Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Phishing Attacks Detection by Using Artificial Neural Networks

Majeed Jasim Nabet[1]*, Loay E. George[2]

1 Informatics Institute for Postgraduate Studies, Baghdad, Iraq

2 University of Information Technology and Communications, Baghdad, Iraq

*Corresponding Author: Majeed Jasim Nabet

DOI: https://doi.org/10.52866/ijcsm.2023.02.03.013
Received May 2023; Accepted July2023 ; Available online August 2023

ABSTRACT: Today's world is heading towards complete digital transformation, and with all its advantages, this transformation involves many risks, the most important of which is phishing. This paper proposes a system that classifies the email as phishing or legitimate. Initially, the samples were brought from different data sets, and then the system extracts the features from all parts of the email. The proposed system uses one of the machine learning algorithms (K-means algorithm) to select the valuable features; the proposed system uses four methods to calculate the distance in the K-means algorithm. After features selection, The paper uses ANN as a classifier to classify emails into phishing and ham, and the proposed system tunes the parameters of ANN to obtain a high percentage of accuracy. The proposed system gave an accuracy equal to 99.4%.

Keywords: Phishing Attacks, K-means, Euclidean Distance, Manhattan Distance, Artificial Neural Networks

1. INTRODUCTION

With the ubiquity of the Internet today, society uses Internet products for various things, such as sharing knowledge, socializing, and conducting multiple financial activities, including purchases, advertising, and sending money [1,2]. This state has led to the emergence of cybercrime; Cybercrime is using computers, communication devices, or networks as a tool for illicit purposes. Such as phishing, identity fraud, theft of financial or credit card payment data, and theft and sale of corporate data [3,4]. Phishing is a crime where perpetrators send fake emails that appear to be from popular and trusted brands or organizations and ask for personal credentials such as bank passwords, usernames, phone numbers, addresses, credit card details, etc. [5,6]. Phishing detection can be classified according to different methods [7], such as List-based approaches; The URLs are saved in the database if these URLs are considered phished sites. When a new URL is used, it compares with the URLs in the database, and if it matches, it is prevented by the browser and is saved in the database for future use. This technique is called blacklisting, and whitelisting is a technique for keeping legitimate URLs in a database and checking for new ones. This technique is characterized by high response time and high detection accuracy. The disadvantages of this technology are that it cannot detect zero-Day attacks, and the size of the databases is significant [8]. Heuristic-based approaches, This technique does not use search by URL's name but common characteristics of phishing sites, such as The URL string and information of the domain name. Antivirus or intrusion detection systems create and use a database of signatures of known attacks to scan websites. A website is considered a phishing website if its heuristic patterns match the signature in the database. The main disadvantage is that Phishers can hide this feature when creating the following phishing site, and virus signatures are often updated more slowly, resulting in zero-day vulnerabilities [9]. Visual Similarity approaches, a list of regularly attacked pages (domain names and screenshots) is kept to protect users against the possible impersonation of such pages. If a user accesses a page not on the trusted list, its content is evaluated against the trusted ones. If there is a high visual similarity, it is regarded as a phishing page since it impersonates one of the trustworthy pages. The disadvantage of this technique is that it relies on visual similarity only. [10]. The machine learning approach, This technique uses machine learning techniques to detect phishing. It uses a large labeled dataset as input to train a classification model, which then categorizes each input sample into a predefined number of classes [11,12]. The trained models are binary classifiers because there are only two classifications in a phishing detection problem (benign and phishing) [12,13]. Deep learning approach, Neural networks are increasingly being used in phishing detection due to their ability to analyze complex patterns and identify anomalies in data. NNs can be trained on large datasets of phishing and non-phishing samples [14].

