Gupta, B.B., Gaurav, A.., & Chui, K..T. (2023). Convolution neural network (CNN) based phishing attack detection model for e-business in enterprise information systems. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 23* (pp. 719-725). ICEB'23, Chiayi, Taiwan, October 19-23, 2023

Convolution Neural Network (CNN) Based Phishing Attack Detection Model for E-Business in Enterprise Information Systems

Brij B. Gupta [1,*]
Akshat Gaurav [2]
Kwok Tai Chui [3]

*Corresponding author
[1] Department of Computer Science and Information Engineering, Asia University, Taiwan, & Lebanese American University, Beirut, 1102, Lebanon, bbgupta@asia.edu.tw
[2] Ronin Institute, Montclair, NJ, USA, akshat.gaurav@ronininstitute.org
[3] Hong Kong Metropolitan University (HKMU), Hong Kong, jktchui@hkmu.edu.hk

ABSTRACT

The continued prevalence of phishing attacks highlights the need of research into the creation of reliable detection models for this pervasive online danger to e-business. Using a dataset procured from Kaggle, this study proposes a Convolutional Neural Network (CNN)-based method for identifying phishing scams. Using two Conv1D layers, our model successfully distinguishes between safe and harmful websites. Training results were quite encouraging, with a loss of just 0.077525 and an accuracy of 0.972125 throughout the process. These findings validate our CNN-based phishing attack detection model's sturdiness and adaptability. Our results not only provide a useful tool for spotting phishing attacks but also shed light on the possibilities of CNNs, and in particular Conv1D layers, in the realm of cybersecurity. This study is an important contribution to the ongoing effort to counter the rising danger of phishing attempts and improve the safety of e-business users worldwide.

Keywords: CNN, e-business, electronic commerce, phishing.

INTRODUCTION

The rapid expansion of online enterprises has brought us a new age of unmatched ease, accessibility, and trade. Unfortunately, with this digital development has come a strong foe: phishing attacks (Almomani et al., 2022). Individuals and businesses alike face a serious risk while doing business online because of phishing, the fraudulent effort to gain sensitive information (Gupta & Jain, 2020; Meddah & Guerroudji, 2022; Onyebuchi et al., 2022; Singh et al., 2022). Victims of these attacks often suffer financial ruin and have their faith in digital services eroded as a result. The need of creating reliable and effective systems for identifying phishing attempts has grown in response to this growing threat. This research introduces a novel approach to this problem by introducing a Phishing Attack Detection Model built specifically for the context of online businesses using Convolutional Neural Networks (CNNs). The use of CNNs(Almomani et al., 2022; Chopra et al., 2022; Hasib et al., 2021; Kadri et al., 2022; Li et al., 2022; Tembhurne et al., 2022) in cybersecurity, and more especially phishing detection, is an intriguing new direction, and one that we investigate. Investigating the performance of this CNN-based model, trained and tested using Kaggle data, is central to this study. This model's design includes Conv1D layers, which show promise in capturing complex patterns within the dataset and improving the model's ability to identify both established and novel phishing attack routes.

Our research delves further into the confusion matrix and other metrics for measuring the model's efficacy throughout both training and testing phases. Together, these indicators show that the model can accurately distinguish between benign and harmful "Phishing" efforts and "Normal" E-business activity. Protecting E-commerce and the customers that use it is of utmost importance as we go forward in the world of cyber security. This research not only fills that gap, but it also adds to the ongoing discussion on how to best put cutting-edge deep learning methods like CNNs to use in the service of bettering online safety. Our studies are directed at strengthening the underpinnings of E-business by illuminating the complex relationship between technology and security in this field, thereby guaranteeing the continued primacy of trust and safety in the digital era.

RELATED WORK

Researchers concluded that phishing attacks are a significant threat to e-business(Almomani et al., 2022; Gaurav et al., 2022; Gupta & Jain, 2020; Mishra et al., 2018). Banday et a. (Banday & Qadri, 2011) provide an overview of various phishing approaches, including vishing, spear phishing, pharming, keyloggers, malware, and web Trojans. Dadkhah et al. (Dadkhah et al., 2016) discuss different types of phishing attacks and possible detection techniques for them. Jain et al. (Jain et al., 2021) explain the details of various techniques used by phishing attacks to acquire sensitive information and provide the advantages

and disadvantages of the various anti-phishing technologies Overall, the papers suggest that phishing attacks are a complex and evolving threat to e-business and that there are various techniques available to detect and prevent them.

