Computer Modeling in Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.056473

ARTICLE

Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems

Brij B. Gupta[1,2,3,4,*], Akshat Gaurav[5], Varsha Arya[6,7], Razaz Waheeb Attar[8], Shavi Bansal[9], Ahmed Alhomoud[10] and Kwok Tai Chui[11]

[1]Department of Computer Science and Information Engineering, Asia University, Taichung, 413, Taiwan

[2]Symbiosis Centre for Information Technology (SCIT), Symbiosis International University, Pune, 411057, India

[3]Center for Interdisciplinary Research, University of Petroleum and Energy Studies (UPES), Dehradun, 248007, India

[4]University Centre for Research and Development (UCRD), Chandigarh University, Chandigarh, 140413, India

[5]Computer Engineering, Ronin Institute, Montclair, NJ 07043, USA

[6]Department of Business Administration, Asia University, Taichung, 413, Taiwan

[7]Department of Electrical and Computer Engineering, Lebanese American University, Beirut, 1102, Lebanon

[8]College of Business Administration, Management Department, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

[9]Department of Research and Innovation, Insights2Techinfo, Jaipur, 302001, India

[10]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, 91911, Saudi Arabia

[11]Department of Electronic Engineering and Computer Science, Hong Kong Metropolitan University (HKMU), Hong Kong, 518031, China

*Corresponding Author: Brij B. Gupta. Email: bbgupta@asia.edu.tw

Received: 23 July 2024 Accepted: 26 September 2024 Published: 31 October 2024

ABSTRACT

Phishing attacks present a serious threat to enterprise systems, requiring advanced detection techniques to protect sensitive data. This study introduces a phishing email detection framework that combines Bidirectional Encoder Representations from Transformers (BERT) for feature extraction and CNN for classification, specifically designed for enterprise information systems. BERT's linguistic capabilities are used to extract key features from email content, which are then processed by a convolutional neural network (CNN) model optimized for phishing detection. Achieving an accuracy of 97.5%, our proposed model demonstrates strong proficiency in identifying phishing emails. This approach represents a significant advancement in applying deep learning to cybersecurity, setting a new benchmark for email security by effectively addressing the increasing complexity of phishing attacks.

KEYWORDS

Phishing; BERT; convolutional neural networks; email security; deep learning

1 Introduction

Given the growing integration and interdependence of information systems [1], the present situation of cybersecurity inside corporate information systems raises serious issues. Aiming to defend information assets from threats handled, stored, and transferred by internetworked information systems [2,3], cybersecurity refers to practices connected to enterprise IT security, information security, and cybersecurity [2,3]. Although there is general consensus on safeguarding computer and network systems from harmful malware and unauthorized access, use, disclosure, disruption, modification, or destruction to ensure integrity, confidentiality, and availability of information and information systems [2,4,5]. Reduced dependability of cybersecurity systems causes vulnerabilities in corporate information systems and breaches of data confidentiality, integrity, and availability [6,7]. Cybersecurity is intrinsically dependent on the preservation of business essential interests, human and intellectual capital, trade secrets, proprietary technologies, earnings, and market value [8]. Moreover, the analysis of cyber risks on the economic losses or expenses of the company helps to define the economic efficiency of corporate cybersecurity [9].

Ensuring the security of data, knowledge, and the performance of the industrial value chain depends on cybersecurity strategies being integrated with organizational efforts, information, and communication technologies [10,11]. Given their small, geographical, and familiar reach and financial resources [12], Small and Medium-sized Enterprises (SMEs) are seen as a blind spot in information security and cybersecurity management. The expansion of Secure Knowledge Management (SKM) systems feeds the ongoing need for qualified cybersecurity experts [13]. The manufacturing industry deals with urgent cybersecurity concerns, so automated threat analysis at every level of the company or business becomes very necessary [14]. Better data protection in companies and industrial control systems depends on cybersecurity [15,16]. Generally speaking, every company uses various analytics tools to document cybersecurity events and data breaches, hence adopting a varied posture [17,18]. Indeed, the use of email as a communication medium is important in enterprise information systems as it has evolved into a necessary tool for contemporary communication allowing direct, flexible, and quick information exchange [19,20]. However, the extensive usage of email has also made it a prominent target for many cyberattacks, especially phishing efforts. Using the weaknesses in the information technology system, phishing emails have greatly endangered users [21,22].

Phishing attacks include the use of misleading strategies in online communications wherein credible-looking emails are sent to people to fool them into disclosing sensitive personal information [23,24]. Emails have become susceptible to assaults like spamming [25–27] as their usage for corporate transactions and general communication increases. Furthermore, email correspondence has grown to be a major cybersecurity risk for companies as they allow harmful actions and cybersecurity breaches to be enabled [28,29].

The susceptibility of email to phishing attacks has also been noted in particular circumstances, such as in the distribution automation system, where the dedicated communication backbone can be readily accessed by adversaries, possibly leading to malware intrusion through phishing emails [22,30]. Furthermore, challenging to completely avoid are phishing emails, which use human weakness to imitate [31,32].

Email's importance as a communication medium is underlined even more by its usage in several fields, like mental health, where it is more prone to inadvertent leaks and losses than in-person contact [33–35]. Nonetheless, the susceptibility of email to security and privacy issues still causes great worry as enemies may use it to get private data and damage security [36,37].

In this context, in this study, we introduce a dual-model deep learning framework that leverages the strengths of both Bidirectional Encoder Representations from Transformers (BERT) for nuanced linguistic feature extraction and Convolutional Neural Network (CNN) for accurate and efficient classification to address the pervasive challenge of phishing in enterprise information systems.

The rest of the paper is organized as follows: Section 2 presents the related work. The details of the research methodology are presented in Section 3. Section 4 presents the results. Finally, Section 5 concludes the paper.

2 Related Work

Aiming to reduce the hazards related to false email communications, many frameworks have been suggested for the detection of phishing emails (Table 1). These systems improve phishing email detection and categorization by using cognitive principles, natural language processing, and machine learning, among other approaches. One such framework suggested integrating the Levenberg-Marquardt approach with feed-forward backpropagation using an anti-phishing enterprise environment model. This method uses an artificial neural network (ANN) in corporate settings to identify phishing emails [38]. In order to classify participant performance into either "good" or "bad," Li et al. [39] have also presented a machine learning framework including attribute reduction and 10-fold cross-valuation to grasp user behaviors when phishing attacks take place. Vajrobol et al. [40] proposed a mutual information-based model for phishing attack detection. Jain et al. [41] presented a content and URL analysis-based efficient approach to detect smishing SMS in intelligent systems. Gupta et al. [42] proposed an empowered email classifier for IoT-based systems in Industry 4.0. Heiding et al. [43] used large language models to devise and detect phishing emails. Hussain et al. [44] proposed an effective and lightweight phishing detection method based on multi-variant Convolutional Neural Networks (ConvNets).

Table 1: Summary of approaches for phishing email detection

Reference	Focus	Methodology	Key contributions
[38]	Anti-phishing enterprise model	ANN with feed-forward backpropagation and Levenberg-Marquardt	Detection of phishing emails within enterprise environments.
[39]	Machine learning for user behavior	Attribute reduction and 10-fold cross-validation	Classification of participant performance during phishing attacks into 'good' or 'bad'.
[45]	Fresh-Phish framework	Machine-learning data creation for phishing website detection	Enhanced auto-detection of phishing websites.
[46]	Neural network-based model	Character and word embeddings	Improved detection of deceptive emails.
[47]	Keras word embedding and CNN	Improved RCNN model	Enhanced phishing email detection model.

