

## Neural Network based Phishing Website Detection with Using Firefly Algorithm

Qaderiya Jaafar Mohammed ALHAIDAR<sup>1</sup>, Saja hassan Musa ALDRAJY<sup>2</sup>,  
Mohammad Mahdi BORHAN ELMI<sup>3</sup>

<sup>1,2</sup>Islamic Azad University, Iran

<sup>1</sup><https://orcid.org/0009-0006-8551-3041>

<sup>2</sup><https://orcid.org/0009-0000-6780-2284>

<sup>3</sup>Istanbul Aydin University, Turkey, <https://orcid.org/0009-0007-1705-0214>

**ABSTRACT:** In the digital world, phishing is one of the cybersecurity threats that can cause security and financial problems for operators and internet service providers. Until nowadays, various techniques and methods have been proposed to detect phishing websites. In this paper, a modified multilayer perceptron neural network with firefly meta-heuristic algorithm is used to increase the accuracy of phishing website detection. The phishing databases are taken from the UCI dataset collection. The Phishing Tank dataset under study consists of 31 features in 31 columns and 11055 rows. For missing data values, 3-nearest neighbors method was applied for predicting and in order to reduce the dimensions of the data, the principal component analysis (PCA) method was used. In this situation, 27 out of 31 features were identified as main features. For this purpose, each firefly considered as a candidate which has two dimensions. The first dimension specifies the weight of the neural network links and the second dimension determines their bias. Using firefly algorithm helps multilayer perceptron neural network to get more exploration in the search space so it improves the probability to reach global optimal point in considered problem. The results showed that by using the proposed method, the accuracy of phishing website detection increased by 98.2%. In other words, this method demonstrated that using a combination of multilayer perceptron neural network and firefly metaheuristic algorithm can be an effective solution for detecting phishing websites.

**KEYWORDS:** Phishing website, Neural network, Firefly algorithm, Multi-layer perceptron

### INTRODUCTION

The theft of information or phishing is considered as a one of the main challenges in the field of information technology. This security problem has destructive effects on the security of the Internet and especially financial exchanges. Phishing mostly happens in the case of financial transactions and it takes place in the form of designing a number of fake websites in order to deceive users. To implement it, a phisher or hacker uses social engineering methods to create a sense of trust in users and directs them to fake websites. The appearance of the designed site is very similar to the main site desired by the user, and by typing the user name and password, the user account details of the operators are sent to Fisher. Then, by stating the error of not accessing or updating the site, the user's operation is cancelled and the hacker uses the obtained information to carry out his desired activities. Phishing websites usually have features such as long addresses to prevent users from recognizing the real address, increasing the number of address points more than usual, and using IP addresses.

In order to detect fake sites, various methods have been provided to date, one of which is the use of a blacklist. In this

method, there is a list of fake sites, with the help of which and comparison with legal sites, phishing sites are identified. Despite the easy implementation, using of the existing method is not reliable because it needs to be continuously updated and due to need significant search time in the updated list, its application is limited (Zuraiq and Alkasassbeh, 2019). Another method that has been used is the heuristic method, in which extracted rules are used to identify fake pages, but unfortunately this method is not reliable because phishers constantly change their techn. Unlike the previous methods, the use of data mining in detecting phishing attacks can be more effective because it is able to perform identification operations with higher accuracy with the help of extracting important and effective features (Orunsolu et al, 2022). In Figure 1, the category of cyber-attacks is expressed in the form of a diagram and shows that phishing has the highest share among them (Niakanlahiji et al, 2018). The main contribution of this paper are as follows:

- Using multilayer perceptron neural network for classification of phishing websites from original ones.

- Improving the performance of neural network by using firefly algorithm for regulating the weights and biases of every links in each time step
- Applying principal component analysis for reducing the number of features for increasing the convergence rate and reducing the processing time

In the rest of the paper, at first, explanations about the basic concepts of phishing are presented, then various common methods of identifying fake websites are introduced. In the following, the structure of the neural network and the

algorithm are expressed, and the results of the review of literatures related to the phishing are presented. In the next section, the specifications of the proposed approach as well as the mathematical modeling of it are designed and after introducing performance quality evaluation indices, obtained results are shown. At the end of this paper, the results of the proposed approach with two algorithms ANN and GSA-FDNN have been compared and finally a brief summation of all the steps is reviewed in conclusion.