Recently researchers suggest that deep learning and machine learning models are effective for detecting phishing attacks. Almousa et al. (Almousa et al., 2022) found that deep learning models, including Long Short-Term Memory-based detection models, Fully Connected Deep Neural Network-based detection models, and convolutional neural network-based detection models, achieved high accuracy for phishing website detection. Aljofey et al. (Aljofey et al., 2020) proposed a fast deep learning-based solution model that uses a character-level convolutional neural network (CNN) for phishing detection based on the URL of the website, achieving high accuracy on benchmark datasets. Aljabri et al. (Aljabri & Mirza, 2022) investigated the effectiveness of applying machine learning models in detecting phishing websites and found that the Random Forest algorithm achieved the highest accuracy for both datasets. Atre et al. (Jha et al., 2022) evaluated deep learning models for three phishing detection methods and found that combining results from individual models can improve the effectiveness of detecting cloud-based phishing attacks. Overall, the papers suggest that deep learning and machine learning models are effective for detecting phishing attacks, and different models can be used depending on the specific context and features of the phishing attacks.

PROPOSED APPROACH

The model architecture you've provided is a Convolutional Neural Network (CNN) designed for the task of phishing attack detection in E-business. Let's break down the architecture:

- Two Conv1D (1-dimensional convolutional) layers are the foundation of the model. With 32 output channels and a kernel size of 3, the first Conv1D layer filters the input data using 32 separate filters of size 3. Similar to the first Conv1D layer, but with 128 more output channels, is the second. These layers are critical for detecting elements related to phishing assaults because they capture local trends in the input data.
- Rectified Linear Unit (ReLU) activation functions are used after each Conv1D layer. ReLU allows the network to learn intricate connections in the data by introducing non-linearity via the replacement of negative values with zero.
- MaxPooling1D Layers: These are utilised after the ReLU activation. By choosing the highest value inside a 3-by-3-pixel frame, these layers compress the underlying data set. To make the model more computationally efficient, this aids in downsampling the feature maps while keeping the most important information.
- Flattening the output of higher-dimensional layers into a vector space is the job of the Flatten layer. To link the convolutional layers to the fully connected layers, this is required.
- After the convolutional layers are two fully linked (Linear) layers. To help the model pick up on subtle nuances in the interactions between the simplified feature maps, the first Linear layer introduces a dense layer with 128 units. After the first Linear layer, non-linearity is introduced by ReLU activation. To avoid overfitting, a dropout layer is included following the ReLU activation. Providing the ultimate categorization is a second Linear layer with 2 units as the output layer.
- The model comprises 22,562 parameters in total, all of which may be trained to improve accuracy. The weights and biases of the convolutional and fully connected layers are examples of such parameters.
- Data input sizes of 32 (batch size) and 3 (sequence length) are accepted by the model. There will be 32 total results since the batch size is 32 and there are two types of results ("Normal" and "Phishing").