(Continued)

Table 1 (continued)

Reference	Focus	Methodology	Key contributions
[48]	Multilayer neural network	Focus on Vietnamese email context	Enhanced detection of phishing emails in Vietnamese contexts.
[49]	Text analysis and ML approach	Features specific to phishing	Improved accuracy in phishing email detection.
[50]	Intelligent cybersecurity system	Deep learning techniques	Emphasized the need for more effective phishing detection technology.
[51]	Dynamic and combined technique	THEMIS deep learning algorithm	Effective email phishing prediction.
[52]	Psycholinguistic features analysis	Analysis of emails for deception features	Improved spam and phishing detection technologies.
[53]	Cognitive neuroscience of phishing	Phishing Email Suspicion Test (PEST)	Evaluation of cognitive mechanisms in phishing detection.
[54]	Phishing training program	Short online training program	Highlighted the importance of user training in identifying phishing attempts.
[55]	Multimodal alert system	Development of audio, visual, and haptic phishing email alerts	Potential enhancement of phishing detection through multimodal alerts.

Moreover, Shirazi et al. [45] have created the "Fresh-Phish" framework, a better method for producing current machine-learning data for phishing websites, thus improving phishing website auto-detection. Using character and word embeddings to improve the identification of false emails, Stevanović et al. [46] also presented a neural network-based phishing email detection model. Also, Fang et al. [47] presented a Keras word embedding and convolutional neural network (CNN) enhanced phishing email detecting model. Presenting an association rules-based method for anomaly identification on Controller Area Network (CAN)-bus, D'Angelo et al. [56] designed especially for Software-Defined Networking (SDN), enabled IoT scenarios.

Abid et al. [57] presented a Multi-Level Deep Neural Network suited for this aim. Building on this, Abid et al. [58] presented the ECMT Framework, which provides a complete methodology for IoT security by combining in-memory attribute inspection and advanced neural network designs with hybridized machine learning approaches. These contributions match with larger initiatives to use Information and Communications Technologies (ICT) for sustainable development objectives (Wu et al.) [59], so addressing urgent issues in IoT security. Furthermore, stressing the requirement of

green solutions in managing the enormous data created by IoT networks, Wu et al. [60] underlined the need to tackle environmental issues in big data processing. These papers highlight the many dimensions of IoT research and its consequences for both security and environmental sustainability by bridging the gap between security, sustainability, and technical innovation.

Furthermore, Xuan [48] suggested a multilayer neural network architecture for email phishing detection in an effort to improve phishing email detection in Vietnamese environments. Likewise, Oladimeji [49] provided a text analysis and machine learning method for phishing email detection, including phishing-specific characteristics to improve detection accuracy. Emphasizing the necessity of more efficient phishing detection technology to help to reduce the rising danger of phishing emails, Mughaid et al. [50] have created an intelligent cybersecurity phishing detection system utilizing deep learning methods. Apart from machine learning-based solutions, Hema [51] presented a dynamic and integrated phishing detection method, therefore presenting the THEMIS deep learning algorithm for efficient email phishing prediction. Furthermore, Xu et al. [52] underlined the need to evaluate emails for psycholinguistic traits linked with dishonesty to fine-tune and enhance spam and phishing detecting tools. These systems show the many methods and tools used to handle the phishing email detection problem. Furthermore, various research has looked into the cognitive processes and behavioral signals of false positives and phishing sensitivity. The Phishing Email Suspicion Test (PEST), a lab-based test for assessing the cognitive neuroscience of phishing detection, was first presented by Hakim et al. [53]. This helped to clarify the cognitive processes engaged in phishing detection. Emphasizing the need to teach users to recognize phishing efforts, Weaver et al. [54] investigated the efficacy of a brief online phishing training program meant to assist users in detecting phishing emails.

Moreover, Cooper et al. [55] described the empirical findings of subject-matter experts on the construction of an auditory, visual, and haptic phishing email alert system, thereby stressing the possibilities of multimodal warning systems to improve phishing detection. This research highlights the interdisciplinary character of phishing email detection, including technical, behavioral, and cognitive elements. To speed processing in home UbiHealth systems, Sarivougioukas et al. [61] suggested the integration of fused contextual data with threading technology. Chopra et al. [62] investigated generative adversarial networks for text-to-image synthesis. To improve hazy visibility in visual IoT-driven intelligent transportation systems, Liu et al. [63] used deep network-enabled methods Using deep belief networks with ResNet models for maximum security, Nguyen et al. [64] investigated the use of secure blockchain-enabled Cyber-Physical System (CPS) in healthcare. Gupta et al. [65] presented a fresh method for real-time lexical-based machine learning phishing URL detection. Examining phishing attack strategies, defensive mechanisms, and open research issues, Jain et al. [66] furthermore stressed the need to reduce false information in linked systems, Zhang et al. [67] created a deep learning-based quick fake news detecting model for cyber-physical social services. Together, these projects help to further CPS technology's knowledge and use in many spheres.

3 Research Methodology

Our proposed approach is divided into two phases. The first phase is used to extract the features from the email text, and then the second phase uses the CNN model to classify the features. The details of the phases are as follows:

Algorithm 1: Feature extraction using BERT for phishing email detection

```
Load BERT model and tokenizer
Initialize an empty list for features
procedure ExtractFeatures(text)
    Tokenize text with special tokens and truncation
    Convert tokens to tensor and move to device
    Pass tensor through BERT model
    Extract hidden states
    Concatenate the last four layers of hidden states
    Compute mean of concatenated hidden states
    Append result to the features list
end procedure
for each email text in the dataset do
    Extract features using ExtractFeatures
end for
Combine features into a tensor
Return feature tensor
```

3.1 Feature Extraction

Using the BERT model (Algorithm 1), the first step of our proposed method for phishing email detection uses a feature extraction algorithm. The method starts with pre-trained BERT model loading along with a matching tokenizer. Every email text in the collection is tokenized to guarantee the inclusion of special tokens and adherence to a certain maximum length. After that, the tokenized text is turned into a tensor and run through the BERT model to create a sequence of hidden states for every token.

Our method depends critically on the aggregation of the final four layers of hidden states derived from the BERT model output. We get a rich picture of the contextual interactions within the text by concatenating these layers. We derive a single feature vector that captures the fundamental linguistic traits of the email content by means of the mean of these concatenated hidden states. This vector is subsequently added to a list of features that, when aggregated following completion of processing the whole dataset, forms a master feature tensor.

Input for the next classification step is this tensor, which captures the condensed language core of our dataset. The capacity of our model to distinguish between regular and phishing emails with great accuracy depends on the efficacy of this feature extraction process, which is described in the methodology and provides a strong basis for the difficult job of email classification within corporate information systems.

3.2 CNN for Phishing Email Classification

Building on the strong feature set obtained using BERT, the second step of our proposed method consists in the use of a CNN for the email categorization into “Normal” and “Phishing” categories. First reshaped to fit CNN input criteria, the feature tensor guarantees that every feature vector is suitably arranged for convolutional operation. Then, horizontally concatenated with the matching labels, this ordered dataset becomes a whole dataset suitable for model training and assessment.