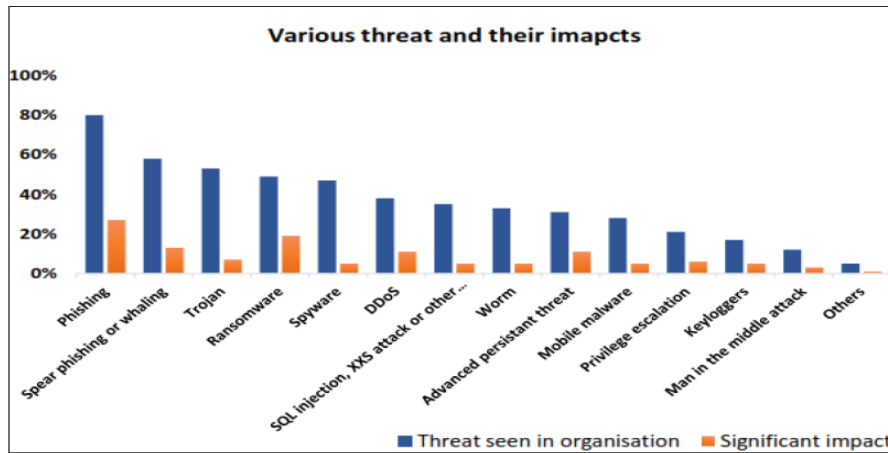


Figure 1. Classifications of cyber attacks. Adapted from Niakanlahiji et al (2018).

### Phishing

Phishing is a type of cyberspace attack where the attacker obtains the user’s information by using various communication methods and designs sites that are very similar to the original sites used by operators so this is very similar to trapping fish (Erdemir and Altun, 2022). The word phishing was first used in 1996, when by sending many e-mails apparently from reputable companies or banks, the victims were directed to illegal sites and important information such as their usernames and passwords were extracted (Al-Ahmadi, 2020). The life cycle of a phishing attack includes planning, regulating, attack, information gathering and frauding (Drury et. al, 2022). There are many types of phishing attacks which the most important are as follows:

- Spear Phishing
- Colony Phishing
- Pop-up phishing
- Pharming Phishing
- Whaling Phishing
- Email Spoofing
- Website Redirects
- Typosquatting
- Watering Hole
- Impersonation & Giveaways
- Malicious Applications
- Text and Voice Phishing

Based on the division of the Anti-Phishing Working Group (APWG), defense mechanisms against phishing attacks are

generally divided into three groups: detection, prevention and correction methods. In order to detect and prevent such attacks, various solutions have been presented which are divided into two levels (Fette et al, 2007). The first level is related to techniques designed to detect and filter phishing e-mails, and in the second level, using address and content evaluation methods, fraudulent sites are identified. In general, phishing attacks have been identified by using approaches such as blacklist & whitelist (Heron, 2009), evaluation of Internet uniform resource locator (URL) (Patil et al, 2018), examination of appearance features and content used in different sites (Chang et al, 2013), and hybrid approaches.

### LITERATURE REVIEW

Due to the widespread use of web pages in people's daily lives, phishing attacks cause a lot of damage to people and organizations. In general, until today, phishing attacks are usually using approaches such as black-white list (Heron, 2009), evaluation of URL (Patil et al, 2018), examination of appearance features and content used in sites (Chang et al, 2013) and hybrid identification approaches. In (Assefa and Katarya, 2022), machine learning based methods were used to classify websites based on their URL. In this regard, 4 different categories were designed and 1,353 different page addresses were classified into three groups: legal, suspicious or phishing site. The aim of the authors (Zhou et al, 2023) were to detect a phishing site by extracting the features by the machine learning algorithm. In this regard, the URL features of the sites were first extracted by a Convolutional based

automatic encoder, and then deep neural network was used for classification. The main challenge observed in the mentioned researchs are overfitting, low accuracy, and

ineffectiveness of machine learning solutions if sufficient training data is not available. A summary of the comparison results of different methods is shown in Table 1.