The literature has extensively covered detecting phishing emails, but there have always been some shortcomings [15,3]. For example, some articles use a data set that contains emails of only one class, or it consists of two classes. Still, the difference between them is very large, or the data set is relatively small, or the data set is from one source,

and this source may be biased. Some works extract features from only one part of the email. Some articles choose the features based on experience or depending on some criteria, such as associating features with persuasion without working with features abstractly. The proposed system extracts the features from all parts of the email using the EMFET tool. The system uses the K-means algorithm to select features significantly impacting classification. The proposed system uses an Artificial Neural Network to classify the samples into legitimate and phishing emails, with tuning parameters for the best result. The paper is structured as Section 2 reviewed some related works; Section 3 explains the proposed model. The proposed model was trained and tested using two commonly used datasets, and the test results are discussed in section 4. Finally, section 5 concludes the contribution of this article.

2. RELATED WORKS

Niu et al., 2017 [16] introduced a hybrid classifier, which uses CS (Cuckoo search) to optimize parameter selection in the kernel function. CS is integrated with a support vector machine to construct a hybrid classifier. Kumar et al., 2020 [17] The researchers used hybrid classification, including SVM and PNN. The researchers have done an XOR operation between SVM outputs and PNN outputs; This work obtained an accuracy of 98.3%. Castillo et al., 2020 [18] used a window of 10 words for analyzing the neighborhood of texts and vectors of different sizes created for testing different NN architectures. Also used the word embeddings as input in the NN, such as backpropagation network, convolutional network, and recurrent neural network. Because this work uses different classifiers, the accuracy ranged from 95.6% to 77.4%, depending on the database and classifier used.

Li et al., 2020 [19] the researchers in this article found that the persuasion principles used in phishing emails are Authority, Reciprocation, and Scarcity; they also extracted the words associated with these principles and then used three machine learning algorithms, which are (K-NN, DT, and Bayes algorithm) to classify the email as phishing or not. The accuracy obtained in this work is equal to 99.42%. Manaswini et al., 2021 [20] This paper used Improved Themis Model. The Improved Themis model has multiple layers to process email data at various levels. After implementing the proposed model, the accuracy was 99.87%. Magdy et al., 2022 [21] The researchers aggregated more than one data set of a different class. Features in this work are extracted based on the content: subject line, body, return address, text, and URL. This work Used Low variance, Principal component analysis, and Chi-squared test to reduce features in the features selection stage. Used ANN, which consists of the input layer, two hidden layers, and an output layer, to classify the samples. The highest accuracy obtained in this work is 99.8%. Mughaid et al., 2022 [22], The researchers used seven methods to classify the samples: a Locally-deep support vector machine, Logistic regression, a Support vector machine, Boosted decision tree, a Neural network, a Decision forest, and Averaged perceptron. Accuracy for classifiers ranges in the range (99.5% to 99.9%), Except the Boosted decision tree is 100%. Butt et al., 2022 [23] the researchers proposed three methods for classifying using SVM, LSTM, and Naive Bayes. The accuracy percentage of the SVM classifier was 99.62%, For the LSTM classifier was 98%, for the NB classifier was 97%.

3. THE PROPOSED SYATEM

This work aggregated two common datasets, SpamAssassin and Naziro. SpamAssassin merged with the Naziro dataset because SpamAssassin only contains legitimate emails, and Naziro contains phishing emails. The SpamAssassin dataset is available at (https://spamassassin.apache.org/old/publiccorpus/), and the Naziro dataset Can be obtained from (https://www.monkey.org/~jose/phishing/). MATLAB is the program used to implement the proposed system. Figure 1 illustrates our proposed model, which consists of four stages:

- Features extraction: At this stage, the proposed system extracts features from all parts of emails.
- Preprocessing stage: In this stage, the proposed system performs the binarization process by knowing the threshold based on the mean, standard deviation, and the highest matching percentage with the actual class.
- Features selection stage: The proposed system uses the K-means algorithm to select the features through which samples can be classified as phishing or not.
- Classification stage: The proposed system classifies the samples into phishing and legitimate samples using ANN.