Algorithm 2: Application of CNN for phishing email classification

```
Reshape feature tensor for CNN input
Concatenate features with labels to form the dataset
procedure SplitDataset(dataset)
    Split dataset into features and labels
    Split into training and testing sets
    Convert splits to tensors
end procedure
procedure ConstructCNNModel
    Define CNN with layers and parameters
    Include ReLU activation and MaxPooling
    Apply flattening layer for FC layers
    Use dropout for regularization
    Define output layer for classification
end procedure
Initialize CNN model
Train model with training data
Evaluate on testing data
```

Following an 80–20 split, the dataset is separated into feature and label sets as stated in the method, which are further split into training and testing subsets. This division has greatly aided in validating the performance of the model on unprocessed data. Tensors-the necessary form for CNN processing are then derived from the training and testing sets. Multiple layers built to capture the hierarchical patterns in the data help to build the CNN model itself. It consists of ReLU activations in convolutional layers followed by max pooling to reduce the feature maps; a flattening step then moves the convolutional output to fully connected layers. Dropout is deliberately included to avoid overfitting and guarantee that the model fits fresh data. Designed for binary classification, the last layer generates the likelihood that an email is a phishing effort. The model adjusts its parameters to reduce classification error after initialisation by means of a training process on the training set. The trained model is then put through the testing set and its performance is evaluated closely. The whole process, as captured in the algorithm, underlines the model's high accuracy in identifying and classifying phishing emails, therefore greatly enhancing the general effectiveness of the phishing detection system in corporate information systems.

4 Results and Discussion

4.1 System Specifications

Six physical core and twelve logical core Intel64 Family 6 Model 151 Stepping 5 processors were running Windows 10 (Version 10.0.226), under which the model development took place. The machine has 16 GB of Random-access memory (RAM), of which 8.25 GB was accessible during the tests, therefore using 51.2% of the total capacity. The storage arrangement included a 50.3% used 500 GB disk. Operating with a low load and a steady temperature of 36°C, the NVIDIA GeForce RTX 3050 with 8 GB of RAM was used for computational chores needing graphics processing unit (GPU) acceleration. Python 3.9 with the following libraries-PyTorch 2.2.1 for deep learning model implementation, Pandas 1.5.3 and NumPy 1.24.3 for data manipulation and preprocessing, and scikit-learn 1.4.1.post1 for other machine learning tools constituted the software environment. This arrangement gave a strong and effective stage for putting the suggested BERT and CNN-based phishing detection system into use and evaluating it.

4.2 Dataset Representations

We used the Kaggle data set to test our proposed model [68]. The dataset is imbalanced, so we used SMOTE to balance the dataset. We created word clouds from the dataset to get an understanding of the textual patterns and frequent terms related to phishing and non-phishing emails.

Fig. 1 displays the word cloud produced from the dataset, therefore stressing the most often occurring terms in the emails. Every word's frequency determines its size; hence, bigger words show more often in the dataset. Independent of their classification, this image offers a summary of the shared language used in emails.

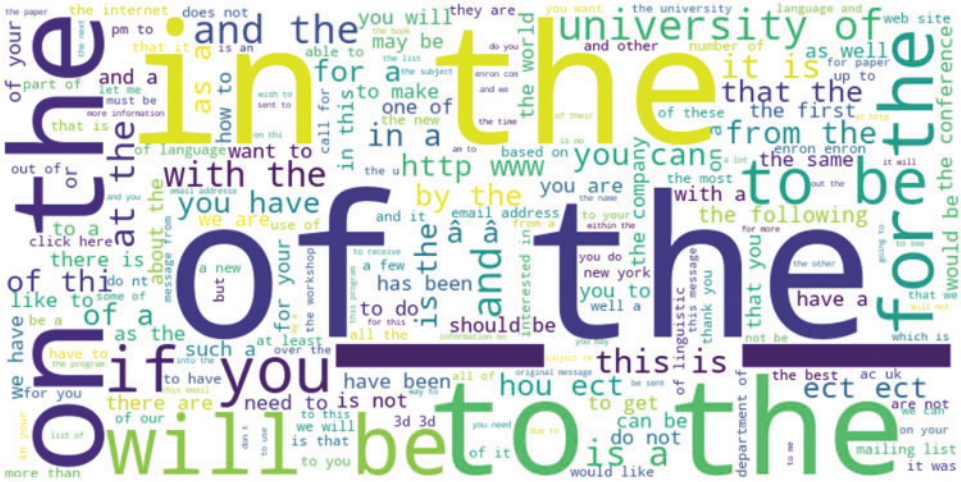

Figure 1: Word cloud

Fig. 2 shows a word cloud emphasizing distinctive phrases and keywords that could point to phishing or safe emails. Comparatively, to more neutral or less common phrases in secure emails, we may see variations in word use patterns that are unique to phishing emails by comparing these two-word clouds-terms associated with urgency or action demands (e.g., "urgent," "account," "login").

Figure 2: Unique word cloud

4.3 Accuracy and Loss Variation

We train the CNN model for five epochs to make it lightweight. Fig. 3 shows the model's performance in terms of both accuracy and loss tracked across the training and test periods. The training loss was 0.1932 with an accuracy of 92.58% at first epoch 0; the testing phase produced a loss of 0.1362 with an accuracy of 94.86%. One could clearly see a development as the program went on. By epoch 4, the training loss dropped to 0.0695 with a matching 97.50% accuracy. The test loss dropped simultaneously to 0.0802, with a test accuracy of 97.39%. The capacity of the model to generalize well from the training data to unknown test data is shown by the continuous decrease in loss and rise in accuracy across the epochs.

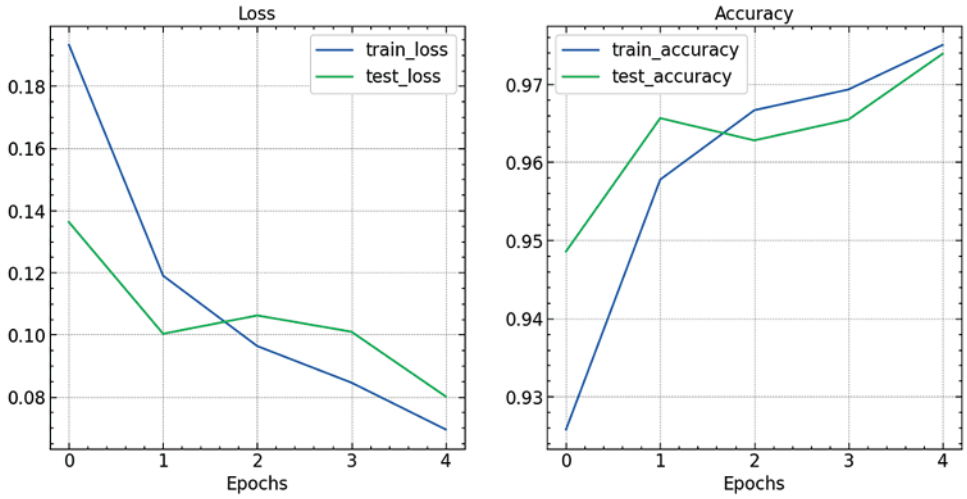

Figure 3: Accuracy and loss

4.4 Classification Report

For both Normal and Fault, the classification report, shown in Fig. 4, offers a thorough analysis of the model's performance across many criteria, including accuracy, recall, and F1-score. For the "Normal" class, which included 1295 samples, the model attained an F1-score of 0.96, a recall of 0.95, and an accuracy of 0.98. With a larger sample size of 2213, the "Fault" class showed somewhat less accuracy at 0.97 but greater recall and F1-score at 0.99 and 0.98, respectively. With an overall accuracy of 97%, the model performed quite strongly for both classes. Further underlining the model's balanced performance are the steady macro average and weighted average scores for accuracy, recall, and F1-score at 0.97. These measures verify that the model is well-calibrated to differentiate between normal and fault situations in the dataset.

