Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Safeguarding Patient Data: Machine Learning for Phishing URL Detection in Healthcare Systems

Alaa Abdulshaheed Mousa[1], Saif Al-Deen H. Hassan[2,*], Mohammed Kareem Rashid[3], Moumal Al-Saady[4]

[1] College of Dentistry, University of Misan, Maysan, Iraq
[2] Department Business Administrator, College of Administration and Economics, University of Misan, Maysan, Iraq
[3] Electronic Computing Center, University of Misan, Maysan, Iraq
[4] Faculty of Science, Engineering and Built Environment, Deakin University, Burwood VIC 3125, Australia

ARTICLE INFO

Article history:
Received 10 January 2025
Received in revised form 3 March 2025
Accepted 2 May 2025
Available online 16 May 2025

Keywords:
Phishing URL detection; machine learning; healthcare cybersecurity; multilayer perceptron; decision tree; Naive Bayes; patient data security

ABSTRACT

Since the healthcare industry depends more and more on digital infrastructure, it is a perfect target for cyberattacks especially phishing. Designed to pass for real healthcare websites, phishing URLs seriously compromise patient data security. Effective strikes may cause disruptions in patient care, financial losses, and Protected Health Information (PHI) breaches. This work investigates the use of machine learning (ML) approaches for robust and accurate phishing URL detection in healthcare systems in order to handle this important problem. We examine a Multilayer Perceptron (MLP) neural network-based detection model and evaluate its performance against known techniques, Decision Tree (DT) and Naive Bayes (NB). Comprehensive URL datasets—more especially, the ISCX-URL-2016 dataset for training and testing—as well as the CIC-InvesBanking-2017 dataset combined with live phishing feeds for validation help to train and rigorously validate the models. With a high accuracy of 87.45% on test data and a precision of 84% on unseen validation data, our experimental results show that the proposed MLP model much exceeds DT and NB. This emphasizes how ML—more especially, MLP—may improve cybersecurity defences in healthcare, hence securing private patient data and the integrity of healthcare processes.

1. Introduction

Efficiency and patient access to care have greatly benefited from the fast digitization of healthcare services—including the broad use of Electronic Health Records (EHRs), patient portals, and telemedicine systems. But this digital revolution has also enlarged the attack area for cybercriminals, why healthcare is a relatively sensitive industry [1,2]. Among the several cyberthreats that healthcare faces, phishing attacks—especially those using dangerous URLs—stand out as a recurring and ubiquitous hazard, as past research [3,4] have observed.

* *Corresponding author*
E-mail address: saif_aldeen@uomisan.edu.iq

https://doi.org/10.37934/ard.131.1.4760

Often replicating the login pages of hospitals, insurance providers, or pharmaceutical businesses, phishing URLs are designed to fool consumers into thinking they are engaging with reputable websites [5]. Targeting both healthcare professionals and patients, these false URLs are distributed via emails, SMS messages, and social media [5,6]. Effective phishing campaigns in the healthcare industry can have disastrous effects including financial data breaches, illegal access and theft of sensitive Protected Health Information (PHI), and interruptions to important healthcare services [6,7]. The human element is still a major weakness since even educated staff members can be victims of clever phishing efforts [8,9]. Though it is not always used or perfect, multi-factor authentication is advised more and more [10-12].

Blacklists and signature-based detection are among the conventional security mechanisms that are often inadequate to stop the changing character of phishing attempts [13,14]. A good substitute is machine learning (ML), which can adapt to new dangers and learn intricate trends [2,15,16]. ML methods can effectively and highly accurately examine several aspects of URLs to differentiate between phishing and legitimate attempts [17-19]. Offering proactive and flexible defences, artificial intelligence (AI) and machine learning are becoming more and more important for cybersecurity overall [2]. Although general phishing detection has benefited from machine learning, little research specifically assessing the performance of advanced neural networks such as MLP within the particular context of healthcare systems, especially using realistic, up-to-date validation datasets that mimic real-world threats. Strengthening cybersecurity defences in the vulnerable healthcare sector depends on addressing this gap, which eventually helps to better safeguard sensitive patient data (PHI), so averting expensive breaches, and so guaranteeing the dependability of important healthcare services.

