Jurnal CoreIT, Vol.10, No.1, June 2024
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Phishing Detection in Deep Learning: Systematic Literature Review

Rahmad Abdillah*[1], Wenni Syafitri[2]
[1]Dept. of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim, Indonesia
[2]Dept. of Informatics Engineering, Universitas Lancang Kuning
rahmad.abdillah@uin-suska.ac.id*, wenni20@gmail.com

Abstract. Phishing is an attack that is harmful to organizations and individuals in cybersecurity. Many researchers use deep learning techniques to detect phishing. However, the proposed techniques still have shortcomings in terms of performance, especially in detecting unknown attacks, even though they have been developed in such a way. Therefore, to gain a more comprehensive understanding of the current state of research on the use of deep learning to detect phishing, a systematic literature review (SLR) is needed. This SLR aims to identify deep learning techniques, performance measures, overfitting techniques, datasets, parameters, phishing types, and recommendations for future phishing detection research. The method used by SLR consists of a research question and research objective, Search strategy, Inclusion and exclusion criteria, and Data extraction and Analysis. Over the past five years, SLR successfully identified 25 quality articles on phishing detection using deep learning. The contribution of this SLR is to provide insight into the current state of research and identify future research areas of phishing detection using deep learning techniques.

Keywords: Detection, Deep Learning, Phishing, Systematic Literature Review.

Received June 2024 / **Revised** June 2024 / **Accepted** July 2024

INTRODUCTION

Phishing is a dangerous attack that targets individuals, organizations, and even countries [1]. Phishing is a fraud in which hackers try to obtain sensitive information, such as login or account information, by impersonating an entity or person with the best reputation through services or communication channels such as email, social media, and others [2]. This act of phishing is in line with criminal law, which is a dishonest act of an individual to gain personal interests or reveal an individual's image [3]. Therefore, researchers have used various phishing prevention techniques like machine learning and deep learning. However, in recent years, researchers have begun improving phishing detection techniques using Deep Learning. This is because Machine Learning spends much time on learning, especially on manual feature engineering [4].

There are many phishing detection studies using deep learning, but very few studies have thoroughly evaluated the ability of deep learning to detect phishing attacks. In particular, research provides a comprehensive systematic review of phishing types, deep learning techniques, performance evaluation, overfitting, parameters used, and datasets.

Table 1. Recent research on phishing detection review using Deep Learning

Researcher	Range	Phishing type/Source	Method	Performance Evaluation	Experiment Parameter		
					Optimization	Overfitting	Dataset
[5]	2017-2023		√	√			√
[6]	2018-2021	√	√	√	√		√
[7]	2016-2020	√	√	√			√
This Study	2020-2024	√	√	√	√	√	√

Very few studies have used this technical aspect scheme to review, and this is because deep learning techniques, such as phishing attack detection, are still very new. Table 1 presents a comparison between our research and related research on the topic of phishing detection using Deep Learning. Related research

DOI : 10.24014/coreit.v10i1.31009

is reviewed from technical aspects, namely phishing type, method, performance evaluation, parameter optimization, overfitting, and dataset. As seen in Table 1, most researchers reviewed the methods, performance evaluation, and datasets used to detect phishing. [5] only focused on reviewing email phishing using deep learning, while [6] added a review of the parameters used for model optimization to detect phishing. The three researchers did not review the critical parameters of the model to detect phishing attacks, namely overfitting. Every phishing researcher must recognize the overfitting techniques used when building phishing detection models. Therefore, the purpose of this research is to provide an in-depth analysis of the use of deep learning techniques to detect phishing through a systematic literature review (SLR) approach with a focus on exploration, namely Phishing type/Source, Method, Performance Evaluation, Optimization parameters, Overfitting and Dataset.

The SLR conducted in this research answers every shortcoming of previous research. This SLR makes a fundamental contribution to the research area of phishing detection using deep learning. SLR makes the following contributions:

1. SLR describes the current status of research on detecting phishing attacks using deep learning.
2. SLR explores deep learning techniques researchers use, especially in the overfitting section.
3. SLR helps identify the limitations of the capabilities of each deep learning technique used to detect phishing.

METHODS

This SLR adopts the methodology used by [8] by making some modifications tailored to phishing detection research. The methodology proposed by [8] is very well-known in SLR-based research. In Figure 1, the SLR process consists of 4 steps: Research question and objective, Search strategy, Inclusion and exclusion criteria, and Data extraction and Analysis.

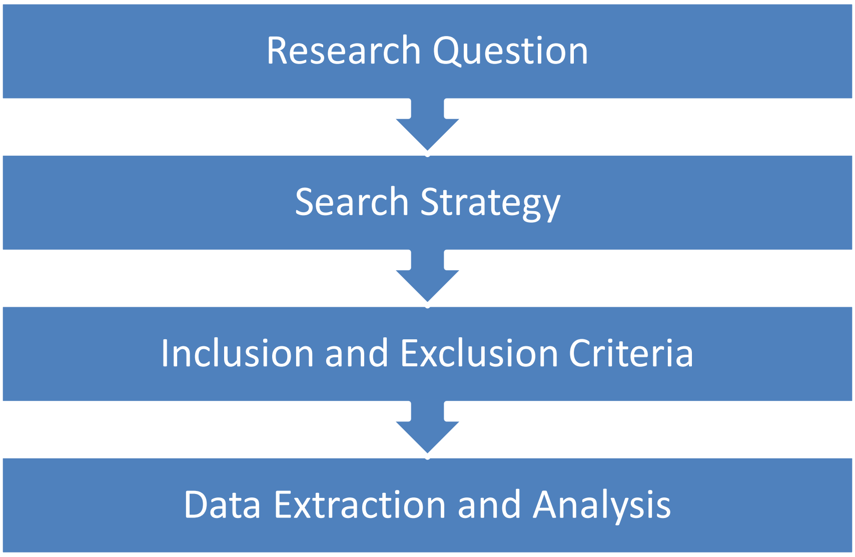
Research Question
Search Strategy
Inclusion and Exclusion Criteria
Data Extraction and Analysis

Figure 1. SLR Methodology

1. Research Question (RQ)

This SLR has six RQs to support the research objectives, namely providing an in-depth analysis of the use of deep learning techniques to detect phishing through a systematic literature review (SLR) approach with an exploration focus on Phishing type/Source, Method, Performance Evaluation, Optimization parameters, Overfitting and Dataset. The following RQs are used in SLR:
RQ1. What types of phishing are widely used as research objects?
RQ2. What Deep Learning techniques are widely used to detect phishing?
RQ3. What evaluation metrics have been used to measure the performance of deep learning?
RQ4. What parameters are used by researchers to improve the performance of deep learning models for phishing detection?