4.5 Confusion Matrix

Confusion matrix in Fig. 5 offers a graphic overview of our suggested model's classification performance. Out of 1295 cases in the "Normal" class, the model accurately predicted 1233 cases, 62 cases misclassified as "Fault," according to the matrix. On the other hand, for the "Fault" class, out of 2213 cases, the model correctly predicted 2183 occurrences, with only 30 cases wrongly labeled as "Normal." This reflects the great capacity of the model to accurately recognize both normal and fault situations, hence producing a high true positive rate for both classes. With few false positives

and false negatives, the modest amount of misclassifications points to the model's dependability and efficacy for the intended phishing detection choreography.

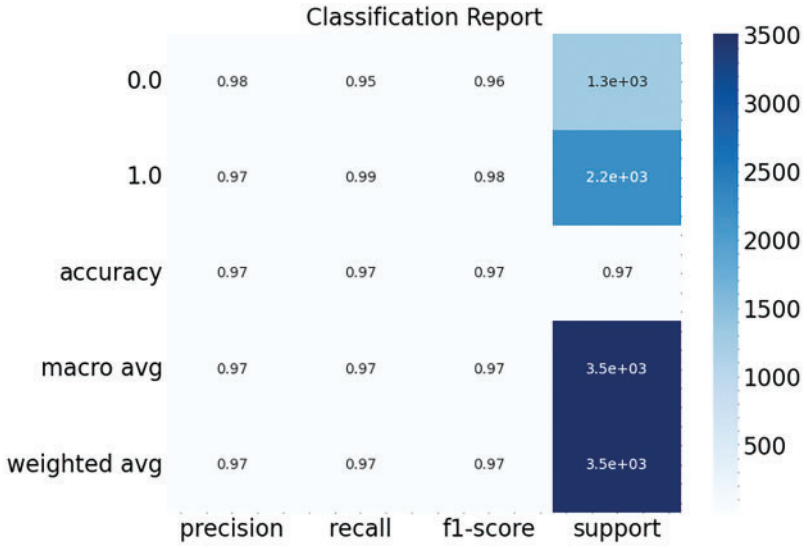

Figure 4: Proposed approach classification report

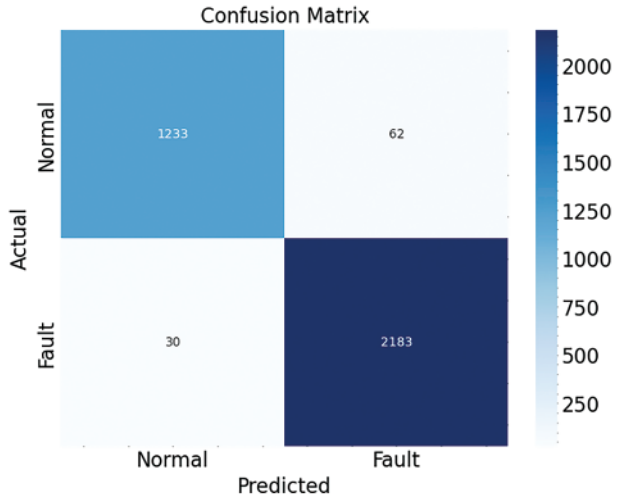

Figure 5: Proposed approach confusion matrix

4.6 Comparative Analysis

We assessed the performance of six common machine learning and deep learning models: Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Logistic Regression, Support Vector Machine (SVM), and a Simple Neural Network in order to compare their efficacy in phishing detection within corporate systems. Confusion matrices and classification reports let the models be evaluated holistically, therefore revealing their strengths and shortcomings.

4.6.1 Gated Recurrent Unit

With 97% the GRU model showed the best accuracy among all the evaluated models. Suggesting that it is equally successful in identifying both sorts of emails, the GRU model maintained high accuracy (0.97), recall (0.97), and F1-scores across both "Normal" and "Fault" classes as shown in the classification report (Fig. 6). With just a limited number of misclassifications, e.g., 46 misclassified "Normal" instances and 66 misclassified "Fault" instance, the confusion matrix (Fig. 7) shows that GRU properly classified most of the cases. The low incidence of false positives and false negatives emphasizes the model's resilience in practical situations where high-accuracy phishing email detection is very vital.

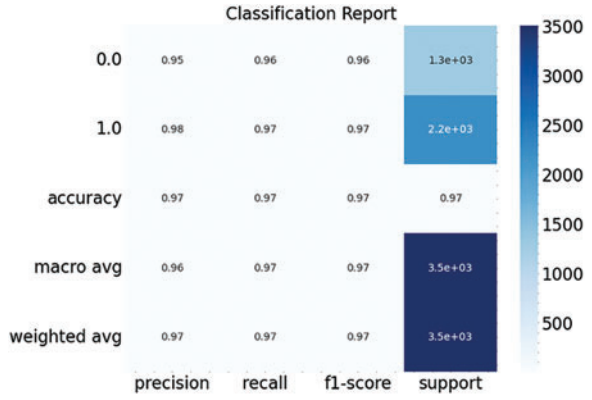

Figure 6: GRU classification report

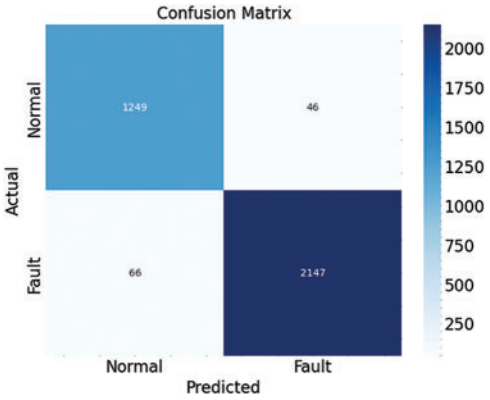

Figure 7: GRU confusion matrix

4.6.2 Long Short-Term Memory

Complementably accurate at 96%, the LSTM model followed the GRU rather closely. Particularly for the "Normal" class, where the F1-score was much lower, the classification report (Fig. 8) shows a somewhat smaller accuracy and recall than GRU. Although LSTM identified most cases very well, it struggled somewhat more than GRU in differentiating between "Normal" and "Fault" emails. Fig. 9 shows this confusion matrix. LSTM is nonetheless a good choice, especially in situations where sequence relationships in data could be more evident.

