ASTEEC Conference Proceeding: Computer Science
3rd International Conference on Information Science and Technology Innovation (ICoSTEC)
July 27, 2024, Yogyakarta, Indonesia

Rectified Linear Unit Function and its Variants: Exploring Options for Web Phishing Classification

Adam Sagara[1], Teddy Surya Gunawan[2], Wanayumini[3]

Magister of Computer Science, Universitas Potensi Utama

[1]adamsagara@protonmail.com

[2]tsgunawan@gmail.com

[3]wanayumini@gmail.com

***Abstract*—In today's digital landscape, the rise of web phishing poses a significant threat to cybersecurity, prompting urgent research into detection and prevention strategies. This study investigates the efficacy of different activation functions within Multilayer Perceptron (MLP) models for detecting phishing websites, utilizing a dataset with 87 features reduced to 60 using Principal Component Analysis (PCA). Evaluation metrics including accuracy, precision, recall, F1-score, and AUC are computed across four activation functions: ReLU, Leaky ReLU, Parametric ReLU, and Exponential Linear Unit (ELU). Results demonstrate consistently high performance across all activation functions, with slight improvements observed with Leaky ReLU (LReLU) and ELU, particularly in precision and F1-score metrics. These findings underscore the robustness and adaptability of MLP models in handling complex classification tasks like phishing detection. Moreover, the study highlights the importance of considering diverse activation functions in model design, offering insights for future optimization and exploration in cybersecurity research.**

***Keywords*—Web phishing; Multilayer Perceptron; Activation functions; Cybersecurity; Phishing detection**

I. INTRODUCTION

In the rapidly evolving digital era, the threat of web phishing has emerged as a major concern in cybersecurity [1]. Malicious actors engage in phishing practices attempting to obtain sensitive information from internet users by masquerading as trustworthy entities [2]. With increasingly sophisticated techniques, phishing attacks can deceive users and lead to significant financial and privacy losses [3]. Hence, research on the detection and prevention of phishing on web platforms has become critically important.

Various studies have been conducted previously to address this issue, such as detecting websites and domains at the browser level [4], identifying phishing websites by retrieving attributes from large clusters of data [5], or employing machine learning to detect phishing websites through an analysis of page layout similarity [6]. Nevertheless, these approaches are deemed insufficient in protecting internet users from web phishing attacks [7]. As a result, some researchers utilize machine learning algorithms to automatically classify whether a website is legitimate or phishing. For instance, Support Vector Machine and Random Forest are employed to ascertain the likelihood of a phishing website [8], while Random Forest and Artificial Neural Network are used to identify phishing websites [9]. Additionally, phishing website detection is conducted using multiple machine learning algorithms [10], and Naive Bayes is employed to classify websites as either phishing or legitimate [11].

The dataset used in this study contains 87 features that have been reduced using Principal Component Analysis (PCA). This research adopts an innovative approach by leveraging the Rectified Linear Unit (ReLU) activation function and its variants in the context of phishing classification using the Multilayer Perceptron algorithm. The classification process is carried out using the Multilayer Perceptron (MLP) algorithm, known for its effectiveness in handling complex classification problems [12]. During the classification process, the performance comparison is conducted between the ReLU activation function and other variants such as Leaky ReLU, Parametric ReLU, and Exponential Linear Unit (ELU). The performance of each model is evaluated using the 10-fold cross-validation technique to generate comprehensive evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic curve (AUC), which are then used to determine the best activation function variant for phishing detection.

The contribution of this research is highly significant in the development of phishing detection techniques on web platforms. The findings can enhance understanding of the effectiveness of various activation functions in the context of phishing classification, and provide valuable insights into the use of MLP models for cybersecurity. It is hoped that the findings of this research will contribute to improving the reliability of phishing detection systems and broadly support efforts to prevent cybercrime..

II. METHOD

A. Dataset

This study utilizes a dataset comprising a list of websites classified as 'legitimate' and 'phishing', obtained from kaggle.com [13]. The dataset consists of 11430 entries with 87 features and 1 target.

The first step in this research involves feature reduction on the dataset using Principal Component Analysis (PCA). The objective of this feature reduction is to decrease the dataset's dimensions for a more efficient and effective classification process.

