following data cleaning, a Minmax scaler was applied to transform the data, aiming to improve its compatibility with deep learning algorithms, preserve the original distribution's shape, and mitigate the influence of feature scales on the optimization process.

Subsequently, the dataset was divided into two groups, with 20% set aside for testing and the remaining 80% going toward training and validation. Using equation 1 below, the values were converted to the testing set and scaled between 0 and 1 using the min-max scaler, which was fitted to the training set (Chou *et al.,* 2023; Huang, Peng & Wu, 2021).

$$MinMaxSclaer\,(v'i) = \frac{i - min_A}{max_A - min_A}\,(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Where $x_i$represents the ith value, max_A and min_A Denote a feature's maximum and minimum values and $new\ \_\ max_A$ and $new\ \_\ min_A$ are the values 0 and 1, respectively. Feature engineering ensued, incorporating a feature selection technique known as Select K-Best (SelectKBest) to enhance model performance, generalization, and computational efficiency and mitigate the impact of irrelevant or noisy features. The score was determined utilizing equation 3.2, as proposed by (Olatunji *et al.,* 2023; Sharifai & Zainol, 2020; Thuy & Wongthanavasu, 2022).

$$x^2 = \sum_{i=1}^{n} \frac{(OFi - EFi)}{EFi} \tag{2}$$

Where OFi is the frequency observed for the feature $F's\ i$ —th value, and EFi is the frequency anticipated for feature $F's\ i$ — th value. The dataset's description and correlation heatmap are presented in Table 3 and Figure 2, respectively.

III. DL Stage: A subset of machine learning, has garnered significant attention in recent years due to advancements in processing power and expanded data storage capacities. These developments have greatly facilitated the application of DL methodologies, which have demonstrated remarkable efficacy across various domains, including image processing, natural language processing, and machine translation, particularly when handling large datasets. Leveraging these advantages, our study adopts two prominent DL algorithms: CNN and BiGRU. The selection of these algorithms is based on the belief that combining different approaches enhances overall accuracy, as demonstrated by (Do *et al.,* 2021; Gupta *et al.,* 2018).

Within this multiclass framework, the CNN component is tasked with extracting highlevel features from the input dataset and is adept at capturing local patterns and features. Subsequently, the BiGRU component sequentially processes these features, considering sequential dependencies and temporal features across different dataset segments. The integration of a fully connected layer atop the BiGRU facilitates final classification. This framework is poised to enhance the accuracy and effectiveness of phishing attack classification by harnessing the complementary strengths of the CNNBiGRU architectures.

Let $X = \{x_1, x_2, \ldots, x_n,\ \}$ represent the input sequence, where x_i, is the word embedding of the feature in the sequence

$$z_i = f\,(\textstyle\sum_{j=1}^{m} w_j \,.\, x_{i+j-1} + b^l\,) \tag{3}$$

Where z_i^l is the output of the I - th convolutional filter at position i, w_j^l are the filter weights, is the bias term, and f is the activation function ReLU.

$$p = max\,(z_1, z_2, \ldots, z_n\ \ m+1) \tag{4}$$

Where p^l the max-pooling layer for the I - th filter

$$h_t = GRU(x_t, h_{t-1}\,) \tag{5}$$

$$\overleftarrow{h}_t = GRU(\underset{\leftarrow}{x_t}, \overleftarrow{h}_{t+1}) \tag{6}$$

$$h_t = [\vec{h}_t\ ;\ \bar{h}_t] \tag{7}$$

Where [. ; .] denotes concatenation

$$h = h^1, h^2, \ldots, h^L \tag{8}$$

where h^l is the output feature vector from the l – th convolutional filter or BiGRU layer.

$$z = hW + b \tag{9}$$

Where W is the weight matrix and b is the bias vector.