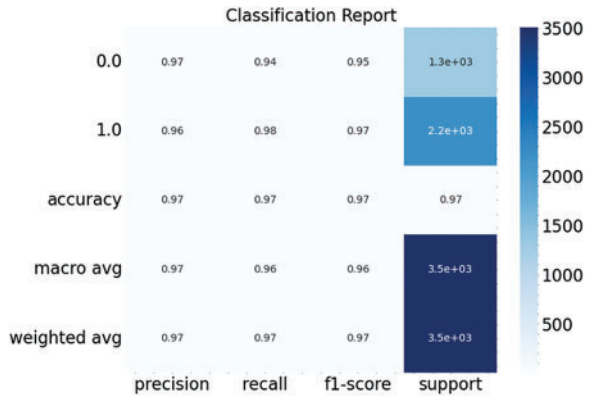

Figure 8: LSTM classification report

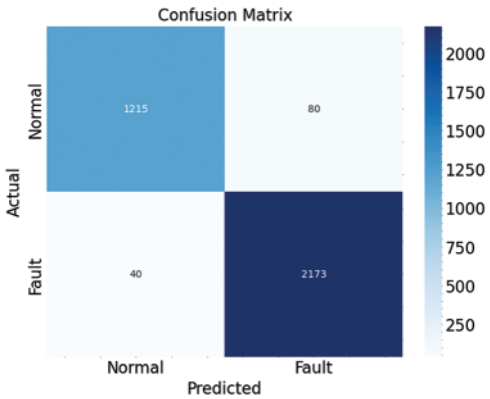

Figure 9: LSTM confusion matrix

4.6.3 Recurrent Neural Network

Though strong, RNNs exhibited a little drop in accuracy when compared to GRU and LSTM; however, they ranked at 96%. Although competent, the classification report (Fig. 10) and confusion matrix (Fig. 11) show that RNNs may not manage long-term dependencies as effectively as GRU or LSTM, hence causing more misclassifications. With a larger incidence of false negatives, the confusion matrix displays more noteworthy misclassification in the "Normal" class. This implies that, especially when correct long-term context knowledge is vital, RNNs may not be as dependable as GRU and LSTM in phishing detection chores even if they can perform well.

4.6.4 Logistics Regression

Logistic Regression attained a 95% accuracy level, as a classic machine learning model. Although recall and accuracy are adequate, the classification report (Fig. 12) demonstrates that deep learning models lead in both respects. Given its lower F1-score, logistic regression especially had difficulty spotting "Fault" events. Examining the confusion matrix (Fig. 13), one finds that many "Fault" events were misclassified as "Normal." In a security setting, where failure to identify phishing emails might have dire consequences, this greater proportion of false negatives is troubling. The performance of logistic regression shows how limited conventional models are in intricate, real-world situations such as phishing identification.

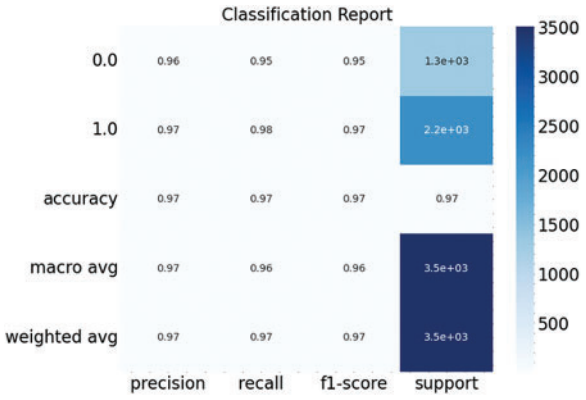

Figure 10: RNN classification report

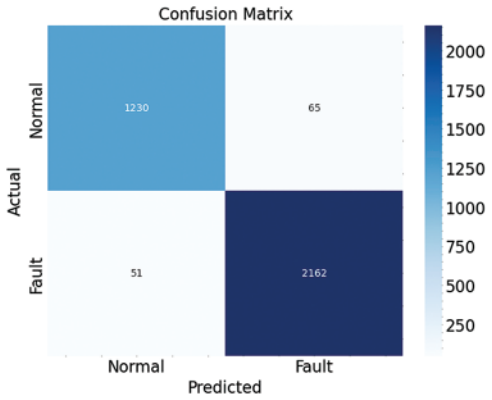

Figure 11: RNN confusion matrix

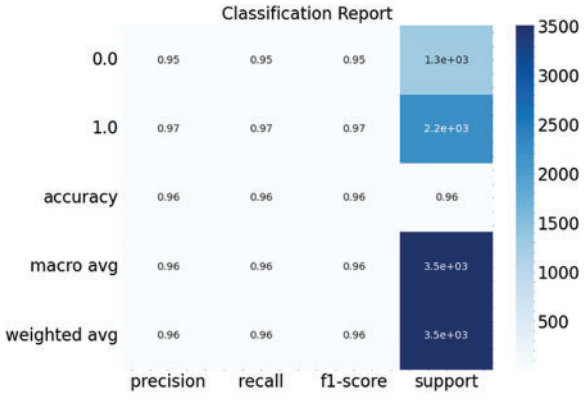

Figure 12: LR classification report

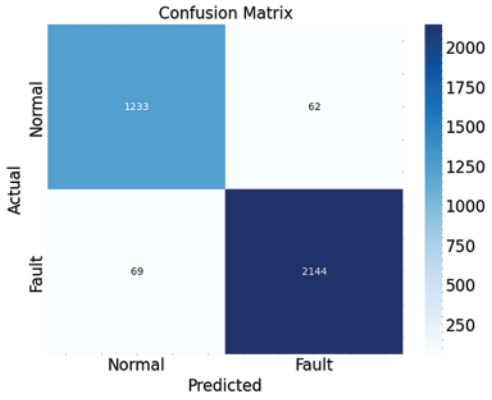

Figure 13: LR confusion matrix

4.6.5 Support Vector Management

With a 95% accuracy, SVM behaved much like Logistic Regression. Though it still falls short of the deep learning models, the classification report (Fig. 14) shows that SVM has a somewhat higher accuracy and recall than Logistic Regression. With a substantial number of misclassified cases, the confusion matrix (Fig. 15) indicates that SVM likewise battled the "Fault" class. Although SVM might be useful in smaller or more organized datasets, given the complex and variable nature of phishing detection in corporate systems it might not be the ideal option.

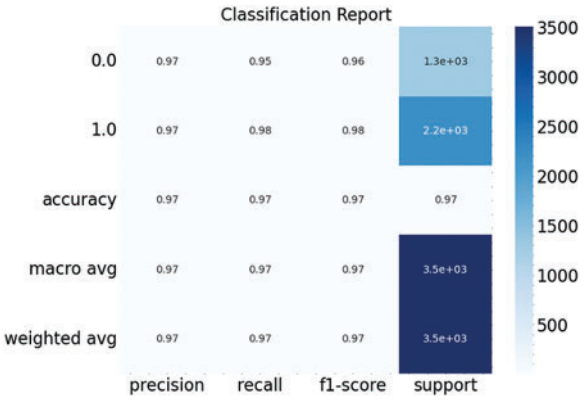

Figure 14: SVM classification report

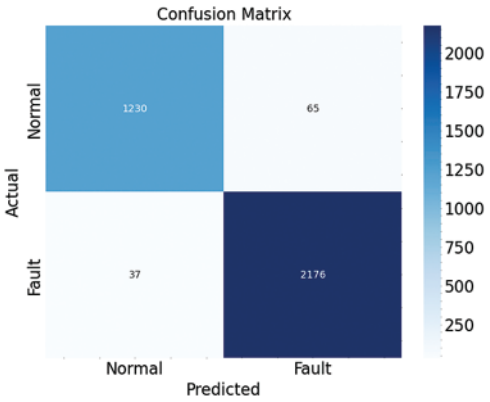

Figure 15: SVM confusion matrix

Now, we performed a comparison study against six conventional models: GRU, LSTM, RNN, Logistic Regression, SVM, and a Simple Neural Network to verify our suggested technique even further. Figs. 16 and 17 provide two important metrics, accuracy and loss, which the comparison focused on.