**Table 1. Comparison of different methods in phishing identification**

Ref	Method	Advantages	Disadvantages
Almoussa et al, 2019	Using a LSTM network to extract global features Using a CNN to detect the effectiveness of parameters Combining these two methods to achieve better performance	High accuracy in detecting Automatically extracting features Have memory	The impossibility of properly training a deep neural network with a small amount of data
Bozkir et al, 2020	Creates URL-based phishing samples using adversarial networks. The generated samples have the ability to fool Blackbox phishing detectors, these samples work to deceive the phishing detection patterns based on Blackbox machine learning.	Ease of implementation	Inefficiency of the network to classify non-image data
Kumar and Indrani, 2021	Two datasets (Phishtank and UCI) are considered for malicious URL detection analysis. In the proposed approach, first the optimal features are extracted from the data set using the DBA-based detector module, and then seventy-five rules are generated using data mining methods.	Using rule reduction algorithm along with classification to predict phishing websites	Long execution time of the proposed method
Elsadiq et al, 2022	Machine learning and deep learning techniques are presented to provide a method capable of identifying phishing websites through URL analysis.	Using URLs Analysis of the frequency of improper details in phishing domains	High complexity and long execution time
Karim et al, 2023	A method based on machine learning models such as Decision Tree (DT), Linear Regression (LR), Random Forest (RF), Naive Bayes (NB), Gradient Boosting Modulator (GBM), K-Neighbors Nearest (KNN) is presented.	Using optimization algorithms	Lack of temporal analysis of the proposed method

**Multilayer perceptron neural network**

Multi-layer artificial neural networks have been used in many researches due to the simplicity of the structure and proper accuracy in information classification. These networks try to create a correct classification of data by imitating the function of the brain and its neurons. Reducing the amount of errors in the classification of phishing attacks actually leads to the separation of fake sites from their original. However, one of the problems of multilayer artificial neural network for the correct classification of websites is the error rate that may not be reduced as much as expected. In fact, the error rate for the classification of phishing attacks is not necessarily minimized by training, and is considered to be a NP-hard optimization problem that does not have a definite solution. For this purpose, multi-layer perceptron (MLP) neural network can be used. A MLP contains at least 3 layers (input, hidden and output layer). This network has the ability to discriminate data that cannot be separated linearly (Noriega, 2005). Perceptrons are actually a special case of artificial neurons that use activation functions such as Heavyside step function.

**Firefly algorithm**

Up To now, about 2,000 different species of fireflies have been registered in the world, most of which produce rhythmic and short flashing lights. There is a Bioluminescence process that causes the phenomenon of night glowing. Researchers believe that this phenomenon occurs to attract the opposite sex for mating and communication with other fireflies in improper situations. Blinking rhythm, rate of change and duration of each blink shape different parts of the communication system. For instance, when fireflies of a particular species mate, the female responds to the male's signal with a unique blinking pattern. Also, in some firefly species such as Photuris, the female has the ability to imitate flashing pattern of other species (for mating). By doing so, the female firefly can trap and eat male of other species that have moved towards the female. The combination of these two important factors makes fireflies visible only from a certain distance. Usually, in the dark of the night, the flashing light of fireflies can be seen from a distance of several hundred meters, which is enough to be observed by other ones and communicate with them. According to the objective

function considered for the algorithm, the flashing light characteristics are modeled. In this regard, the flash intensity of fireflies is determined according to their ranking position in terms of the fitness function defined for them, and then the location of each of them is updated in the next iterations to find the desired optimal solution in the search space. The movement of fireflies is towards other fireflies that have more intensities. In this way, better answers will probably be obtained during the iterations of the algorithm (Yang, 2009)