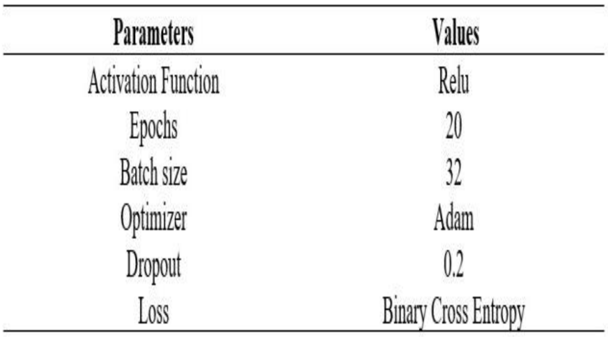
$$\hat{y} = softmaz(z) \tag{10}$$

Where $\hat{y}$ is the predicted probability distribution over the classes.

4. Experimental Setup

To evaluate the suggested hybrid model, a simulation is run on a desktop computer running Windows 11 Pro with a 64-bit operating system on an Intel(R) Core(TM) i5-6300U CPU running at 2.40GHz and 8GB of RAM. The notebook is a Jupitar Notebook 6.4.8.

Table 2: Parameters



Parameters	Values
Activation Function	Relu
Epochs	20
Batch size	32
Optimizer	Adam
Dropout	0.2
Loss	Binary Cross Entropy

4.1 Dataset

A publicly accessible phishing dataset with features taken from both legitimate and phishing websites was used for the studies. The dataset used for evaluating the model, sourced from IEEE Data Port, consists of 15,367 instances, with 7,781 non-phishing and 7,586 phishing samples, providing a balanced distribution. This balance helps prevent model bias and contributes to its strong performance. The 80 extracted features include URL attributes, HTML content, and metadata critical for the CNN-BiGRU architecture, where CNN captures spatial features and BiGRU models temporal patterns. The description of the dataset is presented in

Table 3. Table 3: Description of the Dataset

Dataset	Phishing attack type	Total number of features	Non-phishing	Phishing	Total
IEEE Data Port	Website Uls	80	7781	7586	15367

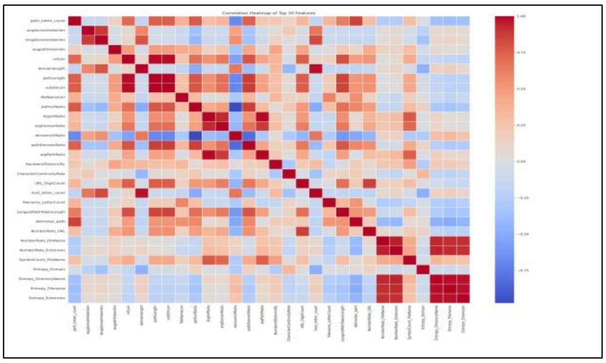


Figure 2: The correlation Heatmap Dataset employed

4.2 Evaluation Metrics

The frameworks performance was appraised using recall, accuracy, F1-score, and precision, These metrics offer an ample assessment of the model's efficiency in distinguishing between phishing and legitimate websites (Abdulrahman et al., 2019; Wabi, et al., 2024). These metrics are represented in the following equation

$$Precision = \frac{TP}{TP+FP} = +\frac{True\ Positive}{Total\ Predicted\ Positive} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} = +\frac{True\ Positive}{Total\ Predicted\ Positive} \quad (12)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} = +\frac{True\ Positive}{Total\ Predicted\ Positive} \quad (13)$$

$$F1-score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} = +\frac{True\ Positive}{Total\ Predicted\ Positive} \quad (14)$$

5. Results and Discussion

The proposed development of a hybridized CNN-BiGRU framework for the detection of website phishing attack was evaluated through an experiment with Python programming language using Jupyter Notebook tool, a cooperative web application for producing and sharing documents that combine Code, Rich Text and Visualizations. This was chosen because of its Ease of Use, Flexibility and Reproducibility. The experimentation of the model was performed with a dataset of 15367 instances, of which 7781 are Non-phishing website instances labelled as 0 and 7586 phishing website instances labelled as 1 with 30 features. To produce a robust hybridized framework, 80% of the datasets were used for training and 20% for testing in each dataset. Also, SelectKbest was used to select the best features passed into the model for training and testing. Table 4 illustrates the performance evaluation of the Hybridized CNN-BiGRU framework for this study, showing the model's performance across various measures.