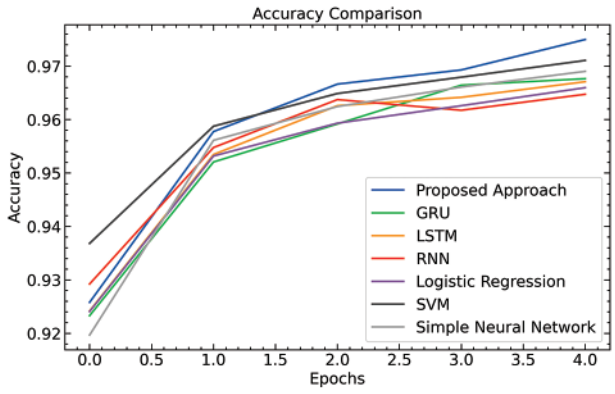

Figure 16: Accuracy comparison

Fig. 16 shows that in terms of accuracy across the training epochs, the suggested method regularly exceeded all other models. The suggested method first had the same accuracy as the other models, but it soon changed to reach the maximum accuracy of over 97.5% by the conclusion of the fifth epoch. Though with somewhat lower ultimate accuracy, the GRU and LSTM models also showed great performance, closely following the suggested method. Although initially competitive, RNN

demonstrated a plateau in accuracy during the third epoch, indicating difficulties in further learning. By comparison, along with the Simple Neural Network, conventional machine learning models such as Logistic Regression and SVM showed slower development across the epochs and began with lower beginning accuracies. These models had lagged by the fifth epoch; Logistic Regression showed the most notable difference.

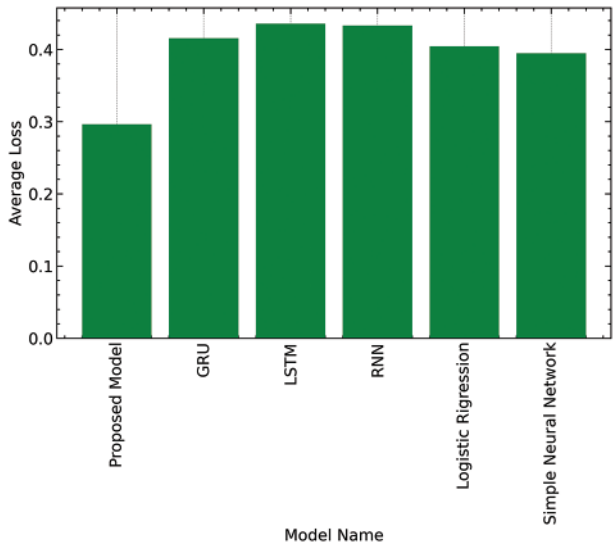

Figure 17: Loss comparison

Given that the suggested method regularly maintained the lowest loss values across the training period, the loss comparison shown in Fig. 17 emphasizes even more, its better performance. Although GRU and SVM exhibited competitive loss values, they were still somewhat greater than the suggested method. While they were good in learning, LSTM and RNN showed modest loss values, meaning that they were not as optimized as GRU or the suggested method in lowering error. Conversely, logistic regression and the simple neural network showed the best loss values throughout the epochs. This emphasizes the restrictions of these models for difficult tasks, including phishing detection, and reflects their reduced power to generalize from the training data, which corresponds with their lower accuracy performance.

5 Conclusion

Using the combined powers of BERT for subtle feature extraction and CNN for precise classification, our work offers a fresh method for phishing email detection within corporate information systems. While CNN has shown great accuracy and recall in classification tests, the BERT model efficiently extracted linguistic features from the email corpus. Based on a 93% accuracy rate, our findings imply that deep learning methods may greatly improve phishing attempt detection. The classification report and confusion matrix reveal the strong resilience of our model, which emphasizes its possible dependability as a tool for cybersecurity in corporate settings. Our proposed models detects the phishing emails efficiently, however, still there is a scope of improvement. In this context, we will try to update the model parameters to detect the phishing email more efficiently. In the future, we will concentrate on testing our model with many datasets.

Acknowledgement: The work described in this paper was supported by a grant from Hong Kong Metropolitan University (RD/2023/2.3). The work was supported Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R 343), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, Kingdom of Saudi Arabia for funding this research work through the project number "NBU-FFR-2024-1092-09".

Funding Statement: The work described in this paper was supported by a grant from Hong Kong Metropolitan University (RD/2023/2.3). The work was supported Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R 343), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, Kingdom of Saudi Arabia for funding this research work through the project number "NBU-FFR-2024-1092-09".

Author Contributions: Final manuscript revision, funding, supervision: Brij B. Gupta, Kwok Tai Chui; study conception and design, analysis and interpretation of results, methodology development: Akshat Gaurav, Varsha Arya, Shavi Bansal; data collection, draft manuscript preparation, figure and tables: Ahmed Alhomoud, Razaz Waheeb Attar. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All data generated or analysed during this study are included in this published article.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Alrubaiq A, Alharbi T. Developing a cybersecurity framework for e-government project in the kingdom of Saudi Arabia. J Cybersecur Priv. 2021;1(2):302–18. doi:10.3390/jcp1020017.
2. Malatji M, Solms S. Cybersecurity capabilities for critical infrastructure resilience. Inf Comput Secur. 2021;30(2):255–79. doi:10.1108/ICS-06-2021-0091.
3. Singh G, Malhotra M, Sharma A. An adaptive mechanism for virtual machine migration in the cloud environment. Int J Cloud Appl Comput. 2022;12(1):1–10. doi:10.4018/IJCAC.297095.
4. Chigada J, Madzinga R. Cyberattacks and threats during COVID-19: a systematic literature review. Sa J Inform Manage. 2021;23(1):1401. doi:10.4102/sajim.v23i1.1277.
5. Gupta T, Panda SP. Cloudlet and virtual machine performance enhancement with clara and evolutionary paradigm. Int J Cloud Appl Comput. 2022;12(1):1–16. doi:10.4018/IJCAC.298322.
6. Maesaroh S, Permana H, Febrianaga P, Pardosi R. Blockchain technology in the future of enterprise security system from cybercrime. Blockchain Front Technol. 2022;2(1):1–8. doi:10.34306/bfront.v2i1.88.
7. Barthwal V, Rauthan MS, Varma R. SMA-LinR: an energy and sla-aware autonomous management of virtual machines. Int J Cloud Appl Comput. 2022;12(1):1–24. doi:10.4018/IJCAC.2022010103.
8. Daoud M, Serag A. A proposed framework for studying the impact of cybersecurity on accounting information to increase trust in the financial reports in the context of Industry 4.0: an event, impact and response approach. Commer Finance. 2022;42(1):20–61. doi:10.21608/caf.2022.251730.

