Jurnal CoreIT, Vol.11, No.1, June 2025
ISSN 2460-738X (Print)
ISSN 2599-3321 (Online)

Optimization Of Social Media Phishing Detection Models

Wenni Syafitri[1*], Guntoro[2], Ahmad Zamsuri[3], Idel Waldelmi[4]
[1,2,3]Dept. of Informatics Engineering, Universitas Lancang Kuning, Indonesia
[4]Dept. of Management, Universitas Lancang Kuning, Indonesia
wenni20@gmail.com (*Corresponding Author), guntoro@unilak.ac.id, ahmadzamsuri@unilak.ac.id, idelwaldelmi@unilak.ac.id

Abstract. Phishing is one of the most dangerous attacks in the cyber world. Very few researchers have focused on social media phishing, although SMS phishing can be related to the messaging features available on various social media platforms. This study will utilize PSO and PCA techniques to optimize the performance of RF in social media phishing. This study will compare the performance of PSO and RF with that of PCA and RF. An optimized phishing message detection model was built using NLP, incorporating TF-IDF for feature extraction, PCA and PSO for feature optimization, and Random Forest as a classifier to distinguish phishing messages from normal messages. The RF model optimized by PSO produces nearly balanced metrics: precision (0.9877), recall (0.9728), and F1 (0.9802), all of which are high. The RF model with PCA optimization achieves a slightly lower Accuracy (0.9639) and the lowest Precision (0.9585). Although there were no significant differences in the classification process, PSO and PCA made a real contribution to future research development.

Keywords: PCA, Phishing, PSO, Random Forest, SMS, Social Media

Received June 2025 / **Revised** June 2025 / **Accepted** June 2025

INTRODUCTION

Phishing is one of the most dangerous attacks in the cyber world. The current trend in phishing involves cyberattacks that target individuals for both material and non-material gain. The impact of phishing attacks on organizations can be paralyzing if the attack targets the organization's operations. Currently, phishing has expanded into the personal sphere through social media platforms, including WhatsApp, Telegram, Facebook Messenger, and Twitter, among others [1]. Several studies have developed various models to address phishing attacks, but not on social media, such as CNN and LSTM models focused on email phishing, Reinforcement Learning models with feature selection [2], and Random Forest, CatBoost, AdaBoost, and Multilayer Perceptron models focused on phishing URLs, and CNN, BiGRU, and GRU models focused on SMS phishing [3].

Very few researchers have focused on social media phishing, although SMS phishing can be related to the messaging features available on various social media platforms, as demonstrated by [3]. The models produced by [3] achieved an accuracy of 95.3% for GRU, 94.42% for BiGRU, and 93.59% for CNN. However, the research [3] has limitations, such as relying on tools to translate English text into Arabic, and its models have never been tested on other datasets; therefore, the accuracy justification only applies to their dataset. Some researchers employ optimization techniques to enhance the accuracy of detecting phishing attacks. Optimization techniques used include Principal Component Analysis (PCA) on phishing URLs [4], Particle Swarm Optimization (PSO) on phishing URLs [5], Hunger Search Optimization (HSO) on phishing URLs [6], multi-objective evolution optimization algorithm (MOE) on phishing websites [7], Chicken Swarm Optimization on phishing websites [8], Manta Ray Foraging Optimization on phishing URLs [9], Brown-Bear Optimization on phishing URLs [10], Horse Herd Optimization [11] on phishing URLs, and dynamic jaya optimization on phishing URLs [12]. These optimization studies are limited to phishing URLs and websites. Therefore, this study will optimize one of the detection techniques, namely Random Forest (RF). RF is used because it has a high detection rate on various types of datasets sourced from phishing websites and emails [13]. Additionally, optimization using PSO successfully improved the performance of Random Forest (RF) by 2% in detecting phishing URLs [5]. [4] Successfully optimized RF for phishing URLs. However, PCA and PSO optimization were not performed for detecting phishing on social media.