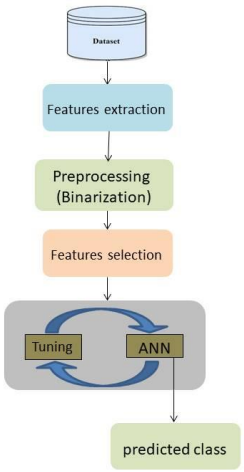

FIGURE 1. The proposed system

3.1 FEATURES EXTRACTION STAGE

Data sets are in the form of actual emails, so the proposed system uses the E-mail Features Extraction Tool (EMFET) to extract features from all parts of emails. The tool is available at (https://github.com/WadeaHijjawi/EmailFeaturesExtraction). It is an open-source and flexible tool to provide the ability to extract the features using any email corpus that have emails with an EML extension. The tool splits the emails' parts (From, To, CC, BCC, Subject, Text Body, and HTML Body) in an output file, and it extracts the selected features in another one as a CSV file. Also, The tool generates An error file if errors occur during extraction. This tool extracts 138 features.

3.2 PREPROCESSING STAGE

At this stage, the proposed system implements binarization to make the data suitable for the neural network. The proposed system Uses the threshold method based on the mean and the standard deviation. Then the proposed system selects the threshold that provides the highest availability of the feature. Figure 2 shows the approach used to implement this stage.

```
For  Each Featuer From 2
Compute Mean And Standard Deviation Of Feature
    For I= -3.2 To 3.2
        Compute Threshold =Mean+Standard Deviation*I
        L=0
        For Each Samples
            If Feature In This Sample > Threshold Then Feature Of Sample Is 1
            Else  Feature Of Sample Is 0
            End If
            If Feature Match Class Then L+1
            End If
        End For
        If L > Biggest Then
            Best Threshold Is Mean+Standard Deviation*I And Best Condition Is Direct
        End If
        L=0
        For Each Samples
            If Feature In This Sample < Threshold Then Feature Of Sample Is 1
            Else  Feature Of Sample Is 0
            End If
            If Feature Match Class Then L+1
            End If
        End For
        If L > Biggest Then
            Best Threshold Is Mean+Standard Deviation*I And Best Condition Is Inversion
        End If
    End For
The Threshold Is Best Threshold And Condition Is Best Condition For Feature
End For
```

FIGURE 2. Binarization approach

3.3 FEATURES SELECTION STAGE

The k-means algorithm is an unsupervised clustering algorithm, but the proposed system uses it in the feature selection process, Where the proposed system makes combinations of binary, triple, quadruple, pentagonal, and hexagonal features and implements the K-means algorithm with these combinations. The matching result of the

algorithm is calculated with the actual classes, and the combinations with a high matching ratio are selected. The proposed system takes the features as pairs (the first feature with the second feature, then the first feature with the third, and so on) and implements the algorithm considering k = 2. After the end of the algorithm, The proposed system examines the centroids of the two clusters. The cluster whose centroid is a phishing sample; is considered a phishing cluster, and the cluster whose centroid is a legitimate sample; is considered a legitimate cluster, then The proposed system calculates the Matching between the predicted class and the actual class, after choosing the top twenty pairs with the highest percentage of Matching and re-work by adding the features to the twenty features so that the assembly becomes a combination of three features and so on.
This work uses four methods to calculate the distance in the K-means algorithm:

i. Euclidean method: where the proposed system uses the Euclidean equation to calculate the distance, which is Eq. (1)

$$d(C,F) = (F_x - C_x)^2 + (F_y - C_y)^2 \quad (1)$$

Where C= centroid; F= feature

Noticed that this method stopped improving at the quadruple combinations, so stopped forming new combinations.

ii. Manhattan method: This method replaces the square with the absolute; it uses Eq. (2)

$$d(C,F) = |F_x - c_x| + |F_y - c_y| \quad (2)$$

Where C= centroid; F= feature

Like the previous one, this method also stopped improving with the quadruple combination.

The proposed system uses another two methods to calculate the distance in the K-means algorithm; the first method that the Euclidean value is divided by the standard deviation to tune the values more, and the second method that the Manhattan value is divided by the standard deviation to tune the values more. Both methods did not give good results.

