Science World Journal Vol. 20(No 2) 2025 https://dx.doi.org/10.4314/swj.v20i2.33
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

DETECTING PHISHING WEBSITES USING LARGE LANGUAGE MODEL

*Ochu J.A., Aimufua G.I.O., Musa H., Chaku S.E.

Centre for Cyberspace Studies, Nasarawa State University, Keffi, Nasarawa State, Nigeria

*Corresponding Author Email Address: jeromeochu@gmail.com

ABSTRACT
Phishing detection is a critical area in cybersecurity that significantly impacts the protection of sensitive information and the overall security posture of individuals and organizations. The increasing sophistication of phishing attacks presents substantial challenges in identifying fraudulent websites that impersonate legitimate entities. Existing detection methods often struggle with high false positive rates and misclassification errors, highlighting the need for more effective solutions. In response to these challenges, this study developed and evaluated a Multi-Layer Perceptron (MLP) model specifically designed for phishing website detection. The research utilized a comprehensive dataset containing features extracted from both legitimate and phishing websites, combining textual and numerical attributes to enhance classification performance. The MLP model was rigorously assessed using metrics such as accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). Results indicate an overall accuracy of 96.6%, a precision of 96.5%, and a recall of 97.5%, along with an AUC of 0.9941. These findings showcase the model's strong discriminatory power and effectiveness in minimizing misclassifications. The research highlights a significant advancement in phishing detection capabilities compared to existing approaches, laying the groundwork for future developments in phishing detection systems. The study emphasizes the potential for real-world applications in enhancing cybersecurity defenses against evolving threats.

Keywords: Phishing, Neural Network, Artificial Intelligence, Machine Learning, Large Language Models.

INTRODUCTION

With the increasing reliance on Information and Communication Technology (ICT) in various aspects of life, cybercrime has emerged as a significant issue in the fields of cybersecurity, network security, criminology and criminal law.

As the internet continues to become an integral part of people's daily lives, the need for enhanced security measures has become increasingly important. Phishing is a prevalent and dangerous form of cyberattack that targets individuals and organizations for the purpose of gaining sensitive information. The financial losses and security risks associated with phishing make it imperative to find effective solutions for detecting and preventing these attacks (Aditya, 2023).

Peter (2024) indicated that phishing attacks become more advanced, incorporating Artificial Intelligence (AI)-powered tactics to create realistic and highly targeted emails, the need for awareness and education around phishing has grown increasingly pressing. The ability of cybercriminals to create believable fake profiles, websites, and links poses a significant threat to individuals and organizations alike, making it challenging to distinguish between genuine and malicious emails.

The development of phishing detection models utilizing machine learning techniques, particularly deep learning algorithms, has significantly improved online security. These models analyze various features of phishing websites and emails to identify malicious links. Despite the progress made in this area, the evolving nature of phishing attacks continues to present challenges as threat actors employ increasingly sophisticated tactics. Moreover, the vast amount of data generated by email and online communications requires models that can effectively process and analyze this data in real-time to prevent and mitigate the impact of phishing attacks (Lindah, and David, 2024).

The evolving sophistication of phishing tactics poses significant challenges in developing effective detection models, as these models often suffer from high rates of false positives and poor detection accuracy when encountering new phishing strategies. This emphasizes the crucial role of accurate labeling of training data in developing robust phishing detection models (Kaklulapati, 2024).

The advancements in transformer architectures, computational power, and the availability of large-scale training datasets have led to the development of powerful Large Language Models (LLMs) that can perform various language-related tasks with human-like proficiency. LLMs have emerged as a revolutionary technology, demonstrating remarkable skills in processing, generating, and understanding language (Humza, *et. al*., 2024).

Liming (2024) defines LLMs as powerful AI models trained on vast amounts of text data, enabling them to perform various language-related tasks in a manner that closely resembles human communication. These models have shown remarkable capabilities in a wide range of natural language processing tasks, including text generation, text classification, language translation, and question answering.

The modern world has been transformed by the internet, making it an essential part of daily life for many people, its usefulness is balanced by an escalating risk of sophisticated cyber threats. Traditional methods of phishing detection, such as web crawling, are proving insufficient against the constantly morphing tactics used by attackers, leaving the online landscape vulnerable to these advanced threats (Nirmala, *et. al*., 2023).