To determine the most optimal number of PCA components, this study applies the elbow method. This stage entails plotting explicit variance against the number of PCA components. In this phase, the research seeks the point where the decrease in explicit variance significantly diminishes, indicating that adding additional components no longer contributes significantly to the data's variance. Figure 1 show the steps on determining the most optima number of PCA components using the elbow method.

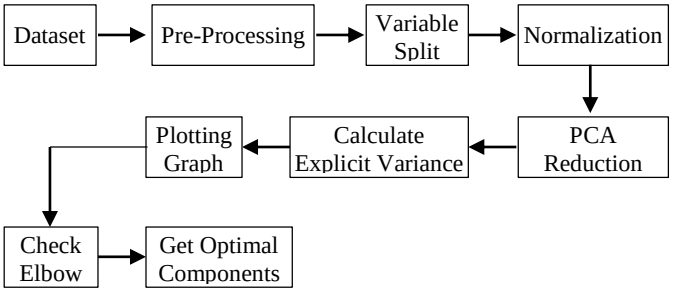

Fig. 1 Applied Elbow Method for Determining Optimum PCA Components

The process of determining the most optimal PCA components in Figure 1 begins with reading the dataset. Following this, data pre-processing is carried out, including removal of irrelevant data, handling missing values, and normalization if necessary. Features from the dataset are separated, and target variables are isolated if present. Subsequently, standardization or normalization of dataset features is performed to ensure uniform value ranges. The PCA algorithm is utilized to reduce dimensions on dataset features, followed by calculating explicit variance for each resulting PCA component. This is followed by creating a plot of explicit variance versus the number of PCA components, where the graph is then examined to identify the point where the decrease in explicit variance significantly diminishes or forms an elbow-like curve. The number of components at that point is designated as the most optimal number of PCA components, which is then used to perform feature reduction on the dataset.

After obtaining the optimal number of components, feature reduction is performed using PCA. This process involves projecting data into a lower-dimensional subspace according to the predetermined number of components. Figure 2 shows the steps on feature reduction using PCA in this research.

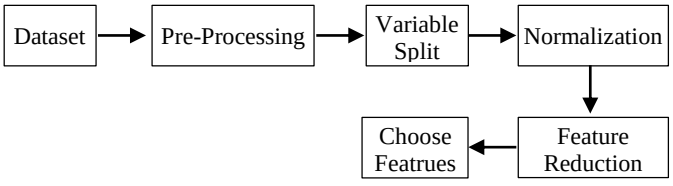

Fig. 2. Applied PCA Method for Feature Reduction

The steps of feature reduction using PCA in Figure 2 begin with reading the dataset to be feature-reduced. Subsequently, data pre-processing is conducted, involving the removal of irrelevant data, handling missing values, and normalizing data if necessary. Features of the dataset are separated, and target variables are also separated if present. Following this, the dataset features undergo standardization or normalization to ensure uniform value ranges. Then, the PCA algorithm is applied to perform dimensionality reduction on the dataset features, with the number of PCA components predetermined based on the elbow method. The subsequent process involves feature reduction, where dataset features are reduced to a lower-dimensional space based on the optimal number of PCA components determined previously. Lastly, feature selection is carried out, where features that are most important or contribute the most to data variance are selected after feature reduction.

B. Web Phising Classification

The reduced feature set is then utilized in the classification process using the Multilayer Perceptron (MLP) algorithm. In this study, four MLP models are designed, each comprising 3 hidden layers with 10 neurons in each layer. The optimization algorithm used is the Adam solver.

Four variations of activation functions were used in this research, namely ReLu, Leaky ReLu, Parametric ReLu, and Exponential ReLu. Each activation function operates differently, but their common goal is to introduce non-linearity into the model, allowing for more complex modeling of relationships between features and targets.