Table 3: Performance Result of Hybridized CNN-BiGRU Framework for the Detection of Website Phishing Attack

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity
99.96	99.92	100	99.92	99.91

Table 4 shows the performance metrics of the Hybridized CNN-BiGRU framework for detecting website phishing attack, which exhibit remarkable effectiveness. The model boasts an accuracy of 99.96%, indicating a highly dependable classification of websites. Its 99.92% precision highlights a minimal false positive rate, accurately identifying phishing websites while seldom misclassifying benign ones. Achieving a perfect recall of 100%, the model ensures all phishing attempts are detected, leaving no malicious sites undetected. The F1-score, at 99.96%, reflects a balanced performance, integrating high precision and recall.

Additionally, the specificity of 99.91% demonstrates the model's capability to recognize benign websites, thus avoiding false alarms correctly. These metrics underscore the model's reliability and efficacy in practical cybersecurity applications. Figure 3 visualizes the performance of the CNNBiGRU Framework and its confusion matrix in Figure 4.

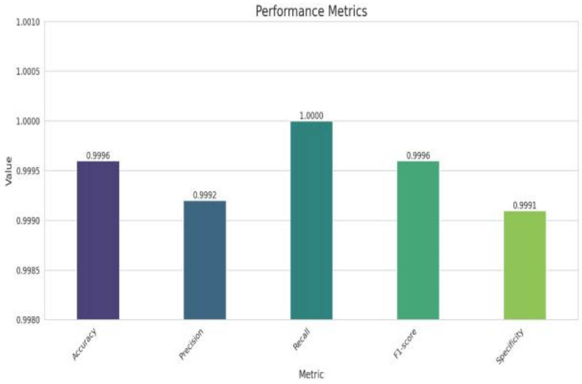


Figure 3: Visualization of the Performance of the Proposed CNN-BiGRU Framework

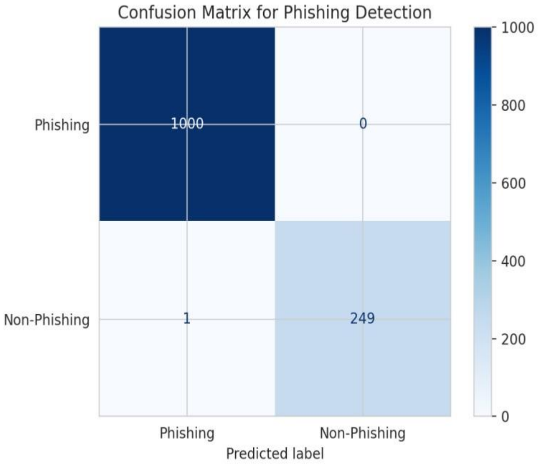
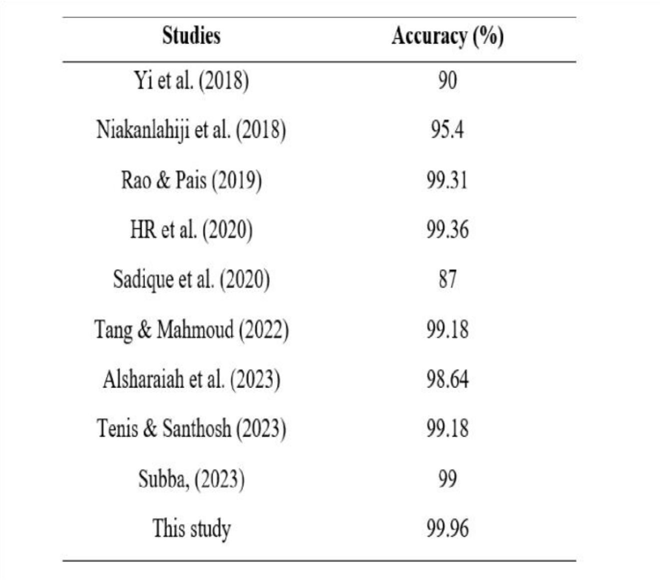


Figure 4: Confusion matrix of the Proposed CNN-BiGRU Framework

5.2 Comparative Analysis of this Study Accuracy with Existing Study Accuracies

Comparison of various studies' accuracy, precision, recall and F1_score with the current research. These studies were analyzed comparatively, focusing on the trends and overall progress as presented in Table 5 and 6 respectively.