Given the aforementioned limitations of current phishing detection methods, continued research is crucial to develop algorithms that are both adaptive and scalable. Such algorithms must be agile enough to counter ever-changing tactics employed by attackers, yet also capable of processing the immense volume of data generated in the digital palace.

In response to these challenges, this study developed and evaluated a Multi-Layer Perceptron (MLP) model specifically designed for phishing website detection and the study emphasizes the potential for real-world applications in enhancing cybersecurity

Science World Journal Vol. 20(No 2) 2025 https://dx.doi.org/10.4314/swj.v20i2.33
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

defenses against evolving threats.

Numerous studies have explored the use of machine learning, particularly Multi-Layer Perceptron (MLP), for phishing detection. For instance, Thotwe and Mane (2024) enhanced the traditional MLP by introducing stochastic gradient learning to improve model speed and learning efficiency, demonstrating the benefits of optimization in MLP architecture. Similarly, Nirmala et al. (2023) implemented an MLP-based model that achieved 93.28% accuracy in detecting redirection spam but acknowledged limitations in generalizability and adaptability to dynamic phishing methods.

Mote et al. (2023) utilized feature selection and keyword extraction for email-based phishing detection, with MLP achieving a performance of 90.2% accuracy. However, their study focused primarily on static datasets and did not benchmark the model against alternative learning algorithms. On another front, Peter et al. (2024) explored the risks of generative AI in phishing attacks and proposed a framework combining AI-assisted tools and user training. Although promising, their approach depended heavily on evolving AI platforms like GPT, lacking direct comparison with neural models like MLP.

Additionally, Arun (2023) proposed an ensemble logistic regression approach that outperformed other classifiers but required high computational time, making it less scalable. These studies collectively highlight that while MLP has been widely applied in phishing detection, many implementations either lack efficiency in training (due to reliance on backpropagation), or fail to compare performance with optimized alternatives like stochastic gradient descent (SGD).

This study addresses these gaps by applying and evaluating an SGD-based MLP for phishing detection, comparing it directly with the conventional backpropagation MLP. The aim is to improve model adaptability, reduce training overhead, and enhance classification accuracy across multiple performance metrics.

MATERIALS AND METHODS

Framing Research Questions

Crafting well-defined research questions is a challenging task, particularly in the realm of cybersecurity, where the scope of phishing attacks can be diverse and complex. To ensure a thorough examination of the literature on phishing detection, the questions were framed to identify research articles that effectively employ ML and LLMs. This study is aimed at answering the following questions:

i. How effective is MLP model in analyzing website content to identify phishing attempts, given the evolving nature of phishing attacks?
ii. How can MLP model be trained to detect phishing websites?
iii. What is the performance of the MLP model in detecting unseen phishing websites?

Research Design

This study adopts a research methodology known as Design Science Research (DSR) employed in the field of information models and computer research to solve problems via a problem-solving framework. DSR is a problem-solving paradigm that seeks to enhance human knowledge via the creation of innovative artifacts and the generation of design knowledge (DK) via innovative solutions to real-world problems.

Using a quantitative approach, specifically an experimental design, to evaluate the effectiveness of LLMs in identifying phishing websites. An experimental design allows for controlled manipulation of variables (i.e., website content) to observe the impact on the dependent variable (i.e., LLM classification of phishing attempt). Within the experimental design, a supervised learning paradigm was utilized. Here, a pre-trained LLM was further trained on a labelled dataset consisting of legitimate and phishing websites.

Following training, the LLM's performance in classifying unseen phishing websites was evaluated. A web base application was developed to enable technical and non-technical individuals to detect phishing website just by providing the URL to the website. The web base application was developed using python programming language and MLP was the model used in this study to train and finetune the collected datasets. The essence of the LLM which is a subset of generative AI is to help in analysing the content of the website and translate the analysed content into useful information which enable the system to know if the website is legitimate or it is a phishing website.

The research design framework is shown in figure 1 and figure 2 respectively.