RQ5. What overfitting prevention techniques have been used to improve the quality of phishing detection models?
RQ6. What datasets have been used to train and test the proposed model?

2. Search strategy

In this process, the search strategy includes using databases such as IEEE, ACM, Science Direct, SpringerLink, and others. This SLR uses the Web of Science (WOS) search engine feature to get articles from these databases. The keywords used in this SLR are "Phishing," "Detection," and "Deep Learning."

3. Inclusion and exclusion criteria

In this process, the criteria that will be used in SLR include document type and, more specifically, the title of articles containing keywords such as "Phishing," "Detection," and "Deep Learning." In addition, document type restrictions are used to obtain articles of higher quality that are relevant to the needs of SLR. Inclusion criteria include unavailable articles, articles irrelevant to RQ, and article types in conferences, book chapters, books, theses, and other articles.

4. Data Extraction and Analysis

This SLR limits articles published in 2020-2024 to get the latest articles. After that, SLR will be extracted into several discussion topics related to RQ, namely Phishing type/Source, Method, Performance Evaluation, Optimization parameters, Overfitting, and Dataset. The extract results will be analyzed quantitatively to get conclusions for each topic of discussion.

RESULT AND DISCUSSION

This research obtained article data from as many as three review articles and 25 research articles. Figure 2 shows the process of obtaining the latest articles and phishing detection using deep learning techniques. Query-related search activities involve filters for article topics, titles, and articles published in 2020-2024. After that, the articles obtained will be included and excluded with criteria such as unavailable articles, articles not relevant to the RQ, and article types in the form of conferences, book chapters, books, thesis, and other articles.

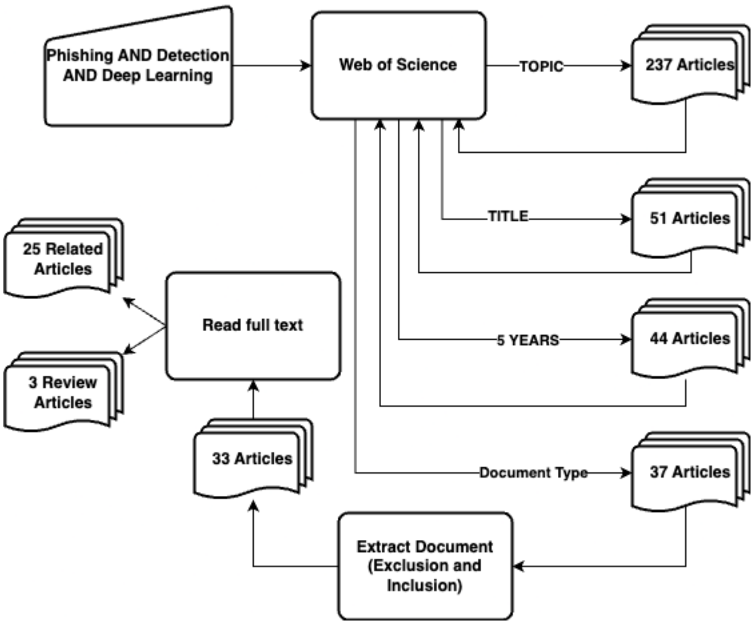

Figure 2. Search Finding

In addition, Figure 3 shows the more dominant discussions of the articles obtained during the SLR. Convolutional neural networks (CNN) are frequently discussed in the articles, and CNN is also the basic model for phishing attack detection.