Table 5: Comparison of the accuracy of various studies with the current research



Studies	Accuracy (%)
Yi et al. (2018)	90
Niakanlahiji et al. (2018)	95.4
Rao & Pais (2019)	99.31
HR et al. (2020)	99.36
Sadique et al. (2020)	87
Tang & Mahmoud (2022)	99.18
Alsharaiah et al. (2023)	98.64
Tenis & Santhosh (2023)	99.18
Subba, (2023)	99
This study	99.96

Table 4 compares the accuracy rates of various studies focused on website phishing attack frameworks, with the most recent research demonstrating significant advancements. Over time, there has been a clear trend of improving accuracy in phishing detection, reflecting the evolution of techniques and technologies. Earlier studies, such as those by Yi et al. (2018) and Sadique et al. (2020), reported lower accuracy rates of 90% and 87%, respectively, highlighting early challenges in identifying phishing threats. As research progressed, studies like those by Niakanlahiji et al. (2018) and Alsharaiah et al. (2023) showed improved accuracy, reaching 95.4% and 98.64%, respectively. More recent work by Rao & Pais (2019), HR et al. (2020), and Tang & Mahmoud (2022) consistently reported accuracy rates above 99%, demonstrating the significant strides made in phishing detection frameworks. The current study, achieving an accuracy of 99.96%, sets a new standard in the field, reflecting the ongoing refinement of detection methods and their growing effectiveness in combating phishing attacks in the latest cybersecurity research, as illustrated in Figure 5.

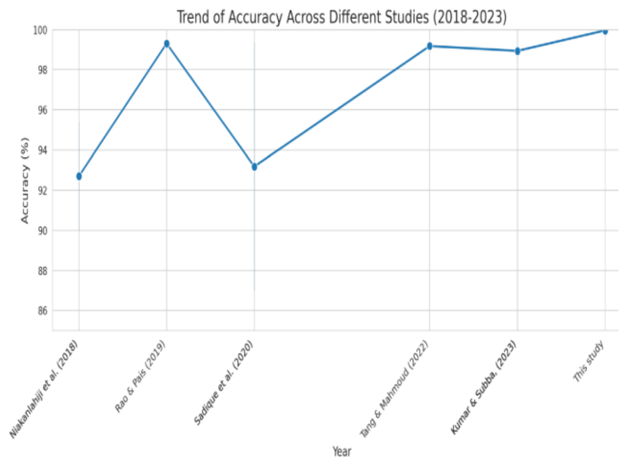


Figure 5: Comparison of the Proposed CNN-BiGRU Framework with the existing Frameworks

Table 6: Comparison of the Accuracy, Precision, Recall and F1_score of various studies with the current research

Author & Year	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
HR et al. (2020)	99.36	98.87	100	99.43
Liu et al., (2021)	Nil	98.86	89.23	93.80
Tenis & Santhosh (2023)	99.20	99.00	99.00	90.00
Alsharaiah et al. (2023)	98.64	98.60	98.70	98.60
Subba, (2023)	98.80	99.2	99.10	99.10
This Study (CNN-BiGRU)	99.96	99.92	100	99.92

Table 6 compares studies focused on phishing detection frameworks, evaluating their accuracy, precision, recall, and F1 score. The current research, employing a CNN-BiGRU model, achieves the highest accuracy at 99.96%, near-perfect precision F1-scores of 99.92%, and a recall of 100%. This highlights a significant advancement over prior studies, showcasing a highly effective model in accurately identifying phishing attacks while minimizing false positives.