All previous methods assumed that the first initial centroid is the mean and standard deviation of the first feature, and the second initial centroid is the mean and standard deviation of the second feature. This is in the binary combination, while the triple combination has added the mean of the third feature to the first initial centroid and added the standard deviation of the third feature to the second initial centroid**.**

3.4 CLASSIFICATION STAGE

This stage consists of two steps, the two steps carried out repeatedly to obtain the best results.

3.4.1 TUNING

In this step, the proposed system tunes the parameters of the neural network after each execution of the neural network; The parameters that the system tunes are:

- Input: In the features selection stage, the paper explained that the proposed system used four methods to calculate the distance in the K-means algorithm, and each method gave us different results. Also, the proposed system formed binary, triple, and quadruple combinations in the first two methods and Pentagonal and hexagonal combinations in the other two methods. Therefore, the proposed system used binary and quadruple combinations for all methods and hexagonal grouping for the other two methods in tuning the inputs of the neural network. And every time the proposed system uses one of the combinations, it tunes the number of features used, ten, twenty, or thirty features, as another type of tuning the inputs of the neural network.
- The ratio of training samples and test samples: The proposed system uses the following proportions to tune the neural network, Fifty percent for training and fifty percent for testing, sixty percent for training and forty percent for testing, seventy percent for training and thirty percent for testing, and finally eighty percent for training and twenty percent for testing.
- The number of neurons in the hidden layer: In this parameter, the system first uses seven neurons in the hidden layer and gradually increases their numbers until they reach fifteen.

3.4.2 ARTIFICIAL NEURAL NETWORK

The proposed system uses the ANN to classify the samples into phishing and legitimate. The ANN consists of an input layer, an output layer, and one hidden layer. The system uses the tool (NPRTOOL) from MATLAB To implement the ANN. The NPRTOOL uses the sigmoid function as an activation function in the hidden layer and the

Scaled conjugate gradient backpropagation (a network training function that updates weight and bias values according to the scaled conjugate gradient method) to train the system. Also, The tool computes the performance that is the mean of cross entropy. The equation of cross entropy is Eq. (3). Figure 3 shows the structure of the ANN.

$$CE = -y * \log p - (1 - y) * \log(1 - p) \quad (3)$$

Where y= probability of output, p= probability of actual class

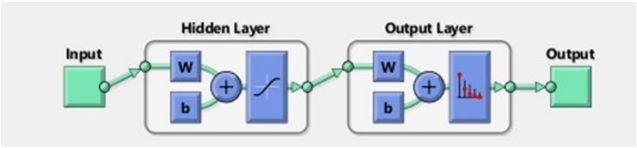

FIGURE 3. -ANN structure

4. RESULTS AND DISCUSSION

The evaluation criteria used in this paper is total accuracy, which is the mean of training accuracy and test accuracy. The training and test accuracies can be obtained from Eq. (4).

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100 \quad (4)$$

Where TP= True Positive, TN= True Negative, FP= False Positive, FN= False Negative.

Figure 4 shows the results obtained from the Euclidean method. The y-axis represents the total accuracy, and the x-axis represents the combinations used. The system compares the results with the result of using all the features to see if the results were close to the result of using all the features. The results showed that in the case of choosing the binary combination and thirty features, the results would be the best. The accuracy of the best result is 99.4%.