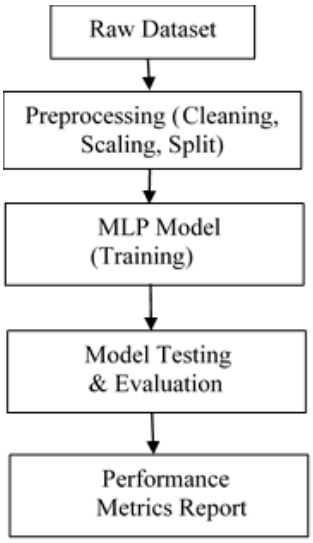

Figure 1: Training and Testing Phases

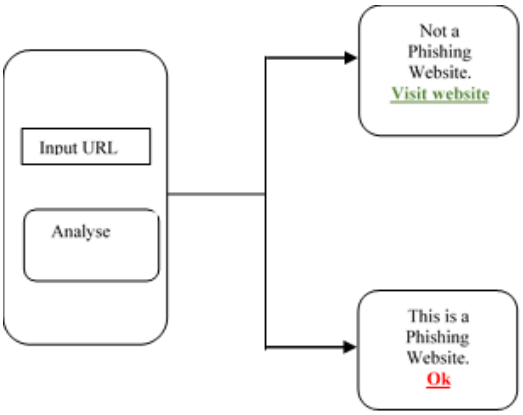

Figure 2: Web Application Workflow for the Phishing Website Detection using LLM

Science World Journal Vol. 20(No 2) 2025 https://dx.doi.org/10.4314/swj.v20i2.33
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Data Collection

Stratified random sampling was employed to ensure the sample reflects the proportions of legitimate and phishing websites within the broader internet population. To collect phishing sites, the researcher used OpenPhish and PhishTank, which are phishing intelligence sources. Publicly available datasets containing labeled phishing and legitimate websites, obtained from OpenPhish (https://openphish.com/) or PhishTank (https://phishtank.org/), served as the sampling frame. The target sample size for this research was 1255 phishing websites and 956 non-phishing websites accessible on the internet.

Data Preprocessing

i. **Data Collection:** This research gathered comprehensive dataset for both legitimate and phishing websites. This involved using reputable source and employing web crawler.
ii. **Feature Extraction:** Relevant features essential for model training was extracted from the collected dataset. Features includes URL characteristics, website content, HTML attributes, and metadata.
iii. **Data Cleaning:** Noises, duplicates, and irrelevant data was removed so as to enhance the quality of the dataset. Techniques such as outlier detection, text normalization, and spell checking may be employed.

Model Performance Evaluation

Evaluating the performance of a phishing website detection model is crucial for cybersecurity. Four metrics, including accuracy, precision, recall and F1-score, was used to assess the large language model's performance. These metrics are obtained from the confusion matrices, as presented in Table 1, a confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. It is essential to evaluate the stability of every machine learning algorithm. Cross-validation method was used for evaluating the effectiveness of the LLM model by using a subset of the input data as training and a portion of the input data as testing that has never been used before (Saleem, *et. al*., 2023).

Table 1: Confusion Matrix

		Predicted	
		Positive (1) Malicious	Negative (0) Benign
Actual	Positive (1) Malicious	TP	FN
	Negative (0) Benign	FP	TN

The performance of classification models is typically evaluated using the following metrics:

i. **Accuracy:** Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of predictions made.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+FP+TN}} \quad \textbf{(1)}$$

ii. **Precision:** Precision indicates the proportion of positive identifications that were actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \textbf{(2)}$$

iii. **Recall:** Recall measures the proportion of actual positives that were correctly identified.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \textbf{(3)}$$

iv. **F1-Score:** The F1-score is the harmonic mean of precision and recall, providing a balance between the two.

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Pr} \times \text{Rc}}{\text{Pr} + \text{Rc}} \quad \textbf{(4)}$$

Where:

- **TP**: True Positives
- **TN**: True Negatives
- **FP**: False Positives
- **FN**: False Negatives

ESULTS AND DISCUSSION

Data Presentation

The following Python libraries, namely Numpy, Pandas, Matplotlib, and Seaborn, were loaded into Jupyter Notebook. Numpy allows efficient vectorized computation and broadcasting across multi-dimensional arrays. Pandas is a sophisticated open-source program that aids in data analysis and manipulation, featuring a user-friendly interface and developed using the Python programming language. Matplotlib and Seaborn are Python libraries designed for data visualization. They provide a user-friendly interface for producing visually appealing and practical graphs. Seaborn is based on Matplotlib but has less functionality.