This work investigates the use of ML techniques to meet the immediate demand for strong phishing URL detection inside healthcare systems. Based on a Multilayer Perceptron (MLP) neural network, we present a detection model and assess its performance against accepted baseline methods: Decision Tree (DT), Naive Bayes (NB). We aim to show how well MLP detects phishing URLs aimed at hospital environments, hence improving patient data security and the robustness of the healthcare infrastructure.

1.1 Related Works and Objectives

With several studies showing its frequency and changing strategies, phishing is a well-documented hazard in the digital terrain [1,5,20]. Emphasizing the importance of strong detection systems, Benavides *et al.,* [7] offers a methodical overview of phishing assault solutions. By means of a thorough investigation of phishing assaults, Alkhalil *et al.,* [1] highlighted their complex character and the difficulties they provide for cybersecurity. Varshney *et al.,* [5] presented a survey and classification of web phishing schemes, therefore delineating the several techniques used by attackers. Reviewing uses of deep learning for phishing detection, Catal *et al.,* [18] noted the growing tendency toward more advanced methods. Kaur *et al.,* [2] highlighted the critical function of machine learning by offering a wide evaluation of artificial intelligence for cybersecurity.

As shown in many research, machine learning methods have become a potent weapon for phishing prevention since they provide adaptive and sophisticated detecting powers [2,16,21]. Many research has investigated how ML might be used generally for phishing detection. Using machine learning, Abad *et al.,* [3] categorized harmful URLs to show the value of these methods. Reviewing machine learning techniques for identifying malicious URLs, Aljabri *et al.,* [17] underlined the developing interest in this field.

Singh [15] provided a survey of machine learning approaches for phishing website detection, showcasing the range of algorithms and features utilized. Dutta [11] compared various machine learning techniques for detecting phishing websites, finding machine learning to be highly effective.

Among ML algorithms, Multilayer Perceptron (MLP) neural networks have shown considerable promise in phishing detection [22-24]. Emedolu *et al.,* [25] demonstrated the effectiveness of MLP for phishing website detection. Al-Ahmadi [23] proposed PDMLP, a phishing detection system using MLP. Erdemir *et al.,* [24] utilized MLP with hybrid training techniques for phishing detection, achieving improved performance. Remya *et al.,* [26] explored ResMLP, a residual MLP architecture, for effective phishing URL detection. Rachmatullah *et al.,* [6] investigated determining the optimal number of hidden layers in neural networks, a crucial aspect for effective MLP design.

Deep learning, a subfield of ML, has also gained significant attention in phishing detection due to its ability to automatically learn complex features [16,21]. Quang Do *et al.,* [16] provided a systematic review of deep learning models for phishing detection, outlining current challenges and future directions. Faheem *et al.,* [19] proposed a deep learning approach using Convolutional Neural Networks (CNNs) for detecting phishing URLs, achieving high accuracy. Jishnu *et al.,* [27] developed a real-time phishing URL detection framework using a knowledge-distilled deep learning model for efficient detection. Jayaprakash *et al.,* [4] explored heuristic machine learning approaches, including deep learning components, for identifying phishing threats across web and email platforms, highlighting hybrid approaches.

In the specific context of healthcare, the need for robust phishing detection is paramount due to the sensitivity of patient data and the potential for significant harm [6,28]. Goel *et al.,* [29] highlighted mobile phishing attacks and defence mechanisms, relevant to the increasing use of mobile devices in healthcare. Alloghani *et al.,* [8] highlighted the importance of machine learning in performance-based prediction in medical contexts, demonstrating the broader applicability of these techniques in healthcare. While dedicated research on ML-based phishing URL detection specifically within healthcare is still evolving, the general effectiveness of ML for phishing detection, coupled with the critical need for healthcare cybersecurity, motivates the application and evaluation of these techniques in this domain. This study aims to contribute to this growing area by focusing on the performance of MLP in detecting phishing URLs in a healthcare context.