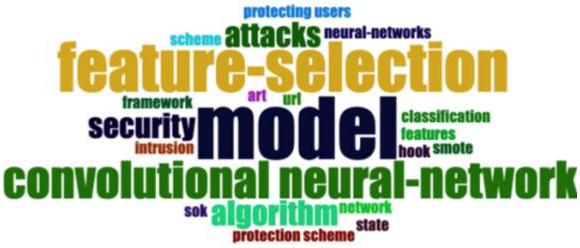

Figure 3. Wordcloud of phishing detection research using deep learning

Literature Survey Findings

Research Paper Publish in 2020

Wang, et al. [9] used the Bidirectional LSTM (BiLSTM) technique to detect phishing websites and the UCI Phishing website dataset as training data. The training data consists of 4898 (phishing) and 6157 (legitimate). In addition, overfitting techniques are used, such as model complexity and parameter hypertuning, and model performance evaluation using accuracy, precision, recall, False Positive Rate, F1-Score, False Negative Rate, True Negative, and True Positive. Somesha, et al. [10] used Deep Neural Network (DNN), Long short-term memory (LSTM), and CNN techniques to detect phishing websites, in addition, PhishTank and Alexa datasets were used as training data with data distribution of 2119 phishing and 1407 legitimate. The model is evaluated using accuracy and error rate measures, while overfitting techniques are used, namely hyperparameters and model complexity. Somesha, et al. [10] suggested using heuristic feature-extracting techniques to get more information in detecting phishing. Rasymas & Dovydaitis [11] used CNN and LSTM techniques to detect phishing websites. The training data used is sourced from PhishTank and has 2585146 data. The model is evaluated with performance measures of Precision, recall, accuracy, f1-score, roc_auc, and techniques used to prevent overfitting, namely early stopping technique and regularization. Feng, et al. [12] used CNN and BiLSTM techniques on PhishTank and Alexa datasets to detect phishing websites. The dataset consists of 24800 normal and 21303 phishing. The model is evaluated using performance measures of Precision, recall, accuracy, f1-score, and False Positive Rate, and the overfitting techniques used are Regularization and hyperparameter tuning. Al-Alyan & Al-Ahmadi [13] used the CNN technique to detect phishing websites with Precision, recall, accuracy, and f1-score performance evaluation measures. They used MUPD and Sahingoz datasets, with a data distribution of 1188695 legitimate and 1218732 phishing. The overfitting techniques used are data handling, hyperparameter tuning, regularization, and model complexity. They suggest trying new data to test the model that has been built. Adebowale, et al. [14] used CNN and LSTM techniques to detect phishing websites. Their training data was sourced from PhishTank and Whois, totaling 10000 data. The built model was evaluated through Precision, recall, accuracy, and f1-score measures, as well as the use of overfitting techniques, namely k-fold cross-validation and holdout cross-validation, hyperparameter tuning, and model complexity. They suggested building a web browser plugin for real-time phishing protection.

Research Paper Publish in 2021

Yang et al. [15] built a phishing detection website using the CNN technique and training data from Alexa, PhisTank, and Yandex. Their training data consisted of 67908 Legitimate and 63159 Phishing. The model was evaluated using Precision, recall, accuracy, f1-score, False Positive Rate, and True Positive Rate measures. Hyperparameter tuning and early stopping were used to overcome overfitting. They suggest building a technique that can extract hidden features.

In contrast to Alzahrani [16], they used the CNN technique to detect phishing websites using training data sourced from Alexa and Phishtank. This dataset consists of 10000 legitimate and 10000 phishing records. The model is evaluated using Accuracy, precision, recall, False Positive, False Negative, and True Negative measures, as well as overfitting prevention using model complexity techniques. They suggested that the model should be tested on real cases.

Research Paper Publish in 2022

Tang & Mahmoud [17] built an Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) model to detect phishing websites. They used PhisStorm, PhishTank, ISCX-URL2016, and Kaggle datasets with data distribution of 429125 legitimate and 236362 phishing. Model evaluation using accuracy, f1-

score, false positive, and false negative. Early stopping, regularization, and hyperparameter tuning were used to prevent overfitting. However, Tang & Mahmoud [17] suggested hardware configuration to maximize detection. Shaiba, et al. [18] used CNN and LSTM to detect phishing websites. They used Alexa and PhishTank datasets as training data with a data distribution of 10817 (benign) and 11037 (phishing). The model evaluation uses Hyperparameter tuning, model complexity overfitting techniques, and Accuracy, Precision, Recall, F1-Score, Specificity, and AUC Score as their proposed model measures. Shaiba, et al. [18] suggested combining multiple algorithms to improve detection. Ogawa, et al. [19] used CNN and BiLSTM techniques to build a phishing website detection model. They used a private dataset with a distribution of 32675 training data and 3619 test data. Model overfitting is addressed using hyperparameter tuning and model complexity, while performance measures use accuracy and false positive rate. Ogawa, et al. [19] suggested adding the results of analyzing information from the Domain Name System (DNS) to improve detection accuracy. Korkmaz, et al. [20]used Generative Adversarial Network (GAN), CNN, and DNN to detect phishing websites. They used training data sourced from PhishTank with a data distribution of 51316 legitimate and 36173 phishing. Cross-validation, Model Complexity, Hyperparameter tuning, and regularization were used to overcome overfitting. In addition, their model was evaluated using the performance measures Sensitivity, Specificity, Precision, F1 Score, Negative Predictive Value, False Positive Rate (FPR), False Discovery Rate, False Negative Rate, Accuracy, and Error Rate. They suggested adding more data and collaboration with other algorithms to improve detection. Elsadig, et al. [21] used BERT and DNN to build models and evaluated their performance using Precision, recall, accuracy, f1-score, True Negative, True Positive, False Positive, and False Negative. Model complexity and regularization as overfitting techniques. They used the Kaggle dataset as training data with a 156422 (bad) and 392294 (good) distribution. In addition, Elsadig, et al. [21] suggested improving phishing detection using dynamic features and CNN techniques. Bu & Kim [22] used Convolutional Recurrent Neural Network (CRNN) and Genetic Algorithm to detect phishing websites. They used training data sourced from Phishstorm, Phishtank, ISCX-URL-2016, and Open directory project datasets with URL data distribution, namely 222541 for phishing and normal. Model evaluation against overfitting uses Cross-validation, hyperparameter tuning, model complexity, and regularization techniques, while model performance evaluation uses accuracy and recall measures. Bu & Kim [22] suggested evaluating deep lerarning performance based on parameter reduction. Almousa, et al. [23] used CNN, LTSM, and RNN to detect phishing websites. The training data used by Almousa, et al. [23] is sourced from Tan, Kumar, UCI, and AZA datasets, with 72605 dataset records. This technique is evaluated using performance measures of precision, recall, f1-score, accuracy, FPR and early stopping, model complexity, and hyperparameter tuning to prevent overfitting.