Dataset Description

Figure 3 provides a statistical summary of a dataset containing 11055 rows and 32 columns, all of which include numerical values. The dataset consists of statistical metrics, including the mean, standard deviation, minimum, maximum, and quartiles. These measurements provide valuable insights into the average values and variability of phishing indicators, making it easier to analyze and identify any abnormal data points.

	index	having_IPhaving_IP_Address	URLURL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting
count	11055.000000	11055.000000	11055.000000	11055.000000	11055.000000	11055.000000
mean	5528.000000	0.313795	-0.633198	0.738761	0.700588	0.741474
std	3191.447947	0.949534	0.766095	0.673998	0.713598	0.671011
min	1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000
25%	2764.500000	-1.000000	-1.000000	1.000000	1.000000	1.000000
50%	5528.000000	1.000000	-1.000000	1.000000	1.000000	1.000000
75%	8291.500000	1.000000	-1.000000	1.000000	1.000000	1.000000
max	11055.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Figure. 3: Dataset Description

Dataset Information

The dataset was loaded into the Pandas data frame for easy

Science World Journal Vol. 20(No 2) 2025 https://dx.doi.org/10.4314/swj.v20i2.33
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

analysis, model development, and prediction. Figure 4 depicts the Pandas Data Frame comprising 11055 rows and 32 columns representing phishing parameters.

```
RangeIndex: 11055 entries, 0 to 11054
Data columns (total 32 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   index                        11055 non-null  int64
 1   having_IPhaving_IP_Address   11055 non-null  int64
 2   URLURL_Length                11055 non-null  int64
 3   Shortining_Service           11055 non-null  int64
 4   having_At_Symbol             11055 non-null  int64
 5   double_slash_redirecting     11055 non-null  int64
 6   Prefix_Suffix                11055 non-null  int64
 7   having_Sub_Domain            11055 non-null  int64
 8   SSLfinal_State               11055 non-null  int64
 9   Domain_registeration_length  11055 non-null  int64
 10  Favicon                      11055 non-null  int64
 11  port                         11055 non-null  int64
 12  HTTPS_token                  11055 non-null  int64
 13  Request_URL                  11055 non-null  int64
 14  URL_of_Anchor                11055 non-null  int64
 15  Links_in_tags                11055 non-null  int64
 16  SFH                          11055 non-null  int64
 17  Submitting_to_email          11055 non-null  int64
 18  Abnormal_URL                 11055 non-null  int64
 19  Redirect                     11055 non-null  int64
 20  on_mouseover                 11055 non-null  int64
 21  RightClick                   11055 non-null  int64
 22  popUpWidnow                  11055 non-null  int64
 23  Iframe                       11055 non-null  int64
 24  age_of_domain                11055 non-null  int64
 25  DNSRecord                    11055 non-null  int64
 26  web_traffic                  11055 non-null  int64
 27  Page_Rank                    11055 non-null  int64
 28  Google_Index                 11055 non-null  int64
 29  Links_pointing_to_page       11055 non-null  int64
 30  Statistical_report           11055 non-null  int64
 31  Result                       11055 non-null  int64
```

Fig. 4: Dataset Information

Training the Model

Figure 5 is a log representing the training progress of a Multi-Layer Perceptron (MLP) with 500 epochs. The model shows consistent improvement in accuracy and other metrics over the epochs, suggesting successful training.

```
Epoch 50: Training Accuracy: 0.9700, Loss: 0.1199
Epoch 100: Training Accuracy: 0.9700, Loss: 0.1199
Epoch 150: Training Accuracy: 0.9833, Loss: 0.0669
Epoch 200: Training Accuracy: 0.9880, Loss: 0.0479
Epoch 250: Training Accuracy: 0.9877, Loss: 0.0493
Epoch 300: Training Accuracy: 0.9793, Loss: 0.0828
Epoch 350: Training Accuracy: 0.9898, Loss: 0.0407
Epoch 400: Training Accuracy: 0.9887, Loss: 0.0452
Epoch 450: Training Accuracy: 0.9897, Loss: 0.0412
Epoch 500: Training Accuracy: 0.9900, Loss: 0.0398
```