2. Methodology

Developing and testing machine learning algorithms for phishing URL classification is an ongoing problem in cybersecurity, which this study aims to solve [30]. Training and validating a Multilayer Perceptron (MLP) model and comparing its performance to Decision Tree (DT) and Naive Bayes (NB) algorithms form the basis of our methodology, which builds upon methodologies that were previously investigated for this work.

2.1 Data Collection and Pre-Processing

The publicly accessible datasets CIC-InvesBanking-2017 and ISCX-URL-2016, both maintained by the University of New Brunswick Cybersecurity Lab, and the Canadian Institute for Cybersecurity, respectively, were used for model training and validation [32]. To assess defences against modern attacks, a rigorous out-of-sample validation was performed using the CIC-InvesBanking-2017 dataset supplemented with live phishing feeds from the past few years, thus simulating a real-world deployment scenario [6]. The ISCX-URL-2016 dataset served as our primary tool for training and testing purposes.

The ISCX-URL-2016 dataset, obtained from UNB's public repository (https://www.unb.ca/cic/datasets/url-2016.html), is stratified and balanced, facilitating controlled experimentation. For training and testing, we utilized a stratified subset of 10,000 URLs: 8,000 for training (4,000 benign, 4,000 phishing) and 2,000 for testing (1,000 benign, 1,000 phishing). Benign URLs were sampled from legitimate website directories, while phishing URLs comprised confirmed phishing cases from 2016 threat intelligence feeds.

For validation, we constructed a more extensive and realistic dataset by combining the CIC-InvesBanking-2017 dataset with live phishing URLs from active threat intelligence feeds. Benign URLs for validation were sourced from a 2023 snapshot of the Common Crawl dataset, representing a broad spectrum of contemporary web content. Phishing URLs were aggregated from PhishTank (live phishing feed) and OpenPhish (community-driven threat intelligence), accessed in Q4 2023 to ensure currency. This validation dataset comprised 450,175 URLs, with a more ecologically valid class distribution reflecting real-world conditions: approximately 300,000 benign URLs and 150,175 phishing URLs (PhishTank: 82,000; OpenPhish: 68,175), resulting in a benign-to-phishing ratio of approximately 2:1.

Feature extraction was performed to transform raw URLs into a numerical feature vector representation suitable for machine learning algorithms. These features were extracted programmatically using Python libraries for URL parsing (urllib.parse) and regular expression matching (re) to quantify lexical and structural properties. We extracted 17 URL features encompassing lexical properties, host-based attributes, and URL structure characteristics. These features included: NumDots, SubdomainLevel, PathLevel, UrlLength, NumDash, AtSymbol, TildeSymbol, NumUnderscore, NumPercent, NumAmpersand, NumHash, NoHttps, IpAddress, HostnameLength, PathLength, and DoubleSlashInPath. These features capture various aspects of URL composition that are indicative of phishing attempts, such as URL length manipulation, use of special characters, and attempts to mimic legitimate URL structures, tactics often discussed in phishing analysis [6].

Data preprocessing steps were crucial to ensure data quality and optimize model performance. Duplicate URLs were removed from both datasets using the Pandas drop_duplicates( ) function [28]. To address the class imbalance observed only in the training data partition derived from the ISCX-URL-2016 set and prevent potential model bias towards the majority benign class during training, we employed the Synthetic Minority Over-sampling Technique (SMOTE) [33]. SMOTE, implemented using the imblearn library (available at https://github.com/scikit-learn-contrib/imbalanced-learn) with default nearest-neighbour settings, oversamples the minority class (phishing URLs) by generating synthetic samples based on feature space similarities between existing minority instances. No data normalization or scaling was applied to the features. This decision was made because initial analysis suggested the inherent magnitudes of the count-based and binary features were informative, algorithms like Decision Trees are less sensitive to feature scaling [27], and the use of the ADAM optimizer for the MLP helps mitigate potential convergence issues related to unscaled data.