9. Zadorozhnyi Z, Muravskyi V, Shevchuk O, Bryk M. Innovative accounting methodology of ensuring the interaction of economic and cybersecurity of enterprises. Market Manage Innov. 2021;5:36–46. doi:10.21272/mmi.
10. Junqueira B, Souza N, Lima V, Gonçalves A, Lepikson H. Learning proposal for cybersecurity for industrial control systems based on problems and established by a 4.0 didactic advanced-manufacturing-plant. J Bioeng Technol Health. 2022;5(1):11–7. doi:10.34178/jbth.v5i1.188.
11. Dwivedi RK. Density-based machine learning scheme for outlier detection in smart forest fire monitoring sensor cloud. Int J Cloud Appl Comput. 2022;12(1):1–16. doi:10.4018/IJCAC.305218.
12. Antunes M, Maximiano M, Gomes R, Pinto D. Information security and cybersecurity management: a case study with smes in portugal. J Cybersecur Priv. 2021;1(2):219–38. doi:10.3390/jcp1020012.
13. Daniel C, Mullarkey M, Agrawal M. Rq labs: a cybersecurity workforce skills development framework. Inform Syst Front. 2023;25(2):431–50.
14. Zarreh A, Wan H, Lee Y, Saygin C, Janahi R. Risk assessment for cyber security of manufacturing systems: a game theory approach. Procedia Manuf. 2019;38:605–12. doi:10.1016/j.promfg.2020.01.077.
15. Haleem A, Javaid M, Singh R, Rab S, Suman R. Perspectives of cybersecurity for ameliorative Industry 4.0 era: a review-based framework. Ind Robot Int J Robot Res Appl. 2022;49(3):582–97. doi:10.1108/IR-10-2021-0243.
16. Kumar K, Thaman J. Improving virtual machine migration effects in cloud computing environments using depth first inspired opportunity exploration. Int J Cloud Appl Comput. 2022;12(1):1–22. doi:10.4018/IJCAC.314209.
17. Walker-Roberts S, Hammoudeh M, Aldabbas O, Aydin M, Dehghantanha A. Threats on the horizon: understanding security threats in the era of cyber-physical systems. J Supercomput. 2020;76:2643–4.
18. Dwivedi RK, Kumar R, Buyya R. Gaussian distribution-based machine learning scheme for anomaly detection in healthcare sensor cloud. Int J Cloud Appl Comput. 2021;11(1):52–72. doi:10.4018/IJCAC.2021010103.
19. Nicolaou C, Matsiola M, Kalliris G. Technology-enhanced learning and teaching methodologies through audiovisual media. Educ Sci. 2019;9:196. doi:10.3390/educsci9010062.
20. Hu B, Gaurav A, Choi C, Almomani A. Evaluation and comparative analysis of semantic web-based strategies for enhancing educational system development. Int J Semant Web Inform Syst. 2022;18(1):1–14. doi:10.4018/IJSWIS.302895.
21. Suzuki Y, Monroy S. Prevention and mitigation measures against phishing emails: a sequential schema model. Secur J. 2021;35:1162–82. doi:10.1057/s41284-021-00318-x.
22. Ismail S, Shishtawy TE, Alsammak AK. A new alignment word-space approach for measuring semantic similarity for arabic text. Int J Semant Web Inform Syst (IJSWIS). 2022;18(1):1–18. doi:10.4018/IJSWIS.
23. Hassandoust F, Singh H, Williams J. The role of contextualization in individuals' vulnerability to phishing attempts. Australas J Inf Syst. 2020;24. doi:10.3127/ajis.v24i0.2693.
24. Almomani A, Alauthman M, Shatnawi MT, Alweshah M, Alrosan A, Alomoush W, et al. Phishing website detection with semantic features based on machine learning classifiers: a comparative study. Int J Semant Web Inform Syst. 2022 Jan 1;18(1):1–24. doi:10.4018/IJSWIS.
25. Abdulhamid S, Shuaib M, Osho O, Idris I, Alhassan J. Comparative analysis of classification algorithms for email spam detection. Int J Comput Netw Inf Secur. 2018;10(1):60–7. doi:10.5815/ijcnis.2018.01.07.
26. Barbosa A, Bittencourt II, Siqueira SW, Dermeval D, Cruz NJ. A context-independent ontological linked data alignment approach to instance matching. Int J Semant Web Inform Syst. 2022;18(1):1–29. doi:10.4018/IJSWIS.295977.
27. Tembhurne JV, Almin MM, Diwan T. Mc-DNN: fake news detection using multi-channel deep neural networks. Int J Semant Web Inform Syst. 2022;18(1):1–20. doi:10.4018/IJSWIS.295553.

28. Michael A, Eloff J. Discovering "insider it sabotage" based on human behaviour. Inf Comput Secur. 2020;28:575–89. doi:10.1108/ICS-12-2019-0141.
29. Li S, Qin D, Wu X, Li J, Li B, Han W. False alert detection based on deep learning and machine learning. Int J Semant Web Inform Syst. 2022 Jan 1;18(1):1–21. doi:10.4018/IJSWIS.297035.
30. Choi I, Hong J, Kim T. Multi-agent based cyber attack detection and mitigation for distribution automation system. IEEE Access. 2020;8:183495–504. doi:10.1109/ACCESS.2020.3029765.
31. Petrič G, Roer K. The impact of formal and informal organizational norms on susceptibility to phishing: combining survey and field experiment data. Telematics Inform. 2022;67:101766. doi:10.1016/j.tele.2021.101766.
32. Bhardwaj A, Kaushik K. Predictive analytics-based cybersecurity framework for cloud infrastructure. Int J Cloud Appl Comput. 2022;12(1):1–20. doi:10.4018/IJCAC.297106.
33. Lustgarten S, Garrison Y, Sinnard M, Flynn A. Digital privacy in mental healthcare: current issues and recommendations for technology use. Curr Opin Psychol. 2020;36:25–31. doi:10.1016/j.copsyc.2020.03.012.
34. Gallo L, Gentile D, Ruggiero S, Botta A, Ventre G. The human factor in phishing: collecting and analyzing user behavior when reading emails. Comput Secur. 2024;139:103671. doi:10.1016/j.cose.2023.103671.
35. Sturman D, Bell EA, Auton JC, Breakey GR, Wiggins MW. The roles of phishing knowledge, cue utilization, and decision styles in phishing email detection. Appl Ergon. 2024;119:104309. doi:10.1016/j.apergo.2024.104309.
36. Kisembo IM, Ocen GG, Bongomin O, Alunyu AE, Nibikora I, Matovu D, et al. An algorithm for improving email security on the android operating system in the Industry 4.0 era. J Eng. 2021;2021(1):4690611. doi:10.1155/2021/4690611.
37. Altwaijry N, Al-Turaiki I, Alotaibi R, Alakeel F. Advancing phishing email detection: a comparative study of deep learning models. Sensors. 2024;24(7):2077. doi:10.3390/s24072077.
38. Sankhwar S, Pandey D, Khan RA, Mohanty SN. An anti-phishing enterprise environ model using feed-forward backpropagation and levenberg-marquardt method. Secur Priv. 2021;4(1):e132. doi:10.1002/spy2.132.
39. Li Y, Xiong K, Li X. Applying machine learning techniques to understand user behaviors when phishing attacks occur. ICST Trans Secur Saf. 2019;6:162809. doi:10.4108/eai.13-7-2018.162809.
40. Vajrobol V, Gupta BB, Gaurav A. Mutual information based logistic regression for phishing url detection. Cyber Secur Appl. 2024;2:100044. doi:10.1016/j.csa.2024.100044.
41. Jain AK, Gupta BB, Kaur K, Bhutani P, Alhalabi W, Almomani A. A content and URL analysisbased efficient approach to detect smishing SMS in intelligent systems. Int J Intell Syst. 2022 Dec;37(12): 11117–41. doi:10.1002/int.23035.
42. Gupta BB, Tewari A, Cvitić I, Peraković D, Chang X. Artificial intelligence empowered emails classifier for internet of things based systems in Industry 4.0. Wirel Netw. 2022;28(1):493–503. doi:10.1007/s11276-021-02619-w.
43. Heiding F, Schneier B, Vishwanath A, Bernstein J, Park PS. Devising and detecting phishing emails using large language models. IEEE Access. 2024;12:42131–46. doi:10.1109/ACCESS.2024.3375882.
44. Hussain M, Cheng C, Xu R, Afzal M. CNN-Fusion: an effective and lightweight phishing detection method based on multi-variant convnet. Inf Sci. 2023;631:328–45. doi:10.1016/j.ins.2023.02.039.
45. Shirazi H, Haefner K, Ray I. Improving auto-detection of phishing websites using fresh-phish framework. Int J Multimed Data Eng Manag. 2022;9:1–14. doi:10.4018/IJMDEM.20180101.
46. Stevanović N. Character and word embeddings for phishing email detection. Comput Inform. 2022;41:1337–57. doi:10.31577/cai_2022_5_1337.
47. Fang Y, Zhang C, Huang C, Liu L, Yue Y. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. IEEE Access. 2019;7:56329–40. doi:10.1109/ACCESS.2019.2913705.