**METHODOLOGY**

To implement a high-accuracy classification, the data must be properly pre-processed. Otherwise, the classification algorithm will not be properly trained. So, the first step for pre-processing the data is to replace the missing data, because the accuracy of the neural network is greatly reduced by applying the neural network to model without enough information, and sometimes the entire training network may be disrupted. To address this challenge, the method of calculating the average value of the neighbors of the missing data is used. In this situation, using the nearest neighbor algorithm, first 3 neighbors of the mentioned data are identified and then by calculating their average, the amount of missing data is estimated. This helps the neural network in improving the training phase. In the next step, with the help of principal component analysis (PCA), less important data is removed to increase the processing speed. In this regard, the importance of data is evaluated based on their variance. In the next step, data normalization is implemented to control their effectiveness in the process of training network weights. For this purpose, Equation 1 has been used. In the last step, the data is divided into two groups for testing and training. 70% of the data was used to train the neural network and determine the values of weights and biases, and 30% was used to test and evaluate the performance of the neural network. In other words, in the optimization stage, the weights and bias of the neural network are adjusted to increase the accuracy of the network. This process is known as neural network optimization.

$$v' = (v - min_A) \frac{newMax - newMin}{maxA - minA} + (newMin) \tag{1}$$

One of the most famous weight optimization algorithms is the gradient descent algorithm. In this algorithm, the gradient (partial derivative) of the cost function is used to update the weights. In the proposed approach, instead of using the error back-propagation algorithm in the training process of the neural network, the firefly optimization algorithm is used to increase the level of accuracy of the neural network. In this model, the number of neurons in the input layers is considered as the number of features for each sample and the number of neurons in the output layer is equal to the number of classes. The sigmoid function is used as the activation function, which is expressed by the Equation 2.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Among the problems of the error back-propagation method is the possibility of getting stuck in local optimal points. In this regard, the firefly algorithm has been used to adjust the weight coefficients. It should be noted that the parameters of the firefly algorithm must be adjusted based on the characteristics of the problem under study. In this algorithm, fireflies form the candidates of the population. Each firefly has two dimensions and each dimension provides a value for the corresponding weight and bias. In the following, the classification error function of multilayer perceptron is used as a cost function (Equation 3).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2 \tag{3}$$

After the members of each iteration have been evaluated based on the cost function, it is time to update the location. In meta-heuristic firefly algorithm, in each update, the amount of intensity changes based on the cost function. This means that the member with the best fit has more intensity, and members with less intensity move towards the member with more ones. The mathematical model of fireflies' intensity is given by Equation 4. In this equation,  $\varphi_i(t)$ ,  $\varphi_i(t - 1)$ ,  $J(x_i(t))$  are the new value of luciferin, the previous value of luciferin, and the fitness of  $i^{th}$  firefly's location in the  $t^{th}$  iteration of the algorithm, and  $\rho$  and  $\gamma$  are constant coefficients.

$$\varphi_i(t) = (1 - \rho)\varphi_i(t - 1) + \gamma J(x_i(t)) \tag{4}$$

After updating the light level of each firefly, candidates with lower fitness will move to the member with higher fitness with Equation (5). In this equation,  $N_i(t)$  refers to the set of fireflies located in the neighborhood of  $i^{th}$  firefly at  $t^{th}$  time step and  $p_{ij}(t)$  is the Euclidean distance between worm  $i^{th}$  and  $j^{th}$  fireflies. All the mentioned steps for optimization are repeated until the stopping criteria of the algorithm is achieved, then it will choose a candidate as the final solution which has the lowest error for the classifier. Optimization using firefly meta-heuristic algorithm includes the following steps:

- I. Initial population: In this step, the members in the population are defined and introduced. Each candidate is modeled as a firefly, which has two dimensions, weight and bias.
- II. Cost function: The cost function is used to evaluate the fitness of each firefly. In this research, the multi-layer perceptron classification error function is used as the cost function.
- III. Iteration stage: In each iteration of the algorithm, population candidates (fireflies) are evaluated using the cost function. Then, the top members of the population are selected to produce a new

generation. In this stage, the location of candidates are updated to reach the optimal state.