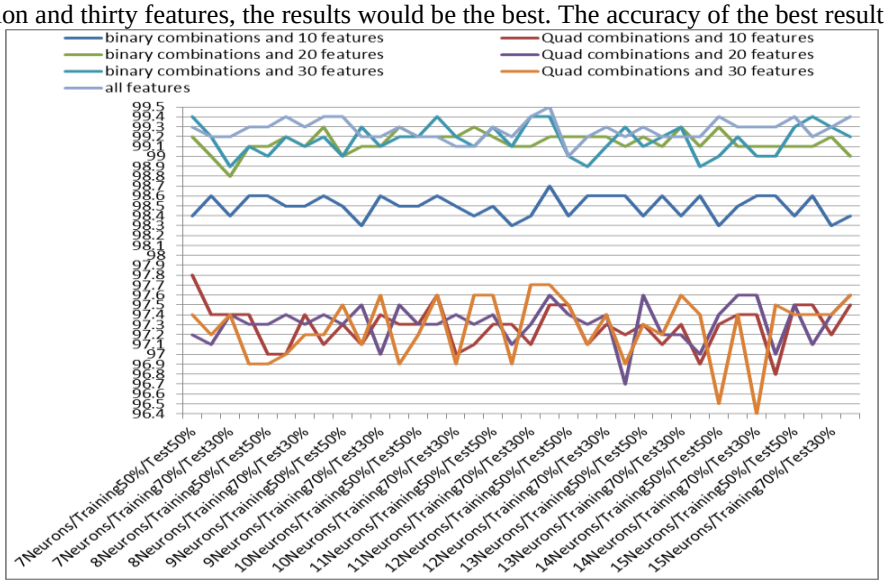

FIGURE 4. The results of the Euclidean method

As for the Manhattan method, Figure 5 shows that choosing the binary combination and twenty features is the best choice. The result reaches 99% if using nine neurons in the hidden layer, the training rate 50%, and the test rate 50%.

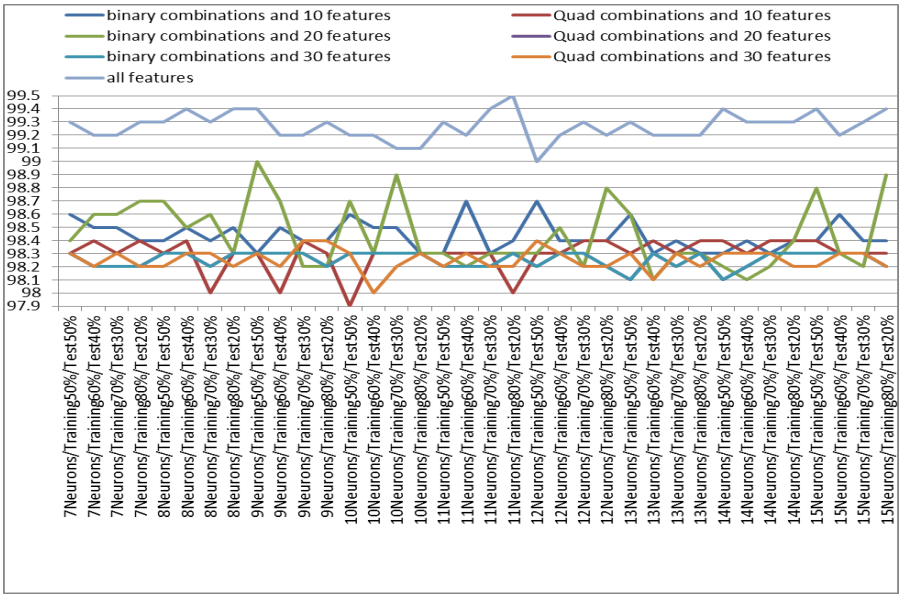

FIGURE 5. The results of the Manhattan method

The results of the third method used in K-means are not Influential. One case is Influential in The fourth method; when using the binary combination and ten features, when the number of neurons in the hidden layer is 9, the training rate is 50%, and the testing rate is 50%, the result will reach 99%.

When comparing the best cases in all methods, it was noted that the Euclidean method was the best as it gave the best results. Total accuracy was obtained equal to 99.4% in more than one case. Figure 6 shows the differences between the methods.