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
DeepLearning                             [32, 2]                   --
├─Conv1d: 1-1                            [32, 32, 3]               1,568
├─ReLU: 1-2                              [32, 32, 3]               --
├─MaxPool1d: 1-3                         [32, 32, 1]               --
├─Conv1d: 1-4                            [32, 128, 3]              4,224
├─ReLU: 1-5                              [32, 128, 3]              --
├─MaxPool1d: 1-6                         [32, 128, 1]              --
├─Flatten: 1-7                           [32, 128]                 --
├─Linear: 1-8                            [32, 128]                 16,512
├─ReLU: 1-9                              [32, 128]                 --
├─Dropout: 1-10                          [32, 128]                 --
├─Linear: 1-11                           [32, 2]                   258
==========================================================================================
Total params: 22,562
Trainable params: 22,562
Non-trainable params: 0
Total mult-adds (M): 1.09
==========================================================================================
Input size (MB): 0.01
Forward/backward pass size (MB): 0.16
Params size (MB): 0.09
Estimated Total Size (MB): 0.25
==========================================================================================
```

Figure 1: Model Architecture

RESULTS AND DISCUSSION

In our research, we used PyTorch to build a Convolutional Neural Network (CNN) for detecting phishing attacks, and we tested it in a CPU setting using a Kaggle notebook. We monitored the model's progress in terms of training loss, training accuracy, test loss, and test accuracy throughout the course of 10 training epochs. The model had a training loss of 0.324944 and a training accuracy of 0.857125 at the beginning of training (epoch 0). The test loss was reduced to 0.176080 on the test dataset, and the test accuracy was improved to 0.938988, both of which are very good results. The model dramatically enhanced itself as training progressed. Training loss was minimised to 0.136915 at the end of period 2 with an accuracy of 0.951250%. Additionally, the test loss was brought down to 0.128674, and the test accuracy was increased to 0.956349. Training accuracy of 0.963875 and test accuracy of 0.964782 were attained in epochs 4 and 5, respectively, demonstrating the model's capacity to generalise its learnings.