Other notable studies, such as HR et al. (2020) and Subba (2023), also demonstrated strong performances, with accuracies exceeding 99% and well-balanced metrics. However, researchers like Liu et al. (2021) displayed lower precision and F1 scores, indicating potential compromises in their detection approaches. Overall, while each study contributes valuable insights to the field, the current research sets a new benchmark with its superior results, which is also depicted in Figure 5.

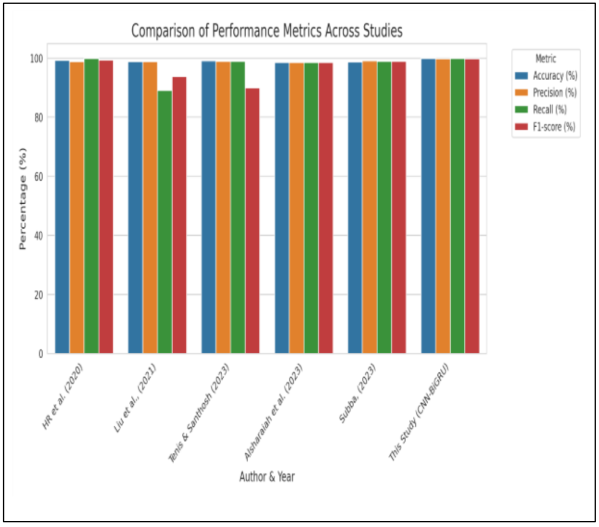
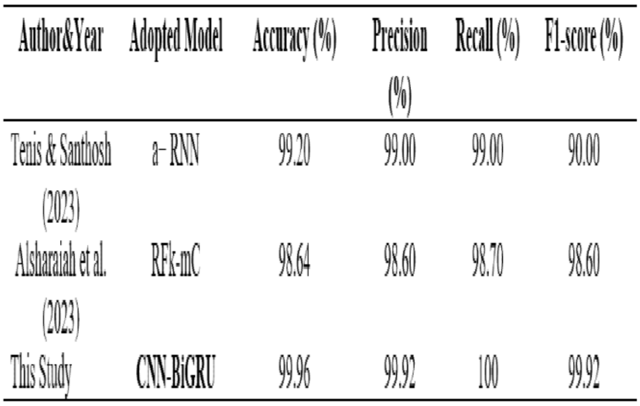


Figure 5: Comparison of the Accuracy, Precision, Recall and F1_score of various studies with the current research

Table 7: Comparison of this Study with the Baselines Studies



Author&Year	Adopted Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Tenis & Santhosh (2023)	a- RNN	99.20	99.00	99.00	90.00
Alsharaiah et al. (2023)	RFk-mC	98.64	98.60	98.70	98.60
This Study	CNN-BiGRU	99.96	99.92	100	99.92

Table 7 compares the performance of three phishing detection frameworks: a-RNN by Tenis & Santhosh (2023), RFk-mC by Alsharaiah et al. (2023), and the CNN-BiGRU framework developed in this study. The CNN-BiGRU consistently outperforms the other frameworks across all metrics. It achieves an accuracy of 99.96%, surpassing Tenis & Santhosh's 99.20% and Alsharaiah et al.'s 98.64%. In terms of precision, the CNN-BiGRU reaches 99.92%, higher than the a-RNN's 99.00% and RFk-mC's 98.60%. Notably, the CNN-BiGRU model achieves a perfect recall of 100%, meaning it detected all phishing attacks, while the other framework report recall values of 99.00% and 98.70%. The CNN-BiGRU also leads in the F1-score with 99.92%, significantly higher than Tenis & Santhosh's 90.00% and Alsharaiah et al.'s 98.60%. Overall, the CNN-BiGRU proves to be more accurate, precise, and reliable than the baseline models, marking a significant improvement in phishing detection. Figure 6 displays a performance comparison of three framework.