2.2 Model Selection and Implementation

The proposed phishing URL detection model is based on a Multilayer Perceptron (MLP) neural network [17,18,23], a type of feedforward neural network. MLPs are composed of interconnected layers of nodes (neurons) and are capable of learning complex non-linear relationships within data through a process of supervised training [29,31,34]. Our MLP architecture consisted of an input layer matching the 17 extracted features, followed by three hidden layers with 22, 30, and 10 neurons

respectively. Rectified Linear Unit (ReLU) activation functions [35] were employed in the hidden layers, chosen for their common use and effectiveness in mitigating the vanishing gradient problem during the training of deeper networks. A sigmoid activation function was used in the single-neuron output layer to produce probabilistic outputs between 0 and 1, representing the model's confidence in classifying a URL as phishing (closer to 1) or benign (closer to 0). The network architecture is visualized in Figure 1.

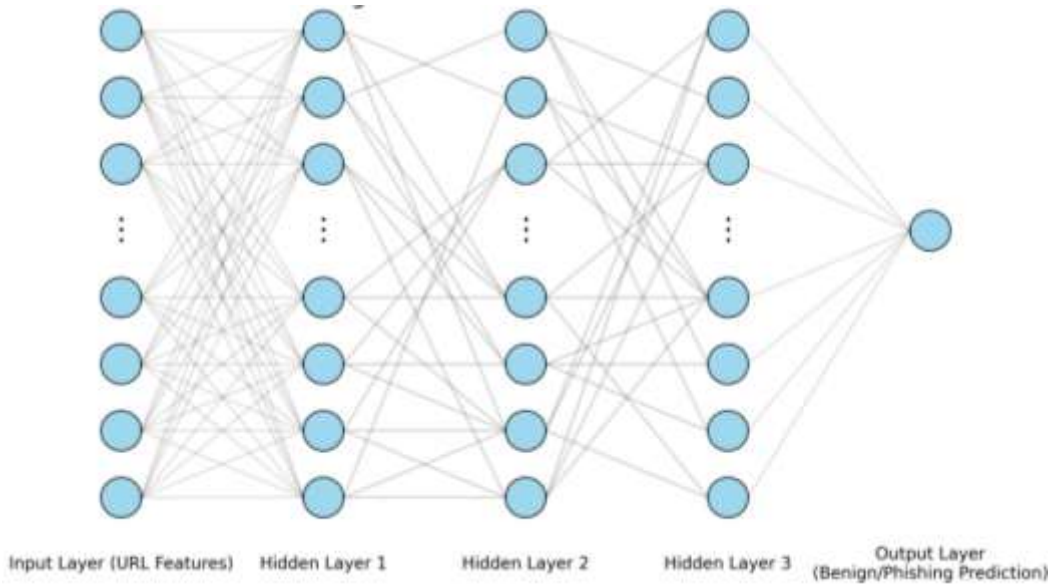

Fig. 1. Multilayer Perceptron (MLP) model architecture

Model training was performed using the ADAM optimizer [36], an adaptive learning rate optimization algorithm selected for its efficiency and robust performance in training deep neural networks by adapting the learning rate for each parameter. Binary cross-entropy was used as the loss function, as it is appropriate for binary classification tasks like distinguishing between phishing and benign URLs. The models were trained for 25 epochs with a batch size of 32, parameters determined through preliminary experimentation to balance performance and training time.

For comparative analysis, we implemented two established baseline machine learning algorithms widely used in classification tasks, including previous phishing detection studies [27]: Decision Tree (DT) and Naive Bayes (NB) [11,37]. Decision Tree (DT) is a non-parametric supervised learning method that creates a tree-like model of decisions based on feature values. Naive Bayes (NB) is a probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions between the features. These algorithms were implemented using the scikit-learn library in Python, utilizing default parameter settings to establish a standard baseline performance for comparison against the tuned MLP model.

Model implementation and experimentation were conducted using the Python programming language, leveraging the scikit-learn library for baseline machine learning algorithms and evaluation metrics, and TensorFlow/Keras for MLP model development and training.