Research Paper Publish in 2023

Prabakaran, et al. [4] used the collaboration of Variational Autoencoder (VAE) and DNN to detect phishing websites. The proposed model uses Kaggle's phishing website dataset as training and test data. This dataset consists of 50000 malicious URLs and 50000 benign URLs. The model is evaluated using the performance measures accuracy, precision, recall, f1-score, true positive rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR). In addition, Hyperparameters and regularization are used to overcome overfitting techniques. However, the limitation of the model built by Prabakaran, et al. [4] is that there are too many false alarms. Benavides-Astudillo, et al. [24] used LTSM, BiLTSM, GRU, and Bidirectional GRU (BiGRU) techniques to detect phishing websites. Benavides-Astudillo, et al. [24] used the Phishload dataset of 9198 phishing and 1176 legitimate. The model was evaluated with precision, recall, and f1-score measures, as well as cross-validation, regularization, and model complexity as techniques to prevent overfitting. Benavides-Astudillo, et al. [24] suggested using other word embedding techniques, such as Fast Text and Word2Vec, to improve detection. Alshingiti, et al. [25]used LSTM and CNN to detect phishing websites using the Canadian Institute for Cybersecurity dataset. This dataset consists of 9800 phishing and 10200 benign. The model used by Alshingiti, et al. [25]was evaluated using accuracy, precision, recall, and f1-score. The overfitting techniques used are Model Complexity, regularization, and hyperparameter tuning. In addition, Alshingiti, et al. [25]suggested improving the model training process and improving feature recognition in the model. While Aldakheel, et al. [26] used the RNN technique to detect phishing websites in the PhisTank database with 10000 phishing and 10000 legitimate records. The model evaluation uses accuracy, precision, recall, and f1-score while overfitting prevention techniques use regularization and hyperparameter tuning. Aldakheel, et al. [26] suggested reducing the bias of the resulting model. Unlike the others, Abdullah Alohali, et al. [27] used a private dataset to train the LSTM technique for phishing detection. Their private dataset consisted of 7053 benign and 5995 phishing. The LSTM was evaluated using accuracy, precision, recall, f1-score, and Jaccard index measures. Abdullah Alohali et al. [27] used hyperparameter tuning, regularization, and model complexity techniques to avoid overfitting. They suggested trying to incorporate some other deep-learning techniques.

Brindha, et al. [3] used RNN and GRU techniques to detect phishing emails. The training data they used was sourced from the Enron email dataset, which consisted of 7781 legitimate emails and 999 phishing emails. They also used hyperparameter tuning and data handling techniques to overcome overfitting. Brindha, et al.'s research [3] suggests combining deep learning models to improve detection. Meanwhile, Atawneh & Aljehani [28] used CNN, LTSM, BERT, and RNN techniques to detect phishing emails. The dataset used by Atawneh & Aljehani [28] is sourced from the enron email dataset with a total of 999 phishing records and 7781 legitimate. Data handling and regularization are techniques to overcome overfitting. Their techniques were evaluated using precision, recall, and f1-score measures. Atawneh & Aljehani [28] suggested analyzing the habits of phishing emails to improve detection.

Research Paper Publish in 2024

Sahingoz, et al. [2] used RNN, Bidirectional Recurrent Neural Network (BRNN), CNN, Artificial Neural Network (ANN), and ATT techniques to detect website phishing attacks. Sahingoz, et al. [2] used PhishTank and commoncrawl datasets as their training and test data. Their dataset consists of 2.3 million phishing records and 2.8 million legitimate records. They used regularization, model complexity, and hyperparameter tuning techniques to anticipate the model built by Sahingoz, et al. [2] overfitting. Their model was tested regarding precision, recall, f1-score, and accuracy. However, Sahingoz, et al. [2] experienced hardware limitations when performing complex configurations of the model. In contrast to Alsubaei, et al. [29], they used ResNeXt-GRU to detect phishing websites. They used a phishing website dataset of 5000 phishing web pages and 5000 legitimate web pages. Their model was evaluated using evaluation measures only. In addition, they used regularization, model complexity, data handling, and hyperparameter tuning techniques. Alsubaei, et al. [29] suggested a combination of several algorithms to improve the accuracy and efficiency of phishing detection techniques.

Altwaijry, et al. [1] used CNN, LTSM, and GRU techniques to detect email phishing. The datasets used are phishing corpus and spam assassin. The dataset records consist of 2278 phishing emails and 4150 legitimate emails. CNN, LTSM, and GRU were evaluated using accuracy, precision, recall, and F1-score. In addition, the overfitting techniques they use are regularization, model complexity, early stopping, and hyperparameter tuning techniques. The obstacle encountered by Altwaijry, et al. [1] is not testing on actual data, besides that, they suggest focusing on analyzing email headers.

Objective Findings

RQ1. Phishing type

In the last five years, phishing detection using deep learning focused on website phishing types (22 Articles) compared to email phishing types (3), as shown in Figure 4. Researchers began using phishing detection techniques in emails in 2022 until 2024. Email phishing can have serious consequences, such as financial loss, identity theft, and even damage to an organization's reputation. Email phishing continues to evolve, following the trend of the times to avoid law enforcement agencies or security systems [28].

In contrast, website phishing detection started in 2020 with a downward trend until 2024. Website phishing is a process to attract people to visit a fake website (website content and URL are made similar to legitimate websites) and persuade them to enter identification information such as usernames, passwords, addresses, social security numbers, personal identification numbers, and other information that can be made to look plausible [4].

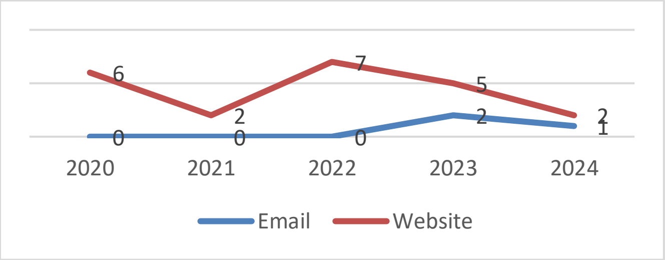

Figure 4. Phishing type

RQ2. Deep learning technique to detect phishing.

Over the past five years, researchers have used deep learning techniques to detect phishing, such as RNN, BRNN, CNN, ANN, LTSM, GRU, DNN, BiGRU, BiLSTM, GAN, and CRNN. Figure 5, 2022 saw the most use of deep-learning techniques such as RNN, CNN, LTSM, GRU, DNN, BiLSTM, GAN, and CRNN. The least used deep learning techniques are GAN (2022), CRNN (2022), BiGRU (2023), ANN (2024), and BRNN (2024). Every year, researchers use CNN techniques, followed by LTSM techniques. The majority of CNN technique usage occurred in 2020 and 2022.