Research on SMS phishing remains limited, despite messaging features on social media serving as a potential channel for this type of phishing. Research [14] discusses phishing detection in COVID-19 emails

DOI: 10.24014/coreit.v11i1.37602

by analyzing URLs, subjects, and message links. Uclassify is used for the automatic classification of content, spam, and language using a semi-supervised learning approach in the health field, employing the Multinomial Naive Bayesian (MNB) algorithm, which accepts binary input. Two experiments demonstrated that MNB achieved high accuracy, specifically 96% (Kaggle Phishing data) and 96.67% for URL identification and 91.6% for hyperlink identification (PhishTank data), with a focus on tokenization, lemmatization, and feature extraction [14]. In [15], deep learning techniques, including CNN, LSTM, RNN, and BERT, were utilized in conjunction with NLP-based feature extraction. BERT and LSTM showed the highest accuracy of 99.61%. Research [16] states that transformer models, such as DistilBERT, BERT, and RoBERTa, outperform traditional models, including Logistic Regression, Random Forest, SVM, and Naive Bayes. roBERTa achieved an accuracy of 0.9943, while SVM had the highest accuracy among traditional methods at 0.9876. Research [17] shows that Federated Learning (FL) performs comparably to centralized learning, with THEMIS achieving an accuracy of 97.9% at epoch 45 and BERT 96.1% at epoch 15. FL has also proven resilient in imbalanced data distributions. To address data imbalance, [18] proposed two algorithms with undersampling: FMPED and FMMPED. These algorithms remove and filter benign emails before training, using an ensemble method, which yields an F1-score and accuracy of 0.9945, an AUC of 0.9828, and a G-mean of 0.9827. Research [19] shows that the combination of FastText and CNN yields the highest accuracy, precision, and F1-score (98.4375%), while FastText with LSTM has the highest recall score (98.9583%). In contrast to [20], they developed the ICSOA-DLPEC model based on the Intelligent Cuckoo Search (CS) optimization algorithm and GRU. After preprocessing and feature extraction with CS, the GRU parameters were refined. The ICSOA-DLPEC model achieved a maximum accuracy of 99.72% in a comparative evaluation.

Therefore, this study will use PSO and PCA techniques to optimize the performance of RF. This study will compare the performance of PSO and RF with that of PCA and RF. PSO optimization on RF involves determining the number of trees, maximum depth, and others. Meanwhile, PCA on RF is used to reduce the number of useless features. This study uses a dataset sourced from SMS data consisting of two data classes: normal and phishing.

Based on the above background, the research question of this study is as follows: How does the performance of phishing detection techniques on social media perform when optimized using PSO and PCA?. The objectives of this study are as follows: to determine the performance of phishing detection techniques on social media when optimized using particle swarm optimization (PSO) and principal component analysis (PCA). The objective of this study can be achieved by comparing RF-PCA and RF-PSO in terms of Accuracy, Precision, and F-score performance metrics.

METHODS

Several studies have successfully demonstrated that optimizing phishing detection algorithms can improve performance, as seen in studies [5] that used PSO and [4] that used PCA. However, their research was not used on social media phishing, but rather to detect phishing URLs. Therefore, this study will utilize PSO and PCA techniques in the RF phishing detection algorithm to optimize the best approach for social media phishing.
In study [5], features extracted from URLs were mined using data mining rules. These features were selected based on the URL structure. The classification of features identified by the data mining rules was performed using the particle swarm optimization (PSO) technique. Feature selection with PSO optimization enables the identification of phishing URLs. By utilizing a large number of rule identifiers, this approach maximizes the true positive rate for identifying phishing URLs. Experiments have shown that feature selection using data mining and PSO is highly effective in identifying phishing URLs based on their structural characteristics.
Additionally, this approach can minimize processing time for identifying phishing websites. Therefore, this approach can be beneficial for identifying such URLs compared to previously proposed contemporary detection models. However, Model [5] has not been tested with more classification algorithms, and this could be a future area of study. Processing time for identifying phishing URLs is also a future aspect of research [5].