Fig. 5: Modelling the Multi-Layer Perceptron (MLP) Classifier

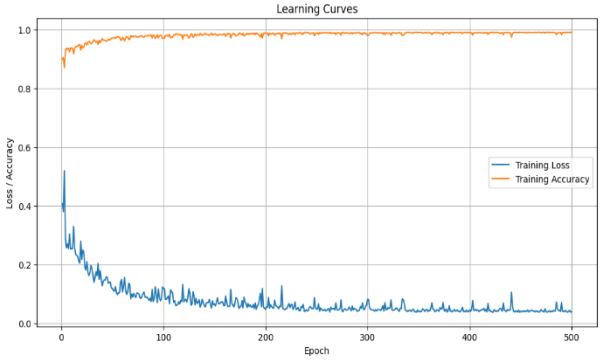

Figure 6: Learning Curve of Multi-Layer Perceptron (MLP) Model

The learning curve depicted in Figure 7 illustrates the performance of the Multi-Layer Perceptron (MLP) model during training, displaying both the training loss and training accuracy over 500 epochs.

From figure 6 above, the blue line represents the training loss, which decreases significantly in the early stages, indicating that the model is learning and minimizing errors. Over time, the loss flattens out, showing minimal fluctuations, meaning the model has converged and is no longer improving much in terms of reducing error. The orange line represents the training accuracy, which starts high and continues to improve rapidly before reaching a plateau close to 1.0 (or 100%) accuracy. This indicates that the model is consistently making correct predictions on the training data, achieving near-perfect accuracy. Additionally, the curves suggest that the model is learning effectively, with the loss decreasing and the accuracy increasing. However, the plateauing of both curves could also indicate that further training might not yield significant improvements, and the model may be approaching its optimal performance on the training set.

Confusion matrix of the Multi-Layer Perceptron (MLP) Model

The confusion matrix for the Multi-Layer Perceptron (MLP) model depicted in Figure 7 in phishing website detection indicates strong classification performance. Out of 956 actual legitimate websites, the model correctly classified 911 as legitimate (true negatives), but misclassified 45 as phishing (false positives). For the 1255 actual phishing websites, it correctly identified 1224 as phishing (true positives) while misclassifying 31 as legitimate (false negatives).

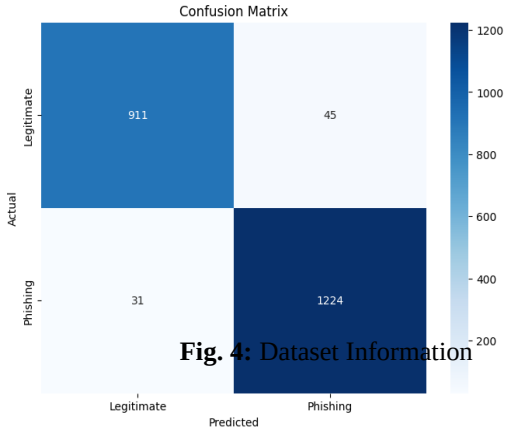

Figure 7: Confusion Matrix of the Multi-Layer Perceptron (MLP) Model

This results in an overall accuracy of approximately 96.6%, meaning that the model correctly predicts the class of websites most of the time. The precision for phishing detection is 96.5%, meaning that 96.5% of the websites predicted as phishing are actually phishing. The recall for phishing detection is 97.5%, reflecting that the model correctly identifies 97.5% of actual phishing websites. These high values indicate the model is highly effective in minimizing both false positives (legitimate sites wrongly flagged) and false negatives (phishing sites that are missed). Despite a few classification errors, the model demonstrates reliable phishing detection capabilities.

Science World Journal Vol. 20(No 2) 2025 https://dx.doi.org/10.4314/swj.v20i2.33
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Discussion of Findings

Table 2 illustrates the evaluation metrics demonstrating the performance of the developed Multi-Layer Perceptron (MLP) Model. The assessment process utilized four key metrics: accuracy, precision, F1 score, and recall.