1. ReLU (Rectified Linear Unit)
   The Rectified Linear Unit (ReLU) is a commonly employed activation function in neural networks, characterized by its ability to output the input directly if positive, and zero otherwise [14]. Its mathematical expression is given by the following formula:
   $$f(x) = \max(0, x) \quad (1)$$
2. Leaky ReLU
   The Leaky Rectified Linear Unit (Leaky ReLU) represents a modification of the standard Rectified Linear Unit (ReLU) activation function within neural networks. It is designed to mitigate the "dying ReLU" issue by permitting a small, non-zero gradient when the input is negative, thus ensuring continuous updates to the weights during training [15]. Its mathematical expression is given by the following formula:
   $$f(x) = \begin{cases} x, & x > 0 \\ 0.01x, & otherwise \end{cases} \quad (2)$$
3. Parametric ReLU
   The Parametric Rectified Linear Unit (Parametric ReLU) is an advanced variation of the standard Rectified Linear Unit (ReLU) activation function, characterized by the introduction of learnable parameters. These parameters govern the slope of the negative segment of the function, affording increased flexibility in the neural network model [16]. Its mathematical expression is given by the following formula:
   $$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & otherwise \end{cases} \quad (3)$$
4. Exponential Linear Unit (ELU)

The Exponential Linear Unit (ELU) represents an alternative to the Rectified Linear Unit (ReLU), offering a solution to the issue of vanishing gradients commonly encountered in neural networks. ELU achieves this by incorporating negative values through the utilization of the exponential function, thereby enhancing the model's ability to capture complex relationships in the data [17]. Its mathematical expression is given by the following formula:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & otherwise \end{cases} \quad (4)$$

C. Model Evaluation

To determine the most suitable activation function variation for phishing classification, evaluation is conducted using the 10-fold cross-validation technique. Evaluation metrics employed include accuracy, precision, recall, F1-score, and area under the Receiver Operating Characteristic curve (AUC). This evaluation aids in assessing the performance of each model and determining the best activation function variation for phishing detection. Formula (5) to (8) shows how to calculate each evaluation metrics.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (5)$$

Accuracy measures the proportion of correctly classified instances out of the total instances [18].

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (6)$$

Precision measures the proportion of true positive predictions out of all positive predictions [19].

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (7)$$

Recall measures the proportion of true positive predictions out of all actual positive instances [20].

$$F1 = \frac{2\ X\ Precision\ X\ Recall}{Precision + Recall} \quad (8)$$

F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall [21].

III. Results and Discussion

The investigation into the optimal number of Principal Component Analysis (PCA) components, employing the Elbow method, yielded a value of 60. This value derives from the juncture where a notable reduction in explicit variance forms an elbow-like curve. Subsequently, this value is adopted as the number of PCA components for the ensuing feature reduction process. Figure 3 elucidates the graphical representation of the relationship between explicit variance values and the number of PCA components.

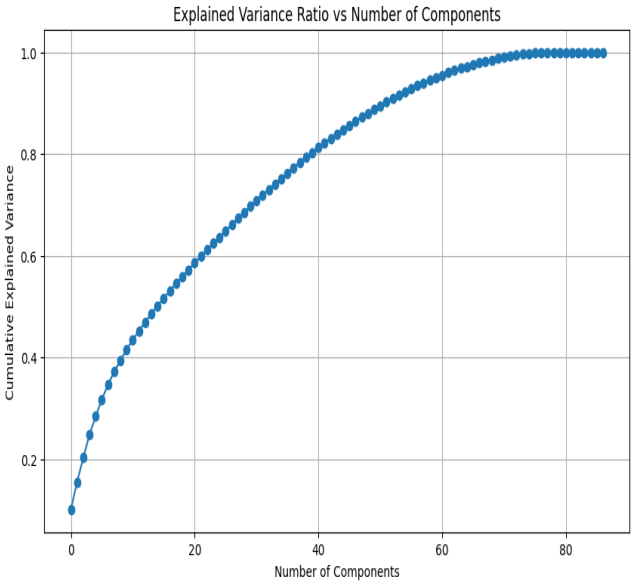

Fig. 3. Explicit Variance Versus PCA Components Plot

After reducing the dataset from its original 87 features to 60 features using the PCA method, the new dataset including the variance values of each PCA component is utilized within the MLP model with various activation functions such as ReLU, Leaky ReLU, Parametric ReLU, and Exponential Linear Unit. The obtained results include accuracy, precision, recall, F1-score, and AUC values as seen in Table 1.