[4] provides a comprehensive examination of feature selection techniques using five diverse datasets. Various methods, including random forest (RF) selection from the model, SelectKBest with chi-square

statistics, principal component analysis (PCA), and recursive feature elimination (RFE), were used. The experiments, with a particular focus on PCA and the fourth dataset, revealed that all four models (RF, decision tree (DT), XGBoost, and multilayer perceptron) achieved 100% accuracy in detecting phishing URL attacks. This underscores the effectiveness of feature selection methods in enhancing a deeper understanding of the role of feature selection in improving the effectiveness of phishing detection systems across various datasets, highlighting the importance of leveraging techniques such as Principal Component Analysis (PCA) to achieve optimal results. However, the model proposed by [4] has the potential for overfitting and requires validation on a broader dataset.

The data collection stage involves gathering data sourced from text messages and messages received on social media platforms, such as WhatsApp, Telegram, X, and others. Once sufficient data has been collected, the next step is to assess each message to determine whether it is phishing or normal.

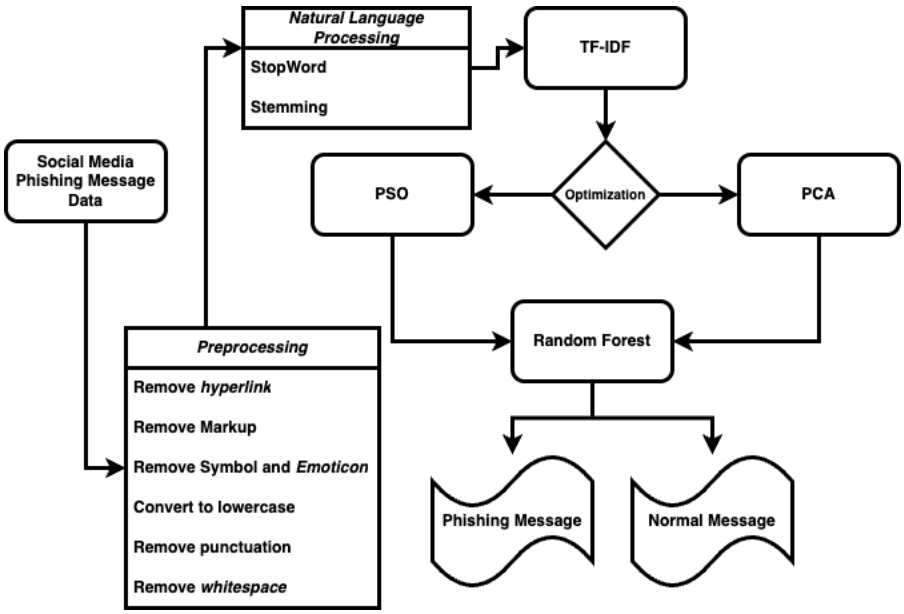

Figure 1. Propose model

In the data analysis stage, data preprocessing is performed to obtain data that is ready for processing in the NLP and TF-IDF processes (see Figure 1). Data preprocessing ensures that all raw data sourced from data collection can be used for model detection. After completing the preprocessing process, the next step is to perform stopword and stemming techniques. The stopword technique removes words that have no meaningful significance, such as hyphens, while the stemming technique returns words to their base form, such as "buying" = "buy," 'selling' = "sell," and so on. After applying stopword and stemming techniques, the next step is to apply the TF-IDF technique, which is a statistical method for measuring the importance of a term in a document relative to the corpus. After the data analysis stage is complete, the next step is to begin implementing the model. At this stage, the designed model is implemented. Model implementation is carried out on Google Colab Pro using the Python programming language. Based on the recommendations from the study [4] on preventing model overfitting, this study applies the Cross-Validation (CV) technique. CV is one of the commonly used techniques for testing models to avoid overfitting.

Additionally, [5] suggests observing the time when implementing the model. This will be relevant when the implementation process is carried out in system development. In the final stage of this study, model analysis was conducted to determine the best optimization technique when using the RF prediction technique. Model analysis considered the Precision, Accuracy, and F-Measure values for each optimization, namely RF-PCA and RF-PSO.

The parameter configuration used in this study is 80% training data and 20% test data. The PCA Optimization configuration used is n_components in PCA = 0.95, Default hyperparameters for RF, and random_state for RF = 42. Meanwhile, PSO optimization uses a configuration that is random_state for RF

= 42, Bounds, W = 0.9, c1 = 0.5, c2 = 0.3, n_particles = 5, iters = 30, scoring, n_estimators, max_depth, min_samples_split, min_samples_leaf.