- IV. Termination criteria: The algorithm terminates when a stopping condition specified in the algorithm is met. Typically, one of the stopping conditions can be reaching a certain number of iterations of the algorithm, or reaching a special target value in the cost function

**Evaluation criteria**

To evaluate the effectiveness of the proposed method, criteria such as precision, accuracy and recall criteria are used. Each

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

of these criteria analyze and examine the results related to the classification method. For example, the accuracy criterion is a criterion that measures all the correct identifications of the classifier from the total set (Equation 5). The precision criterion shows the ratio of correct positive identifications to the correct positive and negative ones (Equation 6), and the recall criterion expresses the ratio of correct positive identifications to the sum of correct positive and negative ones (Equation 7). In these relations, TP is true positive, TN is true negative, FP is false positive and FN is false negative. All these parameters can be calculated using the confusion matrix for each data set.

(5)

(6)

(7)

**RESULTS EVALUATION**

The phishing database used in this paper is obtained from the UCI dataset and is a publicly available collection known as

Phishtank. This database contains 11055 rows and covers 31 features. Some of the identification features are stated in the Table 2.

**Table 2. some features for each sample in Fishtank dataset**

Feature	Description
IP-Address	: Using an IP address instead of a domain name on websites can indicate that someone is trying to steal information.
URL_ Length	: Phishers try to hide the suspicious part of the URL by using a long URL.
Shorting_Service	: Short URLs can be redirected to phishing websites.
Having_ At_ Symbol	: Browsers ignore everything typed before the "@" symbol. Then, the phishers put their fake address after that.
Double_slash_redirecting	: "/" symbol is used in URL to redirect users to another page. If the number of "/" in the URL is more than 7, the website is considered as phishing.
Prefix_Suffix	: Prefix or suffix is added to the URL with a line mark. Using this technique, users cannot notice the difference between original and phishing websites.
Have_Sub_Domain	: Subdomains are separated by dot symbol in URL. If the number of subdomains is more than 1, the website is considered as a fraudulent website.
SSL_final_state	: The content of the certificate used on websites must have a valid issuer (Thawte, Go Daddy, Geo Trust, etc.). The age of the certificate must be more than 1 year. If the HTTPS website has the mentioned feature, it can be considered a legitimate website.

For implementing the proposed model, the system with a Core i7 processor and 8 GB Ram memory are used and MATLAB 2016b is selected for simulation. In order to reduce the dimensions of the data, the principal component analysis (PCA) method was used and the 3-nearest neighbors method was used to determine the missing data values. In this situation, 27 out of 31 features were identified as main features and Equation 1 was used to normalize them. After normalization, all data are in the range between -1 and 1. In order to classify the data, the perceptron neural network layered structure has been created according to the Figure 2.

As can be seen, 27 neurons are selected for the input later according to the number of selected features, 10 layers for hidden layer and 2 layers are chosen for output layer equal to number of classification classes. Sigmoid function is used as activation function and the firefly algorithm with the specifications given in Table 3 is used to train the weights of the neural network. Optimization process of perceptron neural network using firefly algorithm to find proper values of weights and biases is shown in Figure 3. This graph is given for 12 iterations of the firefly algorithm and the error value obtained in each iteration is shown. In the initial