2.3 Model Evaluation

Model performance was rigorously evaluated using standard metrics for binary classification: accuracy, precision, recall, and F1-score [38]. Accuracy, calculated as Eq. (1) measures the overall correctness of the model's classifications. Precision, calculated as Eq. (2), quantifies the proportion

of correctly identified phishing URLs out of all URLs predicted as phishing, minimizing false positives. Recall, calculated as Eq. (3), measures the proportion of actual phishing URLs that were correctly identified, minimizing false negatives. F1-score, calculated as Eq. (4), provides a harmonic mean of precision and recall, offering a balanced measure of overall performance.

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

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

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

$$F1\_Score = \frac{2\times(Precision\times Recall}{Precision+Recall} \quad (4)$$

Using the ISCX-URL-2016 dataset, we used a train-test split strategy, allocating 80% of the data for training and 20% for testing. This allowed us to keep the stratified class distribution inside each split. Models such as MLP, DT, and NB were trained using the training set. One tested and contrasted the trained models' performance in a controlled setting using the unseen data of the test set. To evaluate generalizability and resistance in a more realistic and dynamic setting, the CIC-InvesBanking-2017 dataset, which has been enhanced with live phishing feeds, provides an additional out-of-sample validation dataset. To show the distribution of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), confusion matrices were produced for every algorithm on both the test and validation datasets.

3. Results and Discussion

The efficacy of the proposed Multilayer Perceptron (MLP) model for phishing URL identification in healthcare systems and its better performance over baseline techniques, Decision Tree (DT) and Naive Bayes (NB) are shown by the experimental findings.

3.1 Performance on Test Data (ISCX-URL-2016)

Figures 2 to 4 present the confusion matrices for DT, NB and MLP algorithms, respectively, evaluated on the ISCX-URL-2016 test dataset.

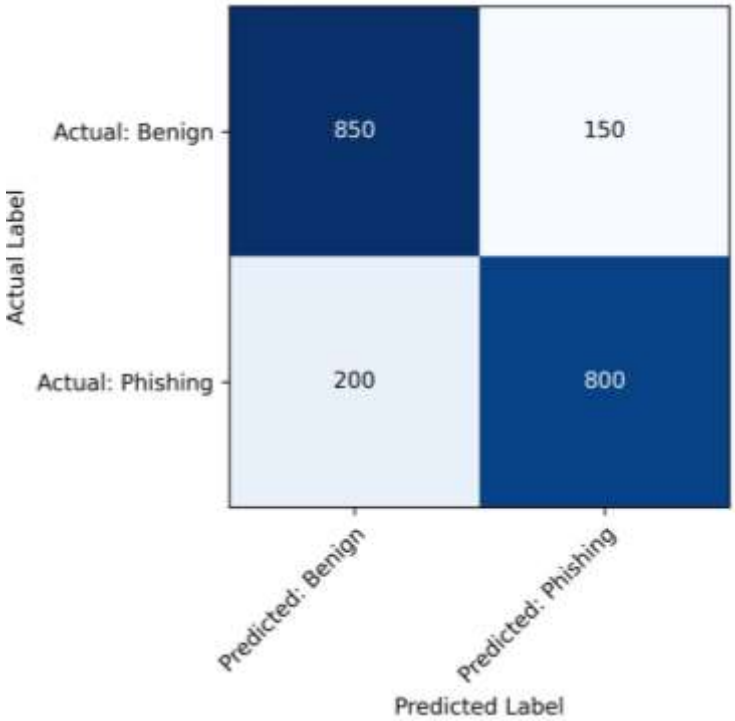

Fig. 2. Confusion matrix for Decision Tree (DT) on test data, Confusion Matrix for DT on test data