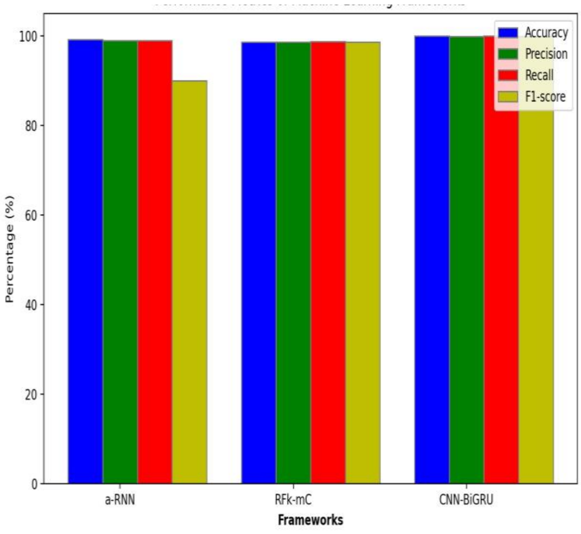


Figure 6: Performance Comparison of three Framework

The CNN-BiGRU framework achieves exceptional results, including 99.96% accuracy, 99.92% precision, and 100% recall. It proves its effectiveness in identifying phishing websites by analyzing spatial and temporal patterns where traditional models often struggle. CNN efficiently extracts spatial data from websites, while BiGRU captures temporal dependencies, enhancing the framework's ability to detect sophisticated phishing patterns. By utilizing deep learning, the framework minimizes the need for manual feature extraction, allowing for greater adaptability to evolving phishing techniques compared to traditional rule-based systems. While the proposed framework is highly effective in detecting phishing websites, improvements in generalization, computational efficiency, and continuous updates will be beneficial to sustain its performance over time.

Conclusion

This study presents a cutting-edge hybrid framework integrating CNN and BiGRU to detect phishing websites. The model effectively captures spatial and temporal features by leveraging the complementary strengths of CNN and BiGRU. Experimental results indicate that the framework achieves outstanding performance metrics, with an accuracy of 99.96%, precision of 99.92%, recall of 100%, and an F1-score of 99.92%. These metrics illustrate the model's robustness and dependability, significantly advancing over traditional and existing phishing detection techniques. Comparative analysis with previous studies further emphasizes the enhanced accuracy and effectiveness of the proposed framework, demonstrating its potential for practical implementation in cybersecurity measures to combat the evolving threat of phishing attacks. Future studies shall focus on using a broader range of datasets, especially those that capture the most recent phishing methods and newly developed phishing sites of rapid evolvement of Phishing tactics. Moreover, optimizing the framework for real-time detection by investigating lightweight models that lower computational requirements without compromising accuracy.

References

Abdulrahman, M. D., Alhassan, J. K., Adebayo, O. S., Ojeniyi, J. A., & Olalere, M. (2019). Phishing Attack Detection Based on Random Forest with Wrapper Feature Selection Method. *International Journal of Information Processing and Communication (IJIPC*, *7*(2), 209–224. http://repository.futminna.edu.ng:8080/jspui/handle/123456789/3109

Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2023). Intelligent phishing detection scheme using deep learning algorithms. *Journal of Enterprise Information Management*, *36*(3), 747– 766. https://doi.org/10.1108/JEIM-01-2020-0036

Alsharaiah, M. A., Abu-Shareha, A. A., Abualhaj, M., Baniata, L. H., Adwan, O., Al-Saaidah, A., & Oraiqat, M. (2023). A new phishing-website detection framework using ensemble classification and clustering. *International Journal of Data and Network Science*, *7*(2), 857– 864. https://doi.org/10.5267/j.ijdns.2023.1.003

Chou, C.-Y., Hsu, D.-Y., & Chou, C.-H. (2023). Predicting the Onset of Diabetes with Machine Learning Methods. *Journal of Personalized Medicine 2023, Vol. 13, Page 406*, *13*(3), 406. https://doi.org/10.3390/JPM13030406

Cuzzocrea, A., Martinelli, F., & Mercaldo, F. (2019). A machine-learning framework for supporting intelligent web-phishing detection and analysis. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3331076.3331087

Do, N. Q., Selamat, A., Krejcar, O., Yokoi, T., & Fujita, H. (2021). Phishing webpage classification via deep learning-based algorithms: An empirical study. *Applied Sciences (Switzerland)*, *11*(19). https://doi.org/10.3390/app11199210

Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, *40*. https://doi.org/10.1016/j.cosrev.2021.100379

Elnagar, S., & Thomas, M. A. (2019). *A Cognitive Framework for Detecting Phishing Websites*. *March*.

Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism. *IEEE Access*, *7*, 56329–56340. https://doi.org/10.1109/ACCESS.2019.2913705

Gupta, B. B., Arachchilage, N. A. G., & Psannis, K. E. (2018). Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems*, *67*(2), 247–267. https://doi.org/10.1007/S11235-017-0334-Z

Hr, M. G., Mv, A., Gunesh Prasad, S., & Vinay, S. (2020). Development of anti-phishing browser based on random forest and rule of extraction framework. *Cybersecurity*, *3*(1). https://doi.org/10.1186/s42400-020-00059-1

Huang, Z., Li, Y., Peng, H., & Wu, J. (2021). An ensemble learning approach for predicting the grade of brain glioma using MRI images. *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2020.102529

Idris, S. O. S. H. A. W. J. O. (2024). Stack Ensemble Model For Detection Of Phishing Website. *IEEE International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, 1–6.

Ivanov, M. A., Kliuchnikova, B. V., Chugunkov, I. V., & Plaksina, A. M. (2021). Phishing Attacks and Protection against Them. *Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021*, 425–428. https://doi.org/10.1109/ElConRus51938.2021.9396693

Khandelwal, P., Konar, J., & Brahma, B. (2020). Training RNN and it's Variants Using Sliding

Window Technique. *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2020*. https://doi.org/10.1109/SCEECS48394.2020.93

Kulkarni, A. D. (2023). Convolution Neural Networks for Phishing Detection. *International Journal of Advanced Computer*

Science and Applications, *14*(4), 15–19. https://doi.org/10.14569/IJACSA.2023.0140403

Kumar, Y., & Subba, B. (2021). A lightweight machine learning based security framework for detecting phishing attacks. *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 184–188.

Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*, *31*(3). https://doi.org/10.1142/S0129065721300011

Liu, D.-J., Geng, G.-G., Jin, X.-B., & Wang, W. (2021). An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment. *Computers and Security*, *110*. https://doi.org/10.1016/j.cose.2021.102421

Liu, D., Lee, J.-H., Wang, W., & Wang, Y. (2019). Malicious Websites Detection via CNN based Screenshot Recognition. *2019 International Conference on Intelligent Computing and Its Emerging Applications (ICEA)*, 115–119. https://doi.org/10.1109/ICEA.2019.8858300

Niakanlahiji, A., Chu, B. T., & Al-Shaer, E. (2018). PhishMon: A machine learning framework for detecting phishing webpages. *2018 IEEE International Conference on Intelligence and Security Informatics, ISI 2018*, 220–225. https://doi.org/10.1109/ISI.2018.8587410

Olatunji, S. O., Alsheikh, N., Alnajrani, L., Alanazy, A., Almusairii, M., Alshammasi, S., Alansari, A., Zaghdoud, R., Alahmadi, A., Basheer Ahmed, M. I., Ahmed, M. S., & Alhiyafi, J. (2023). Comprehensible Machine-Learning-Based Models for the Pre-Emptive Diagnosis of Multiple Sclerosis Using Clinical Data: A Retrospective Study in the Eastern Province of Saudi Arabia. *International Journal of Environmental Research and Public Health*, *20*(5), 4261. https://doi.org/10.3390/IJERPH20054261/S1

Park, A. J., Quadari, R. N., & Tsang, H. H. (2017). Phishing website detection framework through web scraping and data mining. *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2017*, 680–684. https://doi.org/10.1109/IEMCON.2017.8117212

Rao, R. S., & Pais, A. R. (2019). Detection of phishing websites using an efficient feature-based machine learning framework. *Neural Computing and Applications*, *31*(8), 3851–3873. https://doi.org/10.1007/s00521-017-3305-0