48. Xuan C. A framework for vietnamese email phishing detection. Int J Innov Technol Explor Eng. 2019;9(1):2258–64. doi:10.35940/ijitee.A4843.119119.
49. Oladimeji O. Text analysis and machine learning approach to phished email detection. Int J Comput Appl. 2019;182:11–6. doi:10.5120/ijca2019918354.
50. Mughaid A, AlZu'bi S, Hnaif A, Taamneh S, Alnajjar A, Elsoud EA. An intelligent cyber security phishing detection system using deep learning techniques. Cluster Comput. 2022 Dec;25(6):3819–28. doi:10.1007/s10586-022-03604-4.
51. Hema D. A dynamic and combined phishing detection technique. Int J Innov Technol Explor Eng. 2020;9(5):1421–5. doi:10.35940/ijitee.E2819.039520.
52. Xu T, Rajivan P. Determining psycholinguistic features of deception in phishing messages. Inf Comput Secur. 2023;31(2):199–220. doi:10.1108/ICS-11-2021-0185.
53. Hakim ZM, Ebner NC, Oliveira DS, Getz SJ, Levin BE, Lin T, et al. The Phishing Email Suspicion Test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection. Behav Res Methods. 2021 Jun;53:1342–52. doi:10.3758/s13428-020-01495-0.
54. Weaver B, Braly A, Lane D. Training users to identify phishing emails. J Educ Comput Res. 2021;59: 1169–83. doi:10.1177/0735633121992516.
55. Cooper M, Levy Y, Dringus L. Subject matter experts' feedback on a prototype development of an audio, visual, and haptic phishing email alert system. Online J Appl Knowl Manage. 2020;8:107–21. doi:10.36965/OJAKM.2020.8(2)107-121.
56. D'Angelo G, Ficco M, Robustelli A. An association rules-based approach for anomaly detection on can-bus. In: International Conference on Computational Science and Its Applications, 2023; Cham: Springer Nature Switzerland; p. 174–90.
57. Abid YA, Wu J, Xu G, Fu S, Waqas M. Multi-level deep neural network for distributed denial-of-service attack detection and classification in software-defined networking supported internet of things networks. IEEE Internet Things J. 2024;11(14):24715–25. doi:10.1109/JIOT.2024.3376578.
58. Abid YA, Wu J, Farhan M, Ahmad T. ECMT framework for internet of things: an integrative approach employing in-memory attribute examination and sophisticated neural network architectures in conjunction with hybridized machine learning methodologies. IEEE Internet Things J. 2024;11(4):5867–86. doi:10.1109/JIOT.2023.3312152.
59. Wu J, Guo S, Huang H, Liu W, Xiang Y. Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives. IEEE Commun Surv Tutor. 2018;20(3): 2389–406. doi:10.1109/COMST.2018.2812301.
60. Wu J, Guo S, Li J, Zeng D. Big data meet green challenges: greening big data. IEEE Syst J. 2016;10(3): 873–87. doi:10.1109/JSYST.2016.2550538.
61. Sarivougioukas J, Vagelatos A. Fused contextual data with threading technology to accelerate processing in home ubihealth. Int J Softw Sci Comput Intell. 2022;14(1):1–14. doi:10.4018/IJSSCI.285590.
62. Chopra M, Singh SK, Sharma A, Gill SS. A comparative study of generative adversarial networks for text-to-image synthesis. Int J Softw Sci Comput Intell. 2022;14(1):1–12. doi:10.4018/IJSSCI.300364.
63. Liu RW, Guo Y, Lu Y, Chui KT, Gupta BB. Deep network-enabled haze visibility enhancement for visual IoT-driven intelligent transportation systems. IEEE Trans Ind Inform. 2022 Apr 27;19(2):1581–91. doi:10.1109/TII.2022.3170594.
64. Nguyen GN, Le Viet NH, Elhoseny M, Shankar K, Gupta BB, Abd El-Latif AA. Secure blockchain enabled Cyberphysical systems in healthcare using deep belief network with ResNet model. J Parallel Distr Comput. 2021 Jul 1;153:150–60. doi:10.1016/j.jpdc.2021.03.011.
65. Gupta BB, Yadav K, Razzak I, Psannis K, Castiglione A, Chang X. A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment. Comput Commun. 2021 Jul 1; 175(3):47–57. doi:10.1016/j.comcom.2021.04.023.

66. Jain AK, Gupta B. A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterp Inform Syst. 2022;16(4):527–65. doi:10.1080/17517575.2021.1896786.
67. Zhang Q, Guo Z, Zhu Y, Vijayakumar P, Castiglione A, Gupta BB. A deep learning-based fast fake news detection model for cyber-physical social services. Pattern Recognit Lett. 2023 Apr 1;168(4):31–8. doi:10.1016/j.patrec.2023.02.026.
68. Anirudh S, Nishant PR, Baitha S, Kumar KD. An ensemble classification model for phishing mail detection. Procedia Comput Sci. 2024;233(4):970–8. doi:10.1016/j.procs.2024.03.286.