TABLE I
Model Performance Comparison

Model	Accuracy	Precision	Recall	F1	AUC
ReLu	94,73	94,50	94,97	94,57	98,62
LReLu	94,86	95,02	94,73	94,70	98,61
PReLu	94,72	94,67	94,74	94,54	98,65
ELU	94,79	95,00	94,57	94,61	98,63

The comparison of performance among various activation functions in the Multilayer Perceptron (MLP) model for detecting phishing websites indicates that all activation functions yield excellent performance. There is no significant difference in performance among the different activation functions, with all models achieving high accuracy, precision, recall, F1 score, and AUC. However, there is a slight improvement observed in some metrics with the Leaky ReLU (LReLU) activation function, particularly in precision and F1 score, suggesting that LReLU might be slightly more effective in addressing the "dying ReLU" problem. Additionally, the Exponential Linear Unit (ELU) activation function also demonstrates very good results, similar to LReLU, confirming that ELU could be a viable alternative. Therefore, in the context of phishing website detection, the use of various activation functions in the MLP model provides consistent and robust results, with specific choices possibly depending on preferences and specific requirements of the classification task at hand.

These findings underscore the efficacy of employing diverse activation functions within MLP models for phishing website

detection. Additionally, the consistent high performance across various activation functions reflects the robustness and adaptability of the MLP model in handling complex classification tasks. Further exploration could focus on optimizing model parameters or incorporating ensemble methods to enhance performance even further.

IV. CONCLUSIONS

The research explored the optimal number of Principal Component Analysis (PCA) components using the Elbow method, resulting in 60 components, indicative of a significant reduction in explicit variance forming an elbow-like curve. Subsequently, this value was adopted for the feature reduction process. Post-reduction, the dataset, originally comprising 87 features, was transformed into 60 features, incorporating variance values for each PCA component. These reduced features were integrated into the Multilayer Perceptron (MLP) model alongside various activation functions, including ReLU, Leaky ReLU, Parametric ReLU, and Exponential Linear Unit (ELU). Evaluation metrics such as accuracy, precision, recall, F1-score, and AUC were computed, revealing high performance across all activation functions, with slight enhancements observed in metrics like precision and F1-score with Leaky ReLU (LReLU). ELU also demonstrated promising results akin to LReLU, suggesting its potential as an alternative. This underscores the consistent and robust performance of various activation functions within MLP models, with preferences and task-specific requirements possibly influencing activation function selection. These findings emphasize the effectiveness of employing diverse activation functions in MLP models for phishing website detection. Moreover, the study highlights the MLP model's adaptability in handling complex classification tasks, suggesting avenues for further optimization and exploration, such as parameter tuning and ensemble methods, to enhance performance.

REFERENCES

[1] M. Madlenak and K. Kampova, "Phishing as a Cyber Security Threat," in *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, IEEE, Oct. 2022, pp. 392–396. doi: 10.1109/ICETA57911.2022.9974817.

[2] B. Naqvi, K. Perova, A. Farooq, I. Makhdoom, S. Oyedeji, and J. Porras, "Mitigation strategies against the phishing attacks: A systematic literature review," *Comput. Secur.*, vol. 132, p. 103387, 2023, doi: 10.1016/j.cose.2023.103387.

[3] P. M. Dinesh, M. Mukesh, B. Navaneethan, R. S. Sabeenian, M. E. Paramasivam, and A. Manjunathan, "Identification of Phishing Attacks using Machine Learning Algorithm," *E3S Web Conf.*, vol. 399, 2023, doi: 10.1051/e3sconf/202339904010.

[4] N. Nadia, W. Leewando, J. Paulus, and V. Nooril, "Phishing Detection Applications for Website and Domain at Browser using Virustototal API," *Eng. Math. Comput. Sci. J.*, vol. 5, no. 2, pp. 93–96, 2023, doi: 10.21512/emacsjournal.v5i2.9998.

[5] M. R. Natadimadja, M. Abdurohman, and H. H. Nuha, "A Survey on Phishing Website Detection Using Hadoop," *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 237, 2020, doi: 10.32493/informatika.v5i3.6672.

[6] J. Mao *et al.*, "Detecting Phishing Websites via Aggregation Analysis of Page Layouts," *Procedia Comput. Sci.*, vol. 129, pp. 224–230, 2018, doi: 10.1016/j.procs.2018.03.053.

[7] M. E. Pratiwi, T. A. Lorosae, and F. W. Wibowo, "Phishing Site Detection Analysis Using Artificial Neural Network," *J. Phys. Conf. Ser.*, vol. 1140, no. 1, 2018, doi: 10.1088/1742-6596/1140/1/012048.