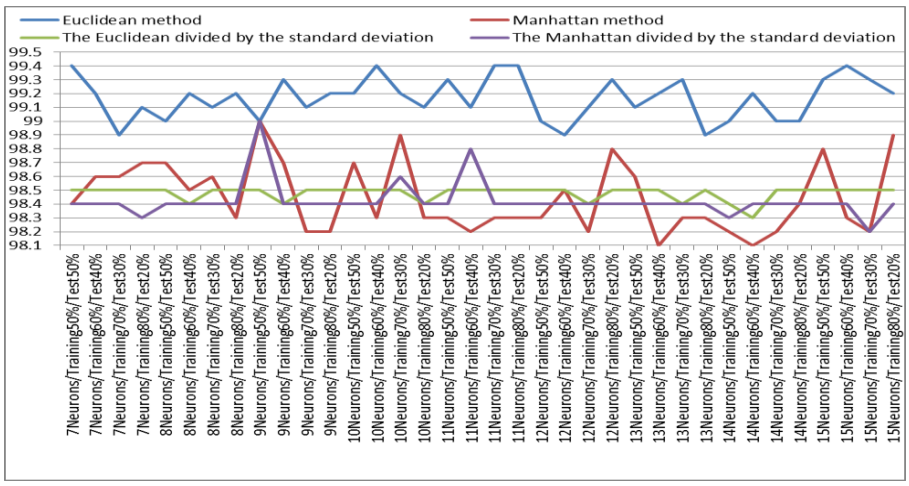

FIGURE 6. The best results in the four methods

Table 1 shows the five cases whose accuracy was 99.4% in the Euclidean method (all cases in binary combinations and 30 features).

Table 1. The best cases

case	No. neurons	Training Ratio	Test Ratio	performance
1	**7**	**50%**	**50%**	**0.0108**
2	**10**	**60%**	**40%**	**0.0214**
3	**11**	**70%**	**30%**	**0.0204**
4	**11**	**80%**	**20%**	**0.0201**
5	**15**	**60%**	**40%**	**0.0179**

This work's results are considered good compared with other researches, as most of those studies did not reach this result, and the use of the hidden single-layer neural network is fast and does not require a long time. Also, the rest of the studies did not tune the parameters of the neural network but only used standard parameter tuning.

5. CONCLUSION

This paper aggregated two data sets from different sources, containing phishing and legitimate emails, for the data to be convergent to know whether the proposed system is effective. The proposed system extracted 138 features from all parts of the email. It subjected them to the K-means algorithm (one of the machine learning algorithms) to reduce the features that will be input to the Artificial Neural Network, and the system used the Artificial Neural Network as a classifier to classify the samples into phishing and legitimate. The proposed system tuned the parameters of the Artificial Neural Network to obtain the highest possible accuracy. The total accuracy of the proposed system reached 99.4%. In the future, the K-means algorithm can be replaced by another machine learning algorithm to improve accuracy, and another type of neural network can be used to increase accuracy.

Funding

None

ACKNOWLEDGEMENT

I thank everyone who supported and encouraged me, My special thanks to my family.

AUTHOR'S DECLARATION

CONFLICT OF INTEREST:

None

I hereby confirm that all figures, tables, and measurements in the manuscript are mine. All obtained figures and geometrical aspects in this study have been drawn and measured via Microsoft Excel ver. 2010 as a way of increasing the correctness of the results.

REFERENCES

[1] M. Almseidin, A. A. Zuraiq, M. Alkasassbeh, and N. Alnidami, "Phishing Detection Based on Machine Learning and Feature Selection Methods," International Journal of Interactive Mobile Technologies, vol. 13, no. 12, p. 171, Dec. 2019, doi: 10.3991/ijim.v13i12.11411.

[2] A. Livara and R. M. Hernandez, An Empirical Analysis of Machine Learning Techniques in Phishing E-mail detection. 2022. doi: 10.1109/iconat53423.2022.9725434.

[3] Kathiravan, Rajasekar, Parvez, Durga, Meenakshi, and Gowsalya, "Detecting Phishing Websites using Machine Learning Algorithm," 7th International Conference on Computing Methodologies and Communication (ICCMC), pp. 5–270, 2023.

[4] P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," Journal of Network and Computer Applications, vol. 210, p. 103545, Jan. 2023, doi: 10.1016/j.jnca.2022.103545.

[5] A. Jain and B. B. Gupta, "Phishing Detection: Analysis of Visual Similarity Based Approaches," Security and Communication Networks, vol. 2017, pp. 1–20, Jan. 2017, doi: 10.1155/2017/5421046.

[6] M. Somesha and AR. Pais, "Classification of Phishing Email Using Word Embedding and Machine Learning Techniques," J Cyber Secur Mobil, pp. 279–320, 2022.