As training progressed, the model continued to improve. The best training loss was 0.077525 and the best training accuracy was 0.972125 in epoch 9. At 0.090748, the test loss was at its lowest, while at 0.967262, the test accuracy was at its highest. These results demonstrate the model's efficiency in differentiating phishing assaults from real E-business endeavours. Its reliability in detecting phishing assaults in real-world settings is supported by its strong generalisation, as seen by its performance on the test dataset. In conclusion, our research highlights the promise of CNN-based models in bolstering online security and shielding E-businesses from ever-evolving dangers. It also demonstrates the viability of incorporating such models in CPU systems, opening them up to a wide range of potential uses in the field of cybersecurity.

Accuracy and Loss Curves

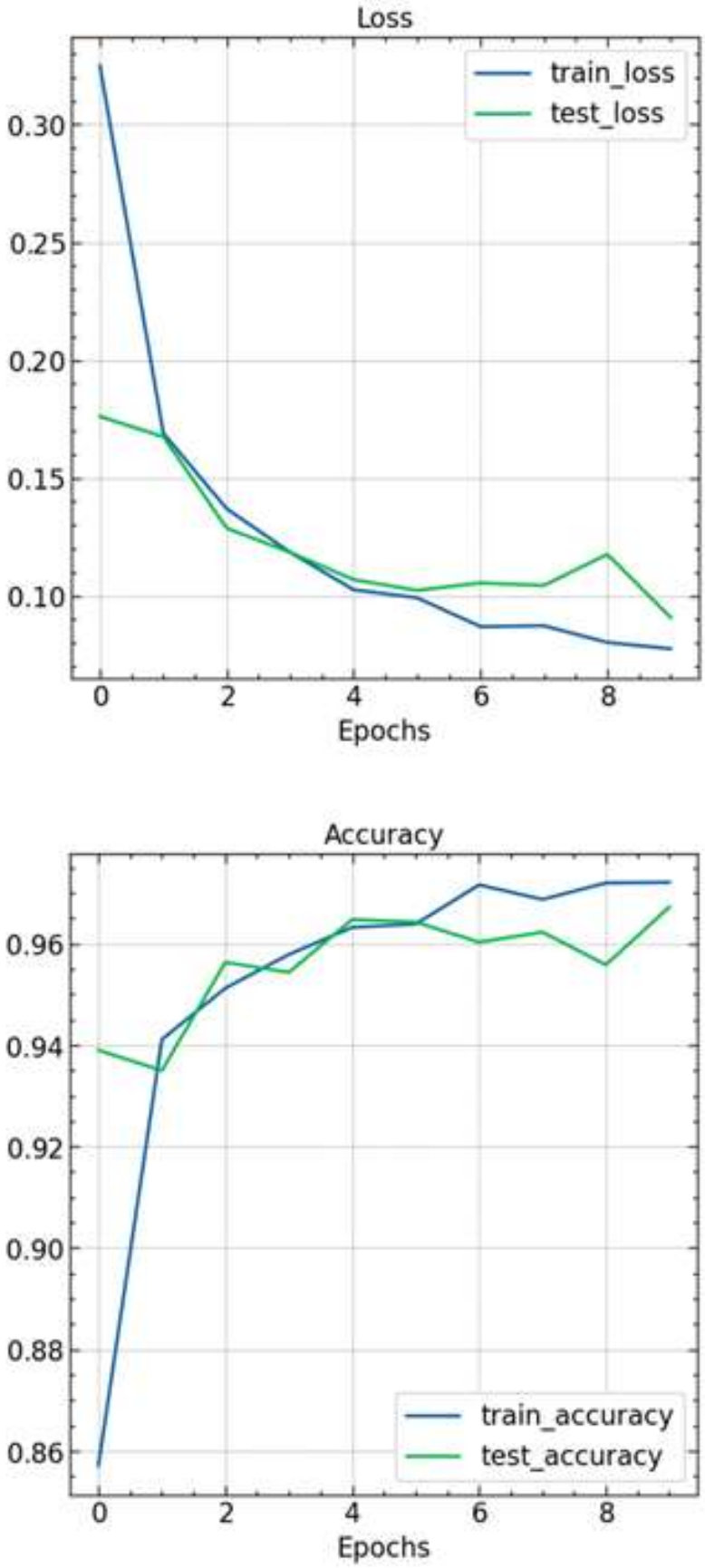

Figure 2: Accuracy and Loss Curve

Classification Report

Our Phishing Attack Detection Model uses convolutional neural networks (CNNs), and the classification report gives a thorough assessment of the model's ability to discriminate between two classes in the dataset: "Normal" and "Phishing." These measures are critical for assessing the model's performance. The "Normal" class has a precision of 0.97, measuring the reliability of the class's positive predictions. When the model classifies an occurrence as "Normal," it is 97% accurate. A recall of 0.96 indicates that 96% of all real-world "Normal" cases are captured by the model, representing the accuracy with which it identifies such instances. For the "Normal" class, the model's F1-score of 0.97 demonstrates a good compromise between accuracy and recall. Altogether, these indicators point to a very high accuracy and dependability in determining "Normal" occurrences. The model is 96% accurate at classifying instances as "Phishing," as shown by the accuracy value of 0.96.

Similarly, the recall for this category is 0.97, meaning that 97% of real-world "Phishing" cases are captured by the model. Again, the model's outstanding performance in identifying "Phishing" cases is shown by the "Phishing" class's F1-score of 0.97. Measured by the percentage of properly categorised occurrences, the model achieves an overall accuracy of 97%. This impressive precision demonstrates the model's prowess in identifying "Normal" from "Phishing" scenarios.