RESULT AND DISCUSSION

The data collected and validated by cybersecurity experts is classified into phishing and non-phishing data. The amount of data successfully collected is 10,316 for the phishing class and 5,634 for the normal class. We utilize a server with Google's A100 specifications to run all processes in this research. The data preprocessing process involved removing most emojis and some non-standard symbols, removing all HTML tags to leave only plain text, converting all characters to lowercase, removing numbers, removing non-word characters, and removing underscores. Additionally, stopwords from the NLTK library and a stemmer from the Sastrawi library were used after data processing.

Figure 2. Word cloud for social media messages

In Figure 3, the initial iterations (0–3) show a significant decrease in cost from 0.0277 to 0.0275. Meanwhile, in the middle iterations (4–12), fine-tuning occurs, reaching a plateau at 0.02728. In the final iterations (13–29), high stability is achieved, with the cost only slightly decreasing to 0.0269 at iteration 29. This indicates that PSO quickly explores the hyperparameter space and identifies a good region within the first 10 iterations. Subsequently, PSO focuses on exploitation (fine-tuning), signaling stable convergence without overfitting. Finally, the final marginal decrease confirms the stability of the solution.

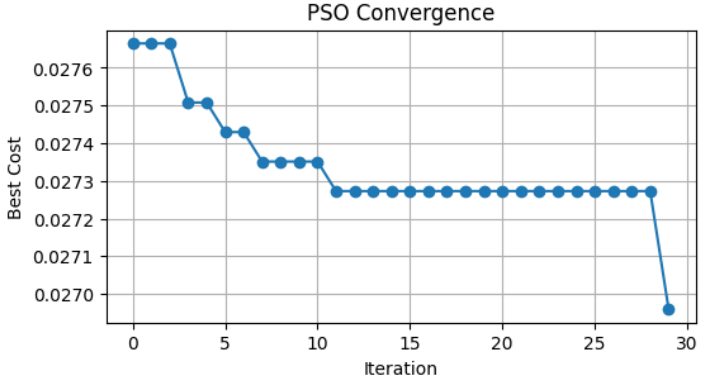

Figure 3. Random Forest Optimization by PSO

Figure 4 shows the confusion matrix for RF optimized using a combination of Gridsearch and PSO. Significantly few normal cases were incorrectly flagged as phishing, meaning that the false alarm rate was very low at 19 cases. However, some phishing data escaped detection in 84 cases, indicating the need for mitigation to reduce false negatives. Overall, the hybrid model provides an excellent balance of precision and recall, making it suitable for scenarios where both false positives and false negatives need to be reduced in the future.

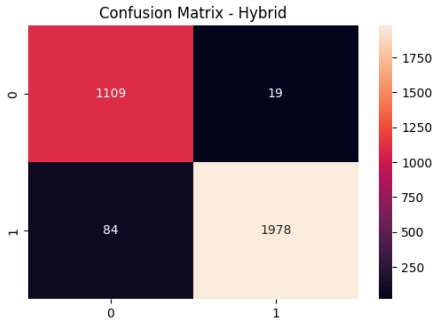

Figure 4. Confusion matrix for RF optimized by Gridsearch and PSO

In Figure 5, the RF model with Grid Search optimization obtained the highest Accuracy value (0.9840) and the best AUC (0.9985). Overall, this model made the fewest classification errors (both False Positives and False Negatives) and was very good at distinguishing between Phishing and normal classes. In addition, the Recall is also high (0.9869), meaning that False Negatives are rarely detected, with nearly all positives being detected. Meanwhile, the RF model with PCA optimization achieves a slightly lower Accuracy (0.9639) and the lowest Precision (0.9585). However, the Recall value (0.9869) is the same as the RF model with Grid Search optimization. This indicates that PCA removes important features, causing the model to generate False Positives more frequently, thereby reducing precision, although the positive detection rate (Recall) remains high.
Meanwhile, the RF model optimized by PSO produces nearly balanced metrics: precision (0.9877), recall (0.9728), and F1 (0.9802), all of which are high. The collaboration between RF optimized by Grid Search and PSO yields the highest precision (0.9905), meaning the fewest false positives. However, recall decreases (0.9593), indicating that there are relatively many false negatives, i.e., some positive cases are being missed.