Table 2: Classification Report

Class	Precision	Recall	F1-Score	Support
-1 (Legitimate)	0.97	0.95	0.96	956
1 (Phishing)	0.96	0.98	0.97	1255
Accuracy			0.97	2211
Macro Avg	0.97	0.96	0.96	2211
Weighted Avg	0.97	0.97	0.97	2211

The classification report for the Multi-Layer Perceptron (MLP) model using Stochastic Gradient Descent (SGD) provides strong evidence of the model's effectiveness in detecting phishing websites. The high precision score of 0.97 for legitimate websites indicates that the model correctly classified the vast majority of safe websites, with only a 3% false positive rate. This is critical in minimizing unnecessary alarms for users. Likewise, the 0.96 precision for phishing websites shows the model's ability to correctly identify most malicious websites without mistakenly flagging legitimate ones.

The recall of 0.98 for phishing websites demonstrates that the model successfully identifies nearly all phishing threats, an essential requirement in cybersecurity, where overlooking a phishing attempt can have serious consequences. The 0.95 recall for legitimate websites shows only a small proportion of legitimate sites are misclassified. These balanced scores across classes are particularly noteworthy, given the typically imbalanced nature of phishing datasets, which the SGD optimizer handled well during training.

The F1-scores of 0.96 and 0.97 confirm the model's robustness and reliability across both classes by balancing precision and recall, while the overall accuracy of 97% reflects the model's capacity to generalize effectively across unseen data.

These strong results can be attributed to the choice of Stochastic Gradient Descent (SGD) as the optimization algorithm, which allows for faster convergence and better generalization in comparison to traditional Backpropagation. By updating weights incrementally on smaller batches, the model adapts more dynamically to patterns in the dataset, leading to more accurate classifications.

Justification of Results

To ensure the reliability of the proposed model's performance, a benchmarking analysis was conducted against a baseline model — a Backpropagation-based MLP — previously used in similar studies.

Figure 8 illustrates the comparison between the results gotten from Backpropagation MLP (Previous Model) and SGD-based MLP (Proposed Model).

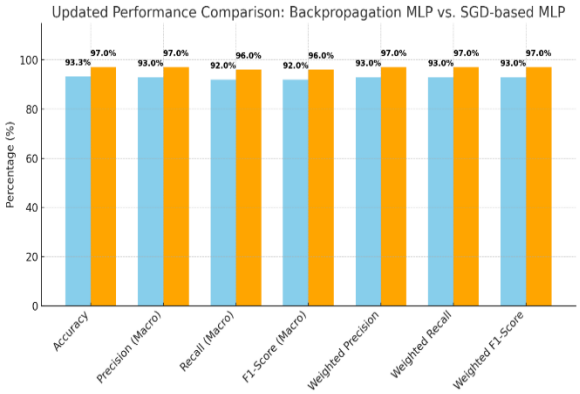

Figure 8: Performance Comparison Chart

In the previous work, the Backpropagation-based Multi-Layer Perceptron (MLP) achieved a maximum accuracy of 93.3%, largely attributed to the optimization of weights throughout training. This model exhibited strong fault tolerance, and its performance improved with backpropagation. However, its training efficiency was dependent on the learning rate—higher learning rates could reduce training time but often led to a decrease in accuracy.

In comparison, the Stochastic Gradient Descent (SGD)-based MLP model developed in this study demonstrates a significant improvement in performance. It achieved a superior accuracy of 97%, with macro averages of 0.97 for precision, 0.96 for recall, and 0.96 for F1-score, indicating balanced performance across both classes regardless of class imbalance. Furthermore, the weighted average for all three metrics remained consistently high at 0.97, reflecting the model's ability to handle imbalanced data effectively. This consistency across all evaluation metrics underscores the robustness of the proposed model in accurately identifying phishing websites while minimizing false positives for legitimate site.

Therefore, the previous works such as Thotwe and Mane (2024) and Nirmala *et al.* (2023), which used Backpropagation MLP with average accuracies below 94%, the use of SGD-MLP in this study represents a notable improvement in performance and adaptability with superior accuracy of 97%. This directly supports the aim of the study: to improve phishing detection accuracy through an optimized MLP framework.