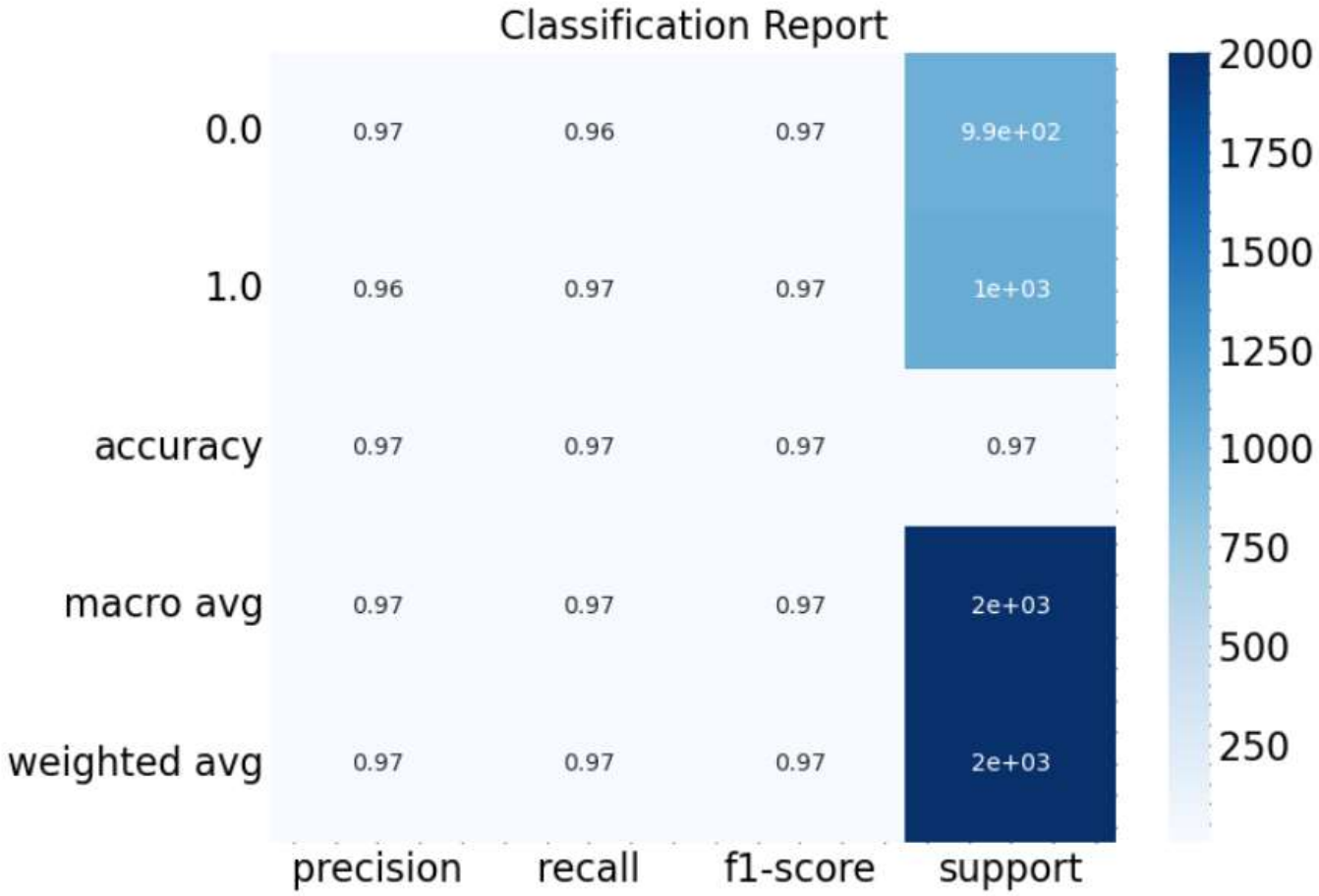

Figure 3: Classification Report

Confusion Matrix

The confusion matrix provides a detailed breakdown of our CNN-based Phishing Attack Detection Model's classification performance for the "Normal" and "Phishing" classes. In this matrix, the two classes are represented as rows and columns: "Normal" and "Phishing."

- True Positives (TP): The model correctly predicted 983 instances as "Phishing."
- True Negatives (TN): It accurately classified 952 instances as "Normal."
- False Positives (FP): The model incorrectly predicted 36 instances as "Phishing" when they were actually "Normal."
- False Negatives (FN): It made 29 incorrect predictions, classifying instances as "Normal" when they were "Phishing."

To break it down further, in the context of our model's performance:

- For "Normal," there were 952 true positives (correctly identified as "Normal") and 36 false positives (incorrectly classified as "Phishing").
- For "Phishing," there were 983 true positives (correctly identified as "Phishing") and 29 false negatives (incorrectly classified as "Normal").

These results show that the model is quite good at distinguishing between "Normal" and "Phishing" cases, with minimal false positives. The large proportion of correct predictions suggests that the model is able to accurately capture "Normal" and "Phishing" scenarios. False positives and false negatives occurred, although at low rates, indicating that the model's performance is highly trustworthy, as is typical in classification tasks. These results corroborate the classification report's claims of high accuracy, recall, and F1-scores, demonstrating the model's ability to accurately discriminate between these two categories crucial to ensuring the safety of online businesses.

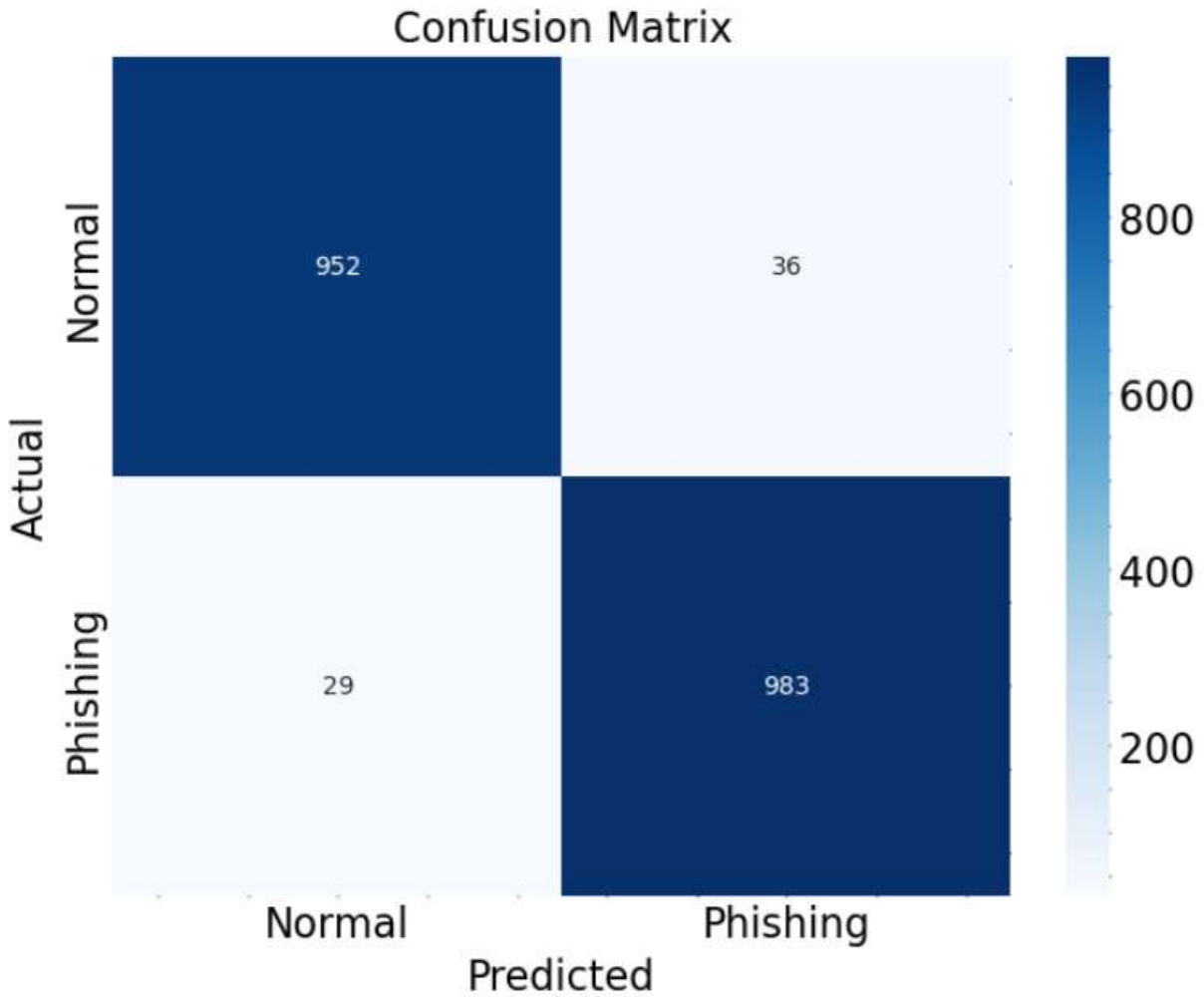

Figure 4: Confusion Matrix

CONCLUSION

In this research, we introduced a Phishing Attack Detection Model built on top of a Convolutional Neural Network (CNN) and optimised for use in online businesses. Our model, developed using a dataset obtained from Kaggle, demonstrates the potential of deep learning in tackling this critical topic, since phishing attempts have become more common in the realm of electronic commerce and need strong and adaptable solutions. Both the training and testing results of our model have been quite