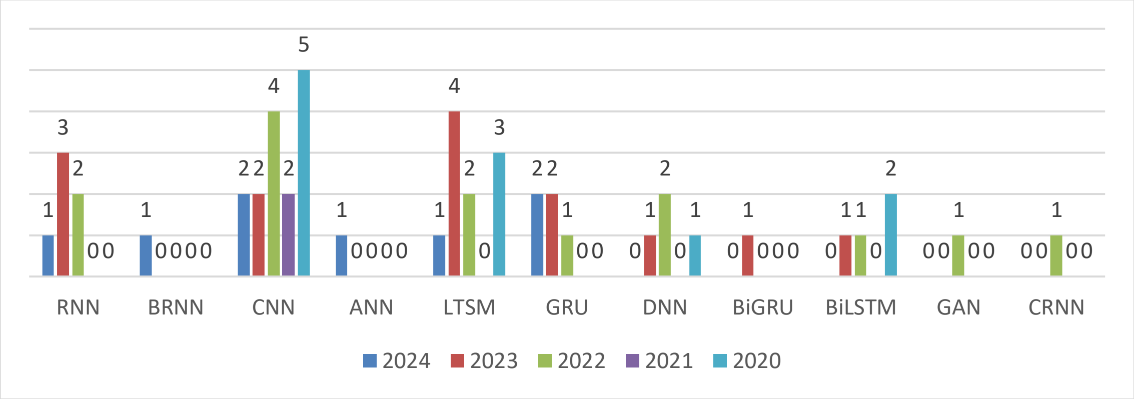

Figure 5. Various Deep Learning Techniques

RQ3. Evaluation metrics to measure deep learning performance

In the last five years, most researchers used 8 out of 14 measures to assess the performance of deep learning techniques, namely Precision, Recall, True Positive, False Positive, F1-Score, True Negative, False Negative, and Accuracy. Precision, Recall, F1-Score, and Accuracy are most widely used to evaluate performance in the last five years. 2022 is the most diverse year using deep learning performance evaluation techniques.

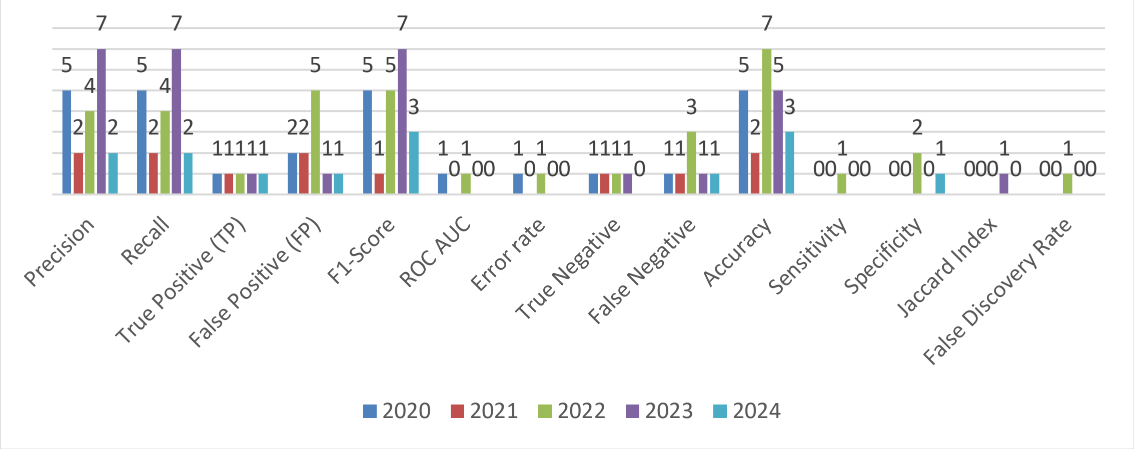

Figure 6. Various performance evaluation

RQ4. Parameters to improve the performance of deep learning models

In the last five years, researchers have used various parameters to improve the performance of deep learning techniques. Parameters such as epoch, learning rate, dropout, batch size, optimizer, loss, and activation function are most commonly used by researchers to improve deep learning performance. 2022, 2023, and 2024 are the most diverse years for using parameters to improve the performance of deep-learning techniques.