	Model	Accuracy	Precision	Recall	F1	AUC
0	RF+PCA	0.963950	0.958549	0.986906	0.972521	0.994467
1	RF+PSO	0.974608	0.987691	0.972842	0.980210	0.996475
2	RF+Grid	0.984013	0.988344	0.986906	0.987624	0.998476
3	RF+Grid+PSO	0.967712	0.990486	0.959263	0.974624	0.995714

Figure 5. Comparison of performance values for each model

In Figure 6, a comparison of ROC and AUC is performed to show the accuracy values of each model. The AUC value in PCA optimization (0.994) indicates that the model is excellent and nearly perfect at separating the two classes after the dimension reduction process. Meanwhile, the PSO value (0.996) indicates that the hyperparameters optimized by PSO successfully maximize discrimination. Similarly, for Gridsearch (AUC = 0.998), the more exhaustive grid search identifies the optimal hyperparameter combination. This indicates very low false-positive rates and false-negative rates across various thresholds. However, the development of the Hybrid model (PSO and Gridsearch) was unable to surpass the AUC value of Gridsearch but was equivalent to the AUC value of PSO. The Grid Search value provides the best separation accuracy, while PSO/Hybrid balances very high accuracy with lower computation time.

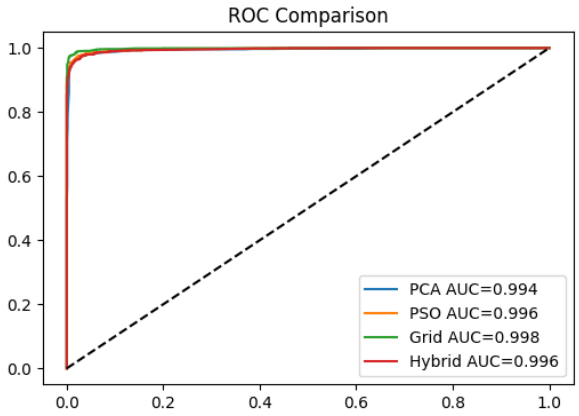

Figure 6. Comparison of RF optimized by PCA, PSO, Gridsearch, and Hybrid (PSO+Gridsearch)

CONCLUSION

This study successfully compared PSO and PCA optimization. Although there were no significant differences in the classification process, PSO and PCA made a real contribution to future research development. Additionally, we implemented a Gridsearch optimization as a comparison to PSO and PCA to determine the optimal combination of optimization methods. In terms of ranking, the best optimization is GridSearch, followed by PSO, and then PCA. This is because the most stable and superior performance across nearly all metrics is achieved with RF using Gridsearch optimization. However, the best combination for detecting all positive cases is either Random Forest (RF) with Gridsearch or RF with Principal Component Analysis (PCA). This study still has limitations during its implementation. Some developments for future research include addressing data imbalance, employing alternative feature selection methods, enhancing interpretability (to disclose the most influential features), and utilizing ensemble learning and building models in real-time systems.

REFERENCES

[1] A. Odeh, Q. A. Al-Haija, A. Aref, and A. A. Taleb, "Comparative Study of CatBoost, XGBoost, and LightGBM for Enhanced URL Phishing Detection: A Performance Assessment," *JISIS*, vol. 13, no. 4, pp. 1–11, Dec. 2023, doi: 10.58346/JISIS.2023.I4.001.

[2] S. S. Patil, N. M. Shekokar, and S. C. Iyer, "Design of Intelligent Feature Selection Technique for Phishing Detection," *IIUMEJ*, vol. 26, no. 1, pp. 254–277, Jan. 2025, doi: 10.31436/iiumej.v26i1.3337.

[3] S. Alsufyani and S. Alajmani, "A Deep Learning for Arabic SMS Phishing Based on URLs Detection," *ijacsa*, vol. 16, no. 1, 2025, doi: 10.14569/IJACSA.2025.0160138.

[4] P. Preeti and P. Sharma, "Enhancing phishing URL detection through comprehensive feature selection: a comparative analysis across diverse datasets," *IJEECS*, vol. 36, no. 2, p. 1182, Nov. 2024, doi: 10.11591/ijeecs.v36.i2.pp1182-1188.