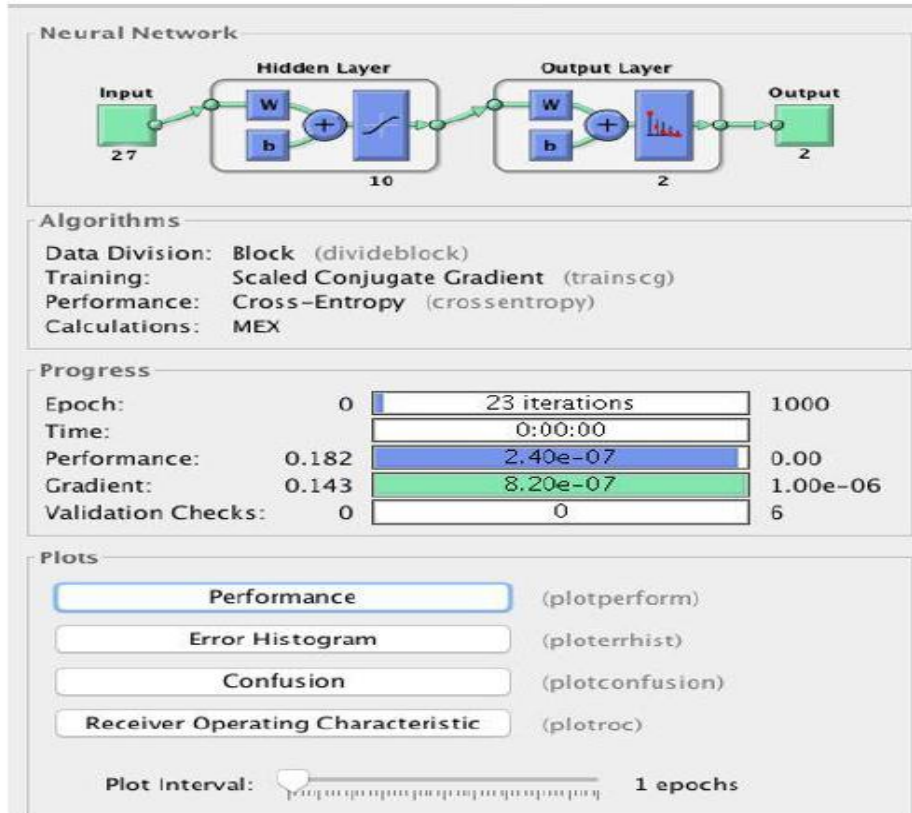
“Neural Network based Phishing Website Detection with Using Firefly Algorithm”

iterations, due to the random selection of weights, the error rate was high, but over time and by updating the values, the

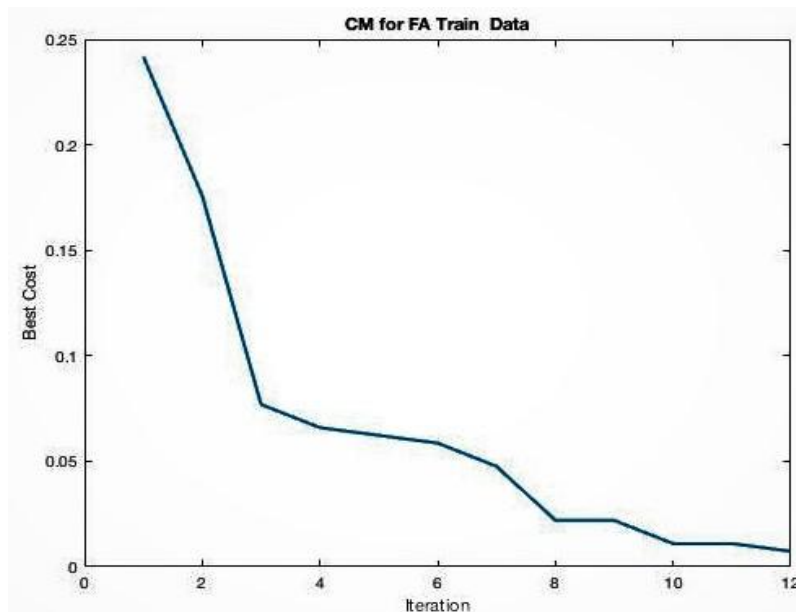
error rate has been reduced and convergence has been achieved.

**Table 3. Parameters of presented algorithm for neural network training**

Target of algorithm	Regulating the weight and bias for each neuron	No. of fireflies	70
Objective function	Mean square error	Stopping criteria	Max. 12 iterations



**Figure 2. Layered structure of neural network.**



**Figure 3. Optimization process of neural network learning with firefly algorithm**

After 12 iterations, the classifier's error rate reached less than 0.05, which indicates the proper training of network weights.