encouraging. Training results show it is quite good at learning and generalising from the data supplied, with a loss of only 0.077525 and an accuracy of 0.972125. Its test loss was 0.090748 and its test accuracy was 0.967262, further demonstrating its superiority in detecting phishing attempts in realistic E-business settings. Our CNN-based model's effectiveness in this setting demonstrates the continued relevance of deep learning methods in the field of cyber security. The use of Conv1D layers also demonstrates their efficiency in collecting complex patterns within the dataset, an important feature for differentiating between benign and harmful cyber-activity. The strategies used by cybercriminals are constantly changing in tandem with the development of e-commerce. Even though phishing attempts are still a major problem, our strategy is a major improvement in terms of protecting online businesses. It has potential uses outside the scope of this investigation and may serve as a springboard for further work in the area of cyber security.

ACKNOWLEDGEMENT

This research work is supported by National Science and Technology Council (NSTC), Taiwan Grant No. NSTC112-2221-E-468-008-MY3.

REFERENCES

Aljabri, M., & Mirza, S. (2022). *Phishing attacks detection using machine learning and deep learning models*. 175–180.

Aljofey, A., Jiang, Q., Qu, Q., Huang, M., & Niyigena, J.-P. (2020). An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics*, *9*(9), 1514. https://doi.org/10.3390/electronics9091514

Almomani, A., et al. (2022). Phishing website detection with semantic features based on machine learning classifiers: A comparative study. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *18*(1), 1–24.

Almousa, M., Zhang, T., Sarrafzadeh, A., & Anwar, M. (2022). Phishing website detection: How effective are deep learning-based models and hyperparameter optimization? *Security and Privacy*, *5*(6), e256.

Banday, M. T., & Qadri, J. A. (2011). Phishing-A growing threat to e-commerce. *arXiv Preprint arXiv:1112.5732*.

Chopra, M., et al.. (2022). Analysis & prognosis of sustainable development goals using big data-based approach during COVID-19 pandemic. *Sustainable Technology and Entrepreneurship*, *1*(2), 100012.

Dadkhah, M., Jazi, M. D., Mobarakeh, M. S., Shamshirband, S., Wang, X., & Raste, S. (2016). An overview of phishing attacks and their detection techniques. *International Journal of Internet Protocol Technology*, *9*(4), 187–195.