[8] M. Al saedi and N. Abbas Flayh, "Phishing Website Detection Using Machine Learning: A Review," *Wasit J. Pure Sci.*, vol. 2, no. 2, pp. 270–281, 2023, doi: 10.31185/wjps.145.

[9] S. A. Khan, W. Khan, and A. Hussain, "Phishing Attacks and Websites Classification Using Machine Learning and Multiple Datasets (A Comparative Analysis)," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12465 LNAI, pp. 301–313, 2020, doi: 10.1007/978-3-030-60796-8_26.

[10] N. Pawar and P. Tijare, "Machine Learning Approach for Detection of Phishing Website," in *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2023, pp. 1694–1700. [Online]. Available: https://www.ijraset.com/best-journal/machine-learning-approach-for-detection-of-phishing-website

[11] R. A. Roni Anagora, R. Rudini, R. T. Rohmat Taufiq, A. D. J. Ahmad Dedi Jubaedi, R. W. Rio Wirawan, and Arman Syah Putra, "The Classification of Phishing Websites using Naive Bayes Classifier Algorithm," *Int. J. Sci. Technol. Manag.*, vol. 3, no. 2, pp. 553–562, 2022, doi: 10.46729/ijstm.v3i2.498.

[12] I. Firmansyah and R. Rosnelly, "Inception-V3 Versus VGG-16 : in Rice Classification Using Multilayer Perceptron," in *2nd International Conference on Information Science and Technology Innovatin (ICoSTEC)*, 2023, pp. 1–5. [Online]. Available: https://prosiding-icostec.respati.ac.id/index.php/icostec/article/view/24

[13] S. Tiwari, "Web page Phishing Detection Dataset," 2021. [Online]. Available: https://www.kaggle.com/datasets/shashwatwork/web-page-phishing-detection-dataset?resource=download

[14] A. D. Rasamoelina, F. Adjailia, and P. Sincak, "A Review of Activation Function for Artificial Neural Network," in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, IEEE, Jan. 2020, pp. 281–286. doi: 10.1109/SAMI48414.2020.9108717.

[15] S. Sharma, S. Sharma, and A. Anidhya, "Activation Functions in Neural Networks," *Int. J. Eng. Appl. Sci. Technol.*, vol. 4, no. 12, pp. 310–316, 2020, [Online]. Available: https://ijeast.com/papers/310-316,Tesma412,IJEAST.pdf

[16] D. Yang, K. M. Ngoc, I. Shin, and M. Hwang, "DPReLU: Dynamic Parametric Rectified Linear Unit and Its Proper Weight Initialization Method," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, 2023, doi: 10.1007/s44196-023-00186-w.

[17] J. C. Chen and Y. M. Wang, "Comparing activation functions in modeling shoreline variation using multilayer perceptron neural network," *Water (Switzerland)*, vol. 12, no. 5, 2020, doi: 10.3390/W12051281.

[18] D. Pardede, Wanayumini, and R. Rosnelly, "A Combination Of Support Vector Machine And Inception-V3 In Face-Based Gender Classification," *Int. Conf. Inf. Sci. Technol. Innov.*, vol. 2, no. 1, pp. 34–39, Mar. 2023, doi: 10.35842/icostec.v2i1.30.

[19] K. F. Margolang, M. Zarlis, and Hartono, "Sentiment Classification on Mandalika MotoGP Event Using K-Means Clustering and Random Forest," in *2nd International Conference on Information Science and Technology Innovatin (ICoSTEC)*, 2023, pp. 65–69.

[20] M. M. Siregar, Roslina, and B. H. Hayadi, "Predicting Non-Performing Loan ' s Risk Level Using K- Means Clustering and K-Nearest Neighbors," in *2nd International Conference on Information Science and Technology Innovatin (ICoSTEC)*, 2023, pp. 187–192. [Online]. Available: https://prosiding-icostec.respati.ac.id/index.php/icostec/article/view/55/55

[21] S. Riyadi, Hartono, and Wanayumini, "Predicting Children's Talent Based On Hobby Using C4.5 Algorithm And Random Forest," in *International Conference on Information Science and Technology Innovation (ICoSTEC)*, Mar. 2023, pp. 182–186. doi: 10.35842/icostec.v2i1.54.