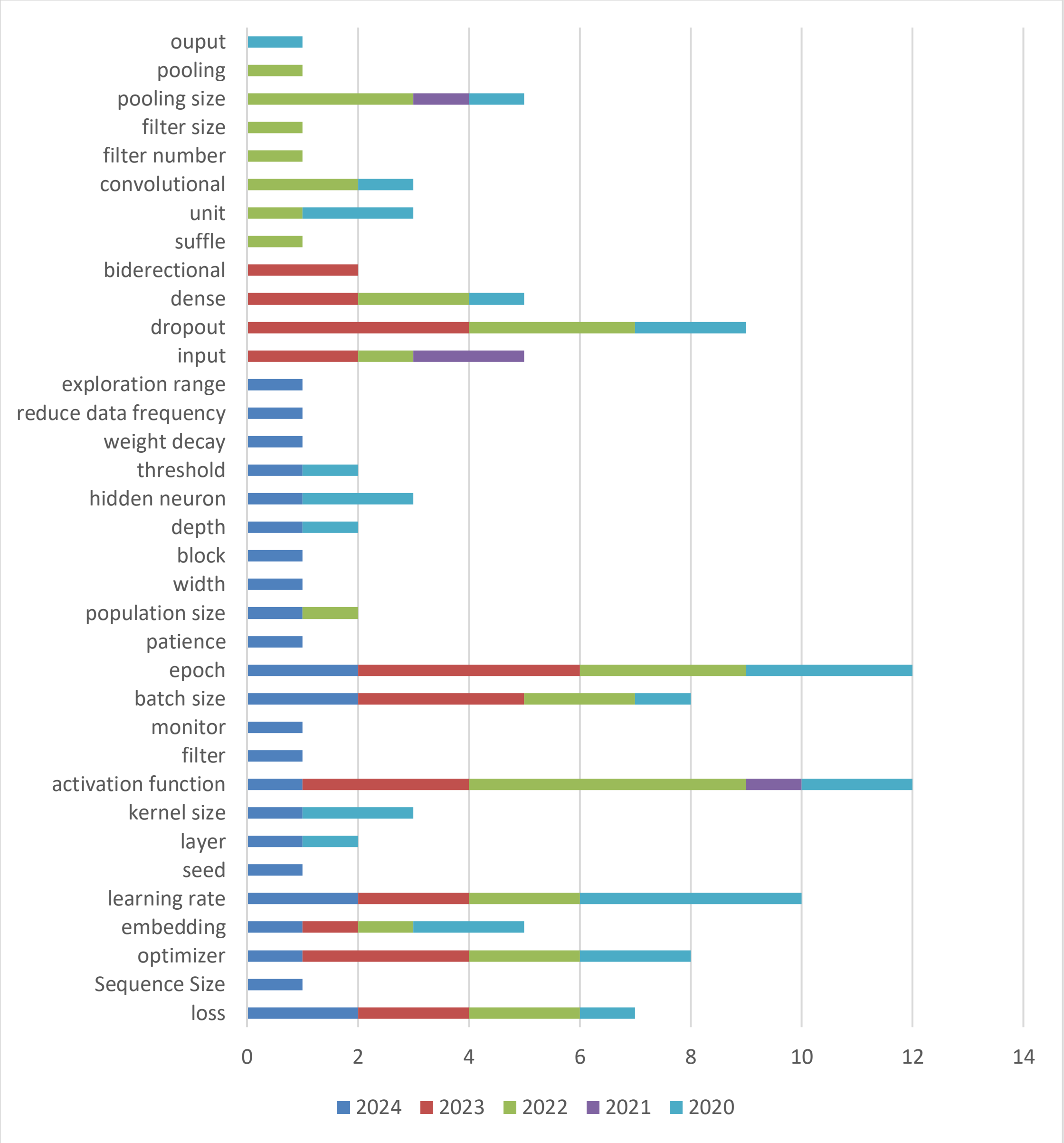

Figure 7. Various parameters to enhance deep learning detection

RQ5. Overfitting technique to improve the quality of phishing detection models

Deep learning researchers always use overfitting techniques to detect and avoid overfitting their proposed models. Researchers have used model complexity and hyperparameter tuning in the last five years to overcome overfitting. After model complexity and hyperparameter tuning, regularization, early stopping, and cross-validation are the most used techniques. 2020 is a year that uses many variants of overfitting techniques.

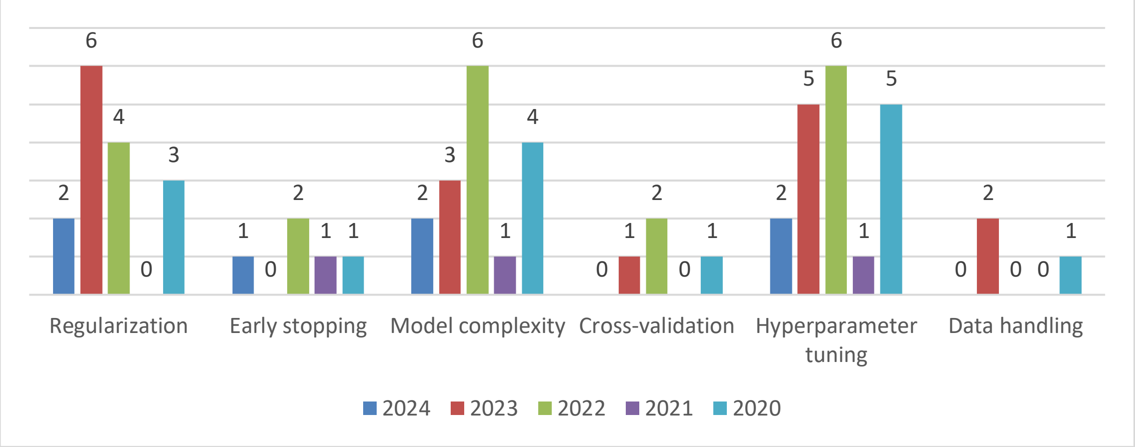

Figure 8. Overfitting techniques in phishing detection

RQ6. Datasets for training and testing the phishing detection model

In Figure 9, the PhishTank dataset is the most used over the last five years. Alexa is the most used dataset after PhishTank, but it is only used in the interval 2020-2022.