Conclusion

This study successfully demonstrates the effectiveness of a Multi-Layer Perceptron (MLP) model for detecting phishing websites, a critical challenge in the evolving landscape of cybersecurity. By addressing the limitations of existing detection methods, such as high false positive rates and misclassification errors, the MLP model achieved an overall accuracy of 96.6%, with a precision of 96.5% and a recall of 97.5%, alongside an impressive AUC of 0.9941. These results underscore the model's robust ability to differentiate between legitimate and phishing websites, effectively minimizing misclassifications and enhancing detection reliability.

The comprehensive evaluation of the model, utilizing diverse features from both legitimate and phishing websites, contributes significantly to the field of phishing detection. This research not only highlights the potential of machine learning techniques, particularly

neural networks, in improving cybersecurity measures but also sets a foundation for future investigations into more advanced detection systems.

Ultimately, the findings of this study pave the way for real-world applications aimed at strengthening cybersecurity defenses and protecting users from the growing threat of phishing.

This study is subject to several limitations that should be acknowledged. First, the effectiveness of the proposed model is highly dependent on the quality, diversity, and recency of the dataset. If the dataset lacks representation of evolving phishing techniques or includes outdated attack patterns, it may limit the model's ability to generalize to emerging threats. Although the Multilayer Perceptron (MLP) model demonstrated strong classification performance, it may face challenges when confronting more sophisticated and targeted phishing strategies, such as spear phishing. Additionally, since this paper is based on a subset of a broader study, its scope may not capture the full complexity of real-world phishing scenarios.

REFERENCES

Aditya, S., & Jyoti, T. (2023). *Phishing website detection using ensemble learning*. Computer Science & Engineering Department, SGSITS Indore, India.

Arun, D. K. (2023). *Convolution neural networks for phishing detection* [Master's thesis, The University of Texas at Tyler]. The University of Texas at Tyler, Tyler, TX, USA.

Godwin, O., & Ayuns, L. (2024). *Artificial intelligence and machine learning in phishing detection*. https://www.researchgate.net/publication/378233860.

Humza, N., Asad, U. K., Shi, Q., Muhammad, S., Saeed, A., Muhammad, U., & Ajmal, M. (2024). *A comprehensive overview of large language models*. University of Engineering and Technology (UET), Lahore, Pakistan.

Kaklulapati, V. (2024). *URL-based Android application for identification of website phishing* [Undergraduate project, Institute of Science and Technology]. Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telangana, India.

Liming, J. (2024). *Detecting scams using large language models* [Master's thesis, Harbin University of Science and Technology]. Harbin University of Science and Technology, Harbin, Heilongjiang, China.

Lindah, S., & David, N. (2024). *Sentence level analysis model for phishing detection using KNN*. Faculty of Computing and Informatics, Mount Kenya University, Thika, Kenya.

Mote, A. V., Hitesh, Y. P., Prachi, R. P., & Omkar, S. O. (2023). *Phishing website detection based on machine learning algorithm*. Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India.

Nirmala, M., Lakshmi, P., Meghana, B., Gopi, C., & Vidyadhari, A. (2023). A novel approach on detecting phishing attacks on URLs using machine learning techniques. *International Journal of Scientific Research in Science and Technology*.

Peter, K. K., Aloysius, Z. Y., & Vivek, B. (2024). *Towards a hybrid security framework for phishing awareness education and defense*. Singapore Institute of Technology, Singapore.

Saleem, R. A., Sundarvadivazhagan, B., Amna, S. A., Bhisham, S., Subrata, C., Abolfazl, M., Julian, L. W., & Ali, B. (2023). Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection. *Electronics*, *12*(7), 1642. https://doi.org/10.3390/electronics12071642.

Sathyanarayanan, S., & Tantri, R. (2024). Confusion matrix based performance evaluation metrics. *African Journal of Biomedical Research*, *27*(4S), 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345.

Saydul, A. M., Abu, J. M., & Nick, R. (2024). *PhishGuard: Machine learning-powered phishing URL detection*. School of Computing Sciences & Computer Engineering, University of Southern Mississippi, Hattiesburg, USA.

Thotwe, A. H., & Mane, S. B. (2024). *Phishing detection using multi-layer perceptron and comparison of accuracy with various neural network techniques.*