[7] D. N. Quang, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," IEEE Access, vol. 10, pp. 36429–36463, Jan. 2022, doi: 10.1109/access.2022.3151903.

[8] P. Kalaharsha and BM. Mehtre, "Detecting Phishing Sites--An Overview," arXiv Prepr arXiv210312739, 2021.

[9] M. Hara, A. Yamada, and Y. Miyake, Visual similarity-based phishing detection without victim site information. 2009. doi: 10.1109/cicybs.2009.4925087.

[10] S. Abdelnabi, K. Krombholz, and M. Fritz, VisualPhishNet: Zero-Day Phishing Website Detection by Visual Similarity. 2020. doi: 10.1145/3372297.3417233..

[11] S. A. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," IEEE Access, vol. 10, pp. 65703–65727, Jan. 2022, doi:

10.1109/access.2022.3183083.

[12] D. M. Divakaran and A. Oest, “Phishing Detection Leveraging Machine Learning and Deep Learning: A Review,” arXiv Prepr arXiv220507411, vol. 20, no. 5, pp. 86–95, Sep. 2022, doi: 10.1109/msec.2022.3175225.

[13] A. A. Nafea, N. Omar, and M. Q. Al-Ani, “Adverse Drug Reaction Detection Using Latent Semantic Analysis,” Journal of Computer Science, vol. 17, no. 10, pp. 960–970, Oct. 2021, doi: 10.3844/jcssp.2021.960.970.

[14] M. Q. Al-Ani, N. Omar, and A. A. Nafea, “A Hybrid Method of Long Short-Term Memory and Auto-Encoder Architectures for Sarcasm Detection,” Journal of Computer Science, vol. 17, no. 11, pp. 1093–1098, Nov. 2021, doi: 10.3844/jcssp.2021.1093.1098.

[15] R. Eckhardt and S. Bagui, “A User-centric Focus for Detecting Phishing Emails,” AI, Mach Learn Deep Learn a Secur Perspect, 2023.

[16] N. Weina, X. Zhang, G. Yang, Z. Ma, and Z. Zhuo, Phishing Emails Detection Using CS-SVM. 2017. doi: 10.1109/ispa/iucc.2017.00160.

[17] A. Kumar, J. M. Chatterjee, and V. García-Díaz, “A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing,” International Journal of Electrical and Computer Engineering, vol. 10, no. 1, p. 486, Feb. 2020, doi: 10.11591/ijece.v10i1.pp486-493.

[18] E.Castillo, S. Dhaduvai, P.Liu, KS.Thakur, A.Dalton, and T.Strzalkowski, “Email threat detection using distinct neural network approaches,” Proceedings for the First International Workshop on Social Threats in Online Conversations: Understanding and Management, pp. 48–55, 2020.

[19] X. Li, D. Zhang, and B. Wu, Detection method of phishing email based on persuasion principle. 2020. doi: 10.1109/itnec48623.2020.9084766.

[20] M.Manaswini and DRN.SRINIVASU, . “Phishing Email Detection Model using Improved Recurrent Convolutional Neural Networks and Multilevel Vectors,” Ann Rom Soc Cell Biol, vol. 25, no. 6, pp. 16674–16681, Jul. 2021.

[21] S. Magdy, Y. Abouelseoud, and M. Mikhail, “Efficient spam and phishing emails filtering based on deep learning,” Computer Networks, vol. 206, p. 108826, Apr. 2022, doi: 10.1016/j.comnet.2022.108826.

[22] A. Mughaid, S. AlZu’bi, A. A. Hnaif, S. Taamneh, A. Alnajjar, and E. A. Elsoud, “An intelligent cyber security phishing detection system using deep learning techniques,” Cluster Computing, vol. 25, no. 6, pp. 3819–3828, May 2022, doi: 10.1007/s10586-022-03604-4.

[23] U. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi, and A. Ahmadian, “Cloud-based email phishing attack using machine and deep learning algorithm,” Complex & Intelligent Systems, vol. 9, no. 3, pp. 3043–3070, Jun. 2022, doi: 10.1007/s40747-022-00760-3.