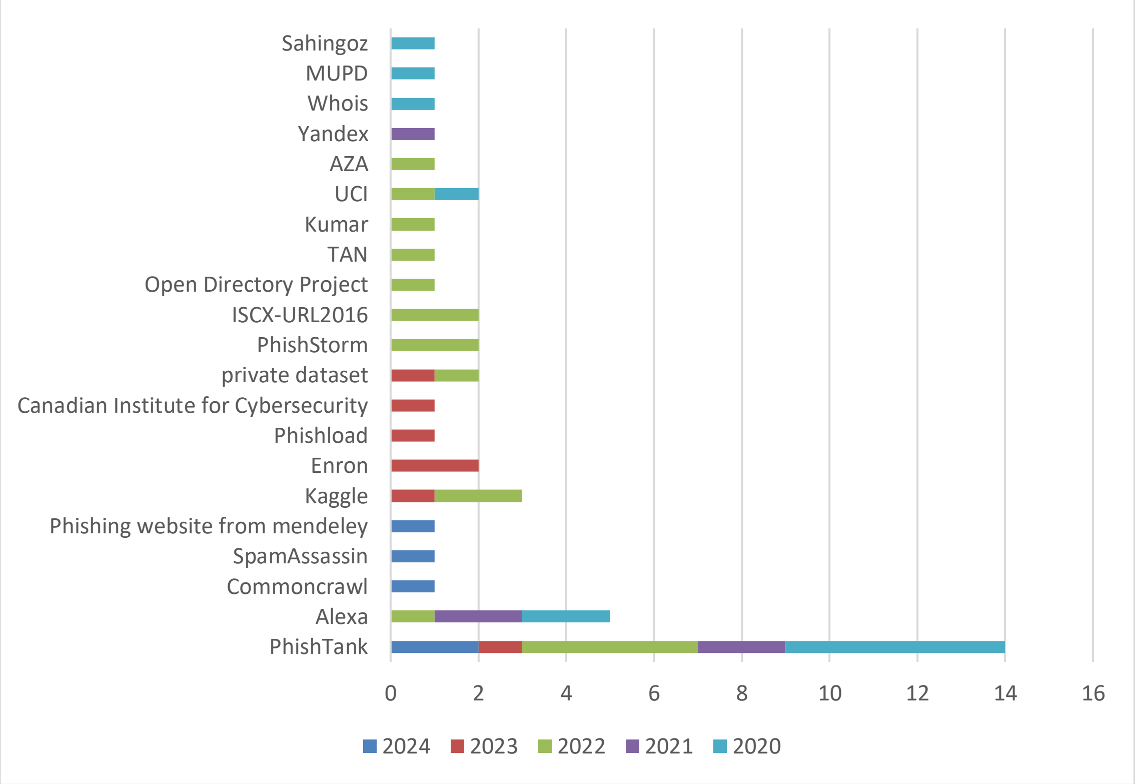

Figure 9. Various datasets in phishing detection using deep learning techniques

Based on Figure 10, the researchers' most widely used dataset size is 10001-50000. While the sizes 1001-10000, 50001-100000, and 100001-1000000 are the most widely used datasets after the size 10001-50000.

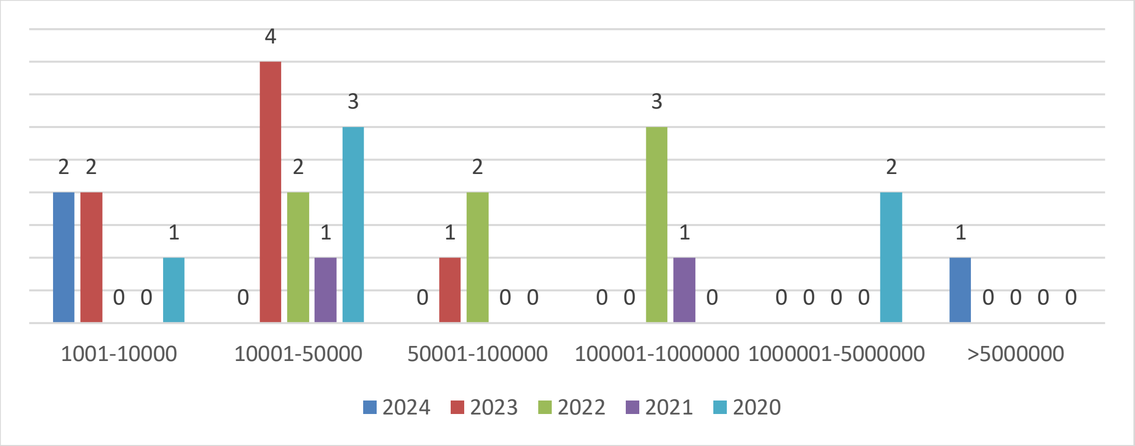

Figure 10. Various sizes of datasets are used as training and test data to detect phishing.

Future Direction

Based on the limitations of researchers over the past five years on the review findings, some suggestions that can be used by researchers who want to develop phishing detection techniques using deep learning, such as models that have been built must be tested on actual data [1][13], Combination of several algorithms or models to improve detection [29][3][27][18][20][21], Optimize the use of hardware to train the model [2][17], reduce the time for model training[25], observe phishing email behavior[28], explore feature extraction techniques[24][25][21][15][10], reduce false alarms[4][26], analyze email headers[1], analyze DNS[19], increase the amount of data[20], analyze model performance against parameter reduction[22], develop realtime phishing detection tools[23][16][14], and compare models with various methods on the proposed dataset[11].

CONCLUSION

This SLR has successfully identified some limitations and future directions for phishing detection using deep learning. Although many researchers in the last five years have shown that the models built to provide the best results in detecting phishing, there is still room for improvement. Some main limitations include implementation in real phishing cases, optimization of hardware usage, exploration of feature extraction techniques, reducing false alarms, analysis of the impact of parameter reduction, email header analysis, and DNS analysis. In summary, although this SLR revealed many limitations and future research directions for phishing detection using deep learning, the limitations encountered can be overcome with further research. The proposed improvements can help improve phishing detection systems' performance, ultimately providing better cybersecurity impact for individuals and organizations.

REFERENCES

[1] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, “Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models,” *Sensors*, vol. 24, no. 7, p. 2077, Mar. 2024, doi: 10.3390/s24072077.

[2] O. K. Sahingoz, E. BUBEr, and E. Kugu, “DEPHIDES: Deep Learning Based Phishing Detection System,” *IEEE Access*, vol. 12, pp. 8052–8070, 2024, doi: 10.1109/ACCESS.2024.3352629.

[3] R. Brindha, S. Nandagopal, H. Azath, V. Sathana, G. Prasad Joshi, and S. Won Kim, “Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification,” *Comput. Mater. Contin.*, vol. 74, no. 3, pp. 5901–5914, 2023, doi: 10.32604/cmc.2023.030784.

[4] M. K. Prabakaran, P. Meenakshi Sundaram, and A. D. Chandrasekar, “An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders,” *IET Inf. Secur.*, vol. 17, no. 3, pp. 423–440, May 2023, doi: 10.1049/ise2.12106.