In Figure 4, the results related to the confusion matrix of the dataset used for training are shown. In this matrix, label 1

## “Neural Network based Phishing Website Detection with Using Firefly Algorithm”

represents the phishing label and label 2 represents non-phishing, the green squares represent the TP and TN criteria, and the red squares represent the FP and FN values. Figure 5

shows the results of the confusion matrix for testing the designed neural network. Each of the levels in this matrix represents a value for one of the evaluation criteria

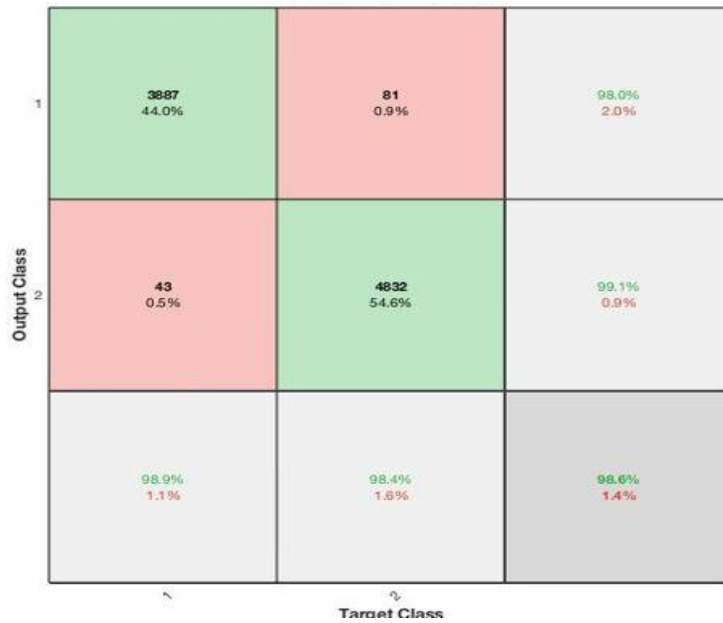


Figure 4. Confusion matrix for training of classifier



Figure 5. Confusion matrix for testing of classifier

The obtained information from the analysis for training and testing are shown in Table 4 and Table 5, respectively. By Comparing the results, we can conclude that the classifier

using the training data has reached 98.6% accuracy, and in the test stage, this accuracy rate reached 98.2%.

Table 4. Result analysis of the training dataset confusion matrix

Accuracy criteria	98%	Negative True detection rate	99.1%
False detection of phishing	0.9%	True detection of original websites	54.6%
True detection of phishing	44%	Overall accuracy	98.6%

**Table 5. Result analysis of the testing dataset confusion matrix**

Accuracy criteria	98.5%	Negative True detection rate	98%
False detection of phishing	0.7%	True detection of original websites	53%
True detection of phishing	45.2%	Overall accuracy	98.2%

**Comparison of results**

In this section, the effectiveness of the proposed approach has been compared with other methods in terms of accuracy and precision criteria. For all these methods, the same dataset was used and the obtained results are shown in Table 6. As it is clear from the results, the proposed approach has better quality results. The reason for the improvement is the training

of perceptron classifier weights using the firefly algorithm. In other words, due to proper preprocessing of data by PCA method and better space search by firefly algorithm, the possibility of getting stuck in local optimal points is reduced and the average of accuracy criteria of the classifier is improved.