Redi, A., & Ernasari, N. (2023). *Efforts to Overcome Web-Based Phishing Crimes in the World of Cyber Crime*. https://doi.org/10.4108/eai.28-10-2023.2341807

Rendall, K., Nisioti, A., & Mylonas, A. (2020). Towards a Multi-Layered Phishing Detection. *SENSORS*, *20*(16). https://doi.org/10.3390/s20164540

Rivki, M., Bachtiar, A. M., Informatika, T., Teknik, F., & Indonesia, U. K. (2024). *Automated Phishing Detection using URLs an Webpages*. *112*. https://doi.org/https://doi.org/10.48550/arXiv.2408.01667

Sadique, F., Kaul, R., Badsha, S., & Sengupta, S. (2020). An Automated Framework for Real-time Phishing URL Detection. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 335–341. https://doi.org/10.1109/CCWC47524.2020.9031269

Sajidha, S. A. (2023). ISCX-URL-2016. *IEEE Dataport.* https://doi.org/https://dx.doi.org/10.21227/xngk-3p42

Saravanan, P., & Subramanian, S. (2020). A Framework for Detecting Phishing Websites using GA based Feature Selection and ARTMAP based Website Classification. *Procedia Computer Science*, *171*, 1083–1092. https://doi.org/10.1016/j.procs.2020.04.116

Sharifai, G. A., & Zainol, Z. (2020). Feature Selection for High-Dimensional and Imbalanced Biomedical Data Based on Robust Correlation Based Redundancy and Binary Grasshopper Optimization Algorithm. *Genes 2020, Vol. 11, Page 717*, *11*(7), 717.

Srinivasan, S., & P, D. (2023). Enhancing the security in cyber-world by detecting the botnets using ensemble classification based machine learning. *Measurement: Sensors*, *25*. https://doi.org/10.1016/j.measen.2022.100624

Subba, B. (2023). A heterogeneous stacking ensemble-based security framework for detecting phishing attacks. *2023 National Conference on Communications, NCC 2023*. https://doi.org/10.1109/NCC56989.2023.10068026

Tang, L., & Mahmoud, Q. H. (2022). A Deep Learning-Based Framework for Phishing Website Detection. *IEEE ACCESS*, *10*, 1509–1521. https://doi.org/10.1109/ACCESS.2021.3137636

Tenis, A., & Santhosh, R. (2023). Modelling of an Adaptive Network Model for Phishing Website Detection Using Learning Approaches. *Fusion: Practice and Applications*, *12*(2), 159–171. https://doi.org/10.54216/FPA.120213

Thuy, N. N., & Wongthanavasu, S. (2022). A Novel Feature Selection Method for HighDimensional Mixed Decision Tables. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(7), 3024–3037. https://doi.org/10.1109/TNNLS.2020.3048080

Wang, W., Zhang, F., Luo, X., & Zhang, S. (2019). PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks. *Security and Communication Networks*, *2019*. https://doi.org/10.1155/2019/2595794

Wei, W., Ke, Q., Nowak, J., Korytkowski, M., Scherer, R., & Woźniak, M. (2020). Accurate and fast URL phishing detector: A convolutional neural network approach. *Computer Networks*, *178*, 107275. https://doi.org/10.1016/J.COMNET.2020.107275

Yi, P., Guan, Y., Zou, F., Yao, Y., Wang, W., & Zhu, T. (2018). Web phishing detection using a deep learning framework. *Wireless Communications and Mobile Computing, 2018*. https://doi.org/10.1155/2018/4678746

Zeng, V., Zhou, X., Baki, S., & Verma, R. M. (2020). PhishBench 2.0: A Versatile and Extendable Benchmarking Framework for Phishing. *Proceedings of the ACM Conference on Computer and Communications Security*, 2077–2079. https://doi.org/10.1145/3372297.3420017

Zhao, H., Lai, Z., Leung, H., & Zhang, X. (2020). *Neural-Network Based Feature Learning: Recurrent Neural Network*. 253- 275. https://doi.org/10.1007/978-3-030-40794-0_12