[5] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, “A Systematic Review on Deep-Learning-Based Phishing Email Detection,” *Electronics*, vol. 12, no. 21, p. 4545, Nov. 2023, doi: 10.3390/electronics12214545.

[6] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, “Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions,” *IEEE Access*, vol. 10, pp. 36429–36463, 2022, doi: 10.1109/ACCESS.2022.3151903.

[7] C. Catal, G. Giray, B. Tekinerdogan, S. Kumar, and S. Shukla, “Applications of deep learning for phishing detection: a systematic literature review,” *Knowl. Inf. Syst.*, vol. 64, no. 6, pp. 1457–1500, Jun. 2022, doi: 10.1007/s10115-022-01672-x.

[8] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering – A systematic literature review,” *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

[9] S. Wang, S. Khan, C. Xu, S. Nazir, and A. Hafeez, “Deep Learning-Based Efficient Model Development for Phishing Detection Using Random Forest and BLSTM Classifiers,” *Complexity*, vol. 2020, pp. 1–7, Sep. 2020, doi: 10.1155/2020/8694796.

[10] 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, no. 1, p. 165, Dec. 2020, doi: 10.1007/s12046-020-01392-4.

[11] T. Rasymas and L. Dovydaitis, “Detection of phishing URLs by using deep learning approach and multiple features combinations,” *Balt. J. Mod. Comput.*, vol. 8, no. 3, pp. 471–483, 2020, doi: 10.22364/BJMC.2020.8.3.06.

[12] J. Feng, L. yang Zou, O. Ye, and J. zhou Han, “Web2Vec: Phishing Webpage Detection Method Based on Multidimensional Features Driven by Deep Learning,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3043188.

[13] A. Al-Alyan and S. Al-Ahmadi, “Robust URL phishing detection based on deep learning,” *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 7, pp. 2752–2768, 2020, doi: 10.3837/tiis.2020.07.001.

[14] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, “Intelligent phishing detection scheme using deep learning algorithms,” *J. Enterp. Inf. Manag.*, vol. ahead-of-p, no. ahead-of-print, Jun. 2020, doi: 10.1108/JEIM-01-2020-0036.

[15] 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, p. 8281, Dec. 2021, doi: 10.3390/s21248281.

[16] S. M. Alzahrani, “Phishing Attack Detection Using Deep Learning,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 21, no. 12, pp. 213–218, Dec. 2021, doi: 10.22937/IJCSNS.2021.21.12.31.

[17] L. Tang and Q. H. Mahmoud, “A Deep Learning-Based Framework for Phishing Website Detection,” *IEEE Access*, vol. 10, pp. 1509–1521, 2022, doi: 10.1109/ACCESS.2021.3137636.

[18] H. Shaiba, J. S. Alzahrani, M. M. Eltahir, R. Marzouk, H. Mohsen, and M. Ahmed Hamza, “Hunger Search Optimization with Hybrid Deep Learning Enabled Phishing Detection and Classification Model,” *Comput. Mater. Contin.*, vol. 73, no. 3, pp. 6425–6441, 2022, doi: 10.32604/cmc.2022.031625.

[19] Y. Ogawa, T. Kimura, and J. Cheng, “Deep-learning-based sequential phishing detection,” *IEICE Commun. Express*, vol. 11, no. 4, pp. 171–175, Apr. 2022, doi: 10.1587/comex.2021XBL0212.

[20] M. Korkmaz, E. Kocyigit, O. K. Sahingoz, and B. Diri, “A Hybrid Phishing Detection System Using Deep Learning-based URL and Content Analysis,” *Elektron. Ir Elektrotechnika*, vol. 28, no. 5, pp. 80–89, Oct. 2022, doi: 10.5755/j02.eie.31197.

[21] M. Elsadig *et al.*, “Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction,” *Electronics*, vol. 11, no. 22, p. 3647, Nov. 2022, doi: 10.3390/electronics11223647.

[22] S.-J. Bu and H.-J. Kim, “Optimized URL Feature Selection Based on Genetic-Algorithm-Embedded Deep Learning for Phishing Website Detection,” *Electronics*, vol. 11, no. 7, p. 1090, Mar. 2022.

[23] M. Almousa, T. Zhang, A. Sarrafzadeh, and M. Anwar, “Phishing website detection: How effective are deep learning-based models and hyperparameter optimization?,” *Secur. Priv.*, vol. 5, no. 6, p. e256, Nov. 2022, doi: 10.1002/spy2.256.

[24] E. Benavides-Astudillo, W. Fuertes, S. Sanchez-Gordon, D. Nuñez-Agurto, and G. Rodríguez-Galán, “A Phishing-Attack-Detection Model Using Natural Language Processing and Deep Learning,” *Appl. Sci.*, vol. 13, no. 9, p. 5275, Apr. 2023, doi: 10.3390/app13095275.

[25] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN,” *Electronics*, vol. 12, no. 1, p. 232, Jan. 2023, doi: 10.3390/electronics12010232.

[26] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. A. Alzahrani, “A Deep Learning-Based Innovative Technique for Phishing Detection in Modern Security with Uniform Resource Locators,” *Sensors*, vol. 23, no. 9, p. 4403, Apr. 2023, doi: 10.3390/s23094403.

[27] M. Abdullah Alohali *et al.*, “Metaheuristics with deep learning driven phishing detection for sustainable and secure environment,” *Sustain. Energy Technol. Assess.*, vol. 56, p. 103114, Mar. 2023, doi: 10.1016/j.seta.2023.103114.

[28] S. Atawneh and H. Aljehani, “Phishing Email Detection Model Using Deep Learning,” *Electronics*, vol. 12, no. 20, p. 4261, Oct. 2023, doi: 10.3390/electronics12204261.

[29] F. S. Alsubaei, A. A. Almazroi, and N. Ayub, “Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics,” *IEEE Access*, vol. 12, pp. 8373–8389, 2024, doi: 10.1109/ACCESS.2024.3351946.