**Table 6. Comparison of results with different methods**

Ref.	Method	Accuracy	Recall
Kumar and Indrani, 2021	ANN	82%	86%
Chang et al, 2013	GSA-FDNN[30]	97%	93%
Presented approach		98.2%	97.6%

**CONCLUSION**

Phishing is a widespread cyberspace scam that can fraud internet operators into giving their information to a profiteer. For example, this information can include login username, password and card number’s information. Phishing has caused many security and financial losses to users, and currently, it is considered the most used online fraud mechanism. In this regard, various methods have been proposed to discover phishing websites. One of the most effective solutions are machine learning-based methods. These methods are able to learn the structural features of a phishing website in the training phase and then identify corresponding phishing websites. In this paper, in order to increase the accuracy for identifying phishing websites, the combination of multilayer perceptron neural network with firefly meta-heuristic algorithm is used. In other words, by modifying the post-propagation function of the error, the weight and bias values of the neural network links are updated with the help of the firefly algorithm. The accuracy obtained using the proposed method has been more than 98%, which indicates the performance quality of the presented approach.

**REFERENCES**

1. Al-Ahmadi, S. (2020). PDMLP: phishing detection using multilayer perceptron. *International Journal of Network Security & Its Applications (IJNSA)* Vol, 12.
2. 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.
3. Assefa, A., & Katarya, R. (2022, March). Intelligent phishing website detection using deep learning.

- In 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1741-1745). IEEE.
4. Bozkir, A. S., Dalgic, F. C., & Aydos, M. (2023). GramBeddings: a new neural network for URL based identification of phishing web pages through n-gram embeddings. *Computers & Security*, 124, 102964.
5. Chang, Ee Hung, Kang Leng Chiew, and Wei King Tiong. "Phishing detection via identification of website identity." In 2013 international conference on IT convergence and security (ICITCS), pp. 1-4. IEEE, 2013.
6. Drury, V., Lux, L., & Meyer, U. (2022, August). Dating phish: An analysis of the life cycles of phishing attacks and campaigns. In *Proceedings of the 17th International Conference on Availability, Reliability and Security* (pp. 1-11).
7. Elsadig, M., Ibrahim, A. O., Basheer, S., Alohal, M. A., Alshunaifi, S., Alqahtani, H., ... & Nagmeldin, W. (2022). Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction. *Electronics*, 11(22), 3647.
8. Erdemir, E., & Altun, A. A. (2022). Website Phishing Technique Classification Detection with HSSJAYA Based MLP Training. *Tehnički vjesnik*, 29(5), 1696-1705.
9. Fette, I., Sadeh, N., & Tomasic, A. (2007, May). Learning to detect phishing emails. In *Proceedings of the 16th international conference on World Wide Web* (pp. 649-656).
10. Heron, S. (2009). Technologies for spam detection. *Network Security*, 2009(1), 11-15.



11. Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., & Joga, S. R. K. (2023). Phishing Detection System Through Hybrid Machine Learning Based on URL. *IEEE Access*, 11, 36805-36822.
12. Kumar, M. S., & Indrani, B. (2021). Frequent rule reduction for phishing URL classification using fuzzy deep neural network model. *Iran Journal of Computer Science*, 4, 85-93.
13. Niakanlahiji, A., Chu, B. T., & Al-Shaer, E. (2018, November). Phishmon: A machine learning framework for detecting phishing webpages. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)* (pp. 220-225). IEEE.
14. Noriega, L. (2005). Multilayer perceptron tutorial. *School of Computing. Staffordshire University*, 4(5), 444.
15. Orunsolu, A. A., Sodiya, A. S., & Akinwale, A. T. (2022). A predictive model for phishing detection. *Journal of King Saud University-Computer and Information Sciences*, 34(2), 232-247.
16. Patil, V., Thakkar, P., Shah, C., Bhat, T., & Godse, S. P. (2018, August). Detection and prevention of phishing websites using machine learning approach. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-5). Ieee.
17. Yang, X. S. (2009, October). Firefly algorithms for multimodal optimization. In *International symposium on stochastic algorithms* (pp. 169-178). Berlin, Heidelberg: Springer Berlin Heidelberg.
18. Zhou, J., Cui, H., Li, X., Yang, W., & Wu, X. (2023). A Novel Phishing Website Detection Model Based on LightGBM and Domain Name Features. *Symmetry*, 15(1), 180.
19. Zuraiq, A. A., & Alkasassbeh, M. (2019, October). Phishing detection approaches. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)* (pp. 1-6). IEEE.