[5] S. M. Alshahrani, "URL Phishing Detection Using Particle Swarm Optimization and Data Mining," *CMC*, vol. 73, no. 3, pp. 5625–5640, 2022, doi: 10.32604/cmc.2022.030982.

[6] 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," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 6425–6441, 2022, doi: 10.32604/cmc.2022.031625.

[7] E. Zhu, Z. Chen, J. Cui, and H. Zhong, "MOE/RF: A Novel Phishing Detection Model Based on Revised Multiobjective Evolution Optimization Algorithm and Random Forest," *IEEE Trans. Netw. Serv. Manage.*, vol. 19, no. 4, pp. 4461–4478, Dec. 2022, doi: 10.1109/TNSM.2022.3162885.

[8] R. C and T. M, "Improved Random Forest Algorithm using Chicken Swarm Optimization for Phishing Website Classification Model," *SSRG-IJEEE*, vol. 10, no. 4, pp. 141–151, Apr. 2023, doi: 10.14445/23488379/IJEEE-V10I4P114.

[9] K. Subashini and D. V. Narmatha, "MANTA RAY FORAGING OPTIMIZATION ALGORITHM WITH DEEP LEARNING ASSISTED AUTOMATED PHISHING URL DETECTION MODEL," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 24, 2023.

[10] B. B. Gupta *et al.*, "A Hybrid CNN-Brown-Bear Optimization Framework for Enhanced Detection of URL Phishing Attacks," *CMC*, vol. 81, no. 3, pp. 4853–4874, 2024, doi: 10.32604/cmc.2024.057138.
[11] P. Hemannth, M. Chinta, S. S. Satya, and P. Sri Aneelaja Devasena, "A Phishing URL Detection Model based on Horse Herd Optimization and Random Forest Algorithms," in *2024 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India: IEEE, May 2024, pp. 926–931. doi: 10.1109/ICPCSN62568.2024.00156.
[12] M. Diviya, M. Subramanian, and D. Gopala Krishnan, "An optimized phishing detection model using hybrid feature selection and a fine-tuned narrow neural network with dynamic jaya optimization to overcome cyberthreats," *Eng. Res. Express*, vol. 7, no. 1, p. 015202, Mar. 2025, doi: 10.1088/2631-8695/ada1a4.
[13] R. Abdillah, Z. Shukur, M. Mohd, T. S. M. Z. Murah, I. Oh, and K. Yim, "Performance Evaluation of Phishing Classification Techniques on Various Data Sources and Schemes," *IEEE Access*, vol. 11, pp. 38721–38738, 2023, doi: 10.1109/ACCESS.2022.3225971.
[14] B. N. Almousa and D. M. Uliyan, "Anti-Spoofing in Medical Employee's Email using Machine Learning Uclassify Algorithm," *IJACSA*, vol. 14, no. 7, 2023, doi: 10.14569/IJACSA.2023.0140727.
[15] 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.
[16] R. Meléndez, M. Ptaszynski, and F. Masui, "Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection," *Electronics*, vol. 13, no. 24, p. 4877, Dec. 2024, doi: 10.3390/electronics13244877.
[17] C. Thapa *et al.*, "Evaluation of Federated Learning in Phishing Email Detection," *Sensors*, vol. 23, no. 9, p. 4346, Apr. 2023, doi: 10.3390/s23094346.
[18] Q. Qi, Z. Wang, Y. Xu, Y. Fang, and C. Wang, "Enhancing Phishing Email Detection through Ensemble Learning and Undersampling," *Applied Sciences*, vol. 13, no. 15, p. 8756, Jul. 2023, doi: 10.3390/app13158756.
[19] P. Yasinta Roesmiatun and A. Zahra, "Enhancing detection of zero-day phishing email attacks in the Indonesian language using deep learning algorithms," *Bulletin EEI*, vol. 14, no. 1, pp. 505–512, Feb. 2025, doi: 10.11591/eei.v14i1.8759.
[20] 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," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 5901–5914, 2023, doi: 10.32604/cmc.2023.030784.