Gaurav, A., et al. (2022). A comprehensive survey on DDoS attacks on various intelligent systems and it's defense techniques. *International Journal of Intelligent Systems*, *37*(12), 11407–11431.

Gupta, B. B., & Jain, A. K. (2020). Phishing attack detection using a search engine and heuristics-based technique. *Journal of Information Technology Research (JITR)*, *13*(2), 94–109.

Hasib, K. M., Towhid, N. A., & Islam, M. R. (2021). Hsdlm: A hybrid sampling with deep learning method for imbalanced data classification. *International Journal of Cloud Applications and Computing (IJCAC)*, *11*(4), 1–13.

Jain, A. K., et al. (2021). A survey of phishing attack techniques, defence mechanisms and open research challenges. *Enterprise Information Systems*, 1–39.

Jha, B., Atre, M., & Rao, A. (2022). *Detecting Cloud-Based Phishing Attacks by Combining Deep Learning Models*. 130–139.

Kadri, O., Benyahia, A., & Abdelhadi, A. (2022). Tifinagh Handwriting Character Recognition Using a CNN Provided as a Web Service. *International Journal of Cloud Applications and Computing (IJCAC)*, *12*(1), 1–17.

Li, S., Qin, D., Wu, X., Li, J., Li, B., & Han, W. (2022). False alert detection based on deep learning and machine learning. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *18*(1), 1–21.

Meddah, I. H., & Guerroudji, F. (2022). Business process discovery using process mining techniques and distributed framework. *International Journal of Cloud Applications and Computing (IJCAC)*, *12*(1), 1–17.

Mishra, A., et al. (2018). Intelligent phishing detection system using similarity matching algorithms. *International Journal of Information and Communication Technology*, *12*(1–2), 51–73.

Onyebuchi, A., Matthew, U. O., Kazaure, J. S., Okafor, N. U., Okey, O. D., Okochi, P. I., Taiwo, J. F., & Matthew, A. O. (2022). Business demand for a cloud enterprise data warehouse in electronic healthcare computing: Issues and developments in e-healthcare cloud computing. *International Journal of Cloud Applications and Computing (IJCAC)*, *12*(1), 1–22.

Singh, A., et al. (2022). Distributed denial-of-service (DDoS) attacks and defense mechanisms in various web-enabled computing platforms: Issues, challenges, and future research directions. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *18*(1), 1–43.

Tembhurne, J. V., Almin, M. M., & Diwan, T. (2022). Mc-DNN: Fake news detection using multi-channel deep neural networks. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *18*(1), 1–20.