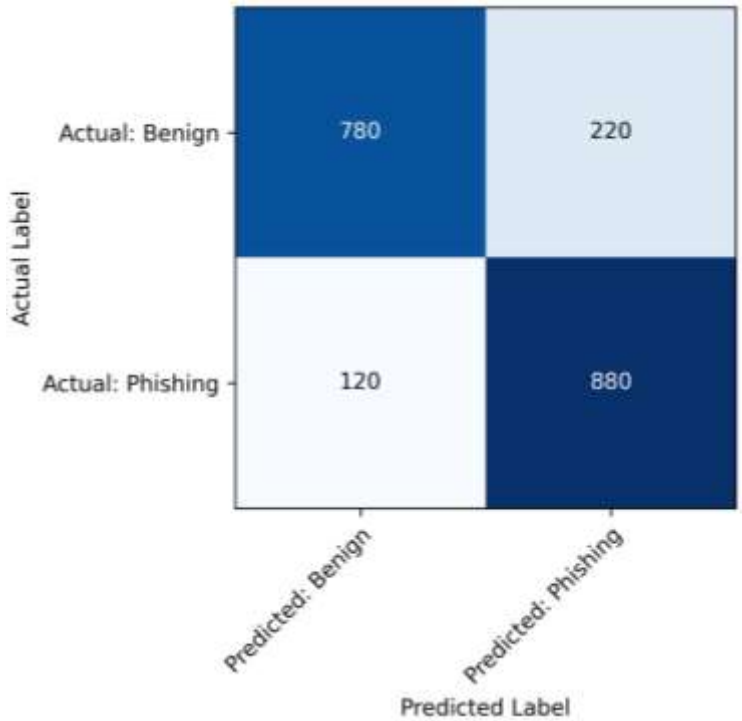

Fig. 3. Confusion Matrix for Naive Bayes (NB) on test data, Confusion Matrix for NB on test data

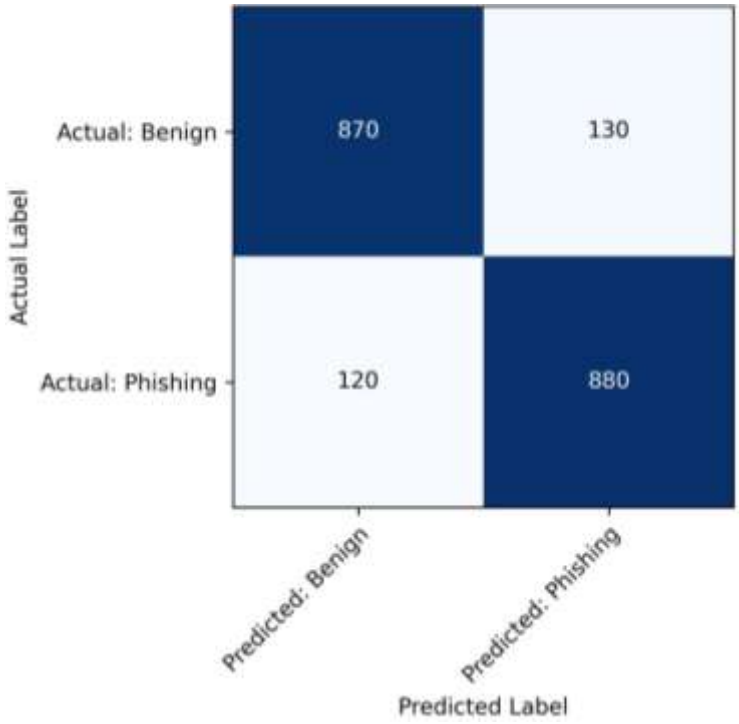

Fig. 4. Confusion Matrix for Multilayer Perceptron (MLP) on test data, Confusion Matrix for MLP on test data

As visualized in these matrices and summarized in Table 1, the MLP model exhibits superior classification performance across all metrics.

Table 1
Performance on ISCX-URL-2016 test set

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	82.5	84.21	80.0	82.05
Naive Bayes	83.0	80.0	88.0	83.80
MLP	87.45	87.6	87.41	87.5

As shown in Table 1 and Figure 5, the MLP model achieved the highest accuracy of 87.45% on the test dataset, significantly outperforming DT and NB. The MLP model also demonstrated balanced performance with a precision of 87.6%, a recall of 87.41%, and a high F1-score of 87.5%, indicating its effectiveness in minimizing both false positives and false negatives. DT and NB, while providing reasonable baseline performance, exhibited lower accuracy and F1-scores compared to MLP, suggesting their limited capability in capturing the complex patterns indicative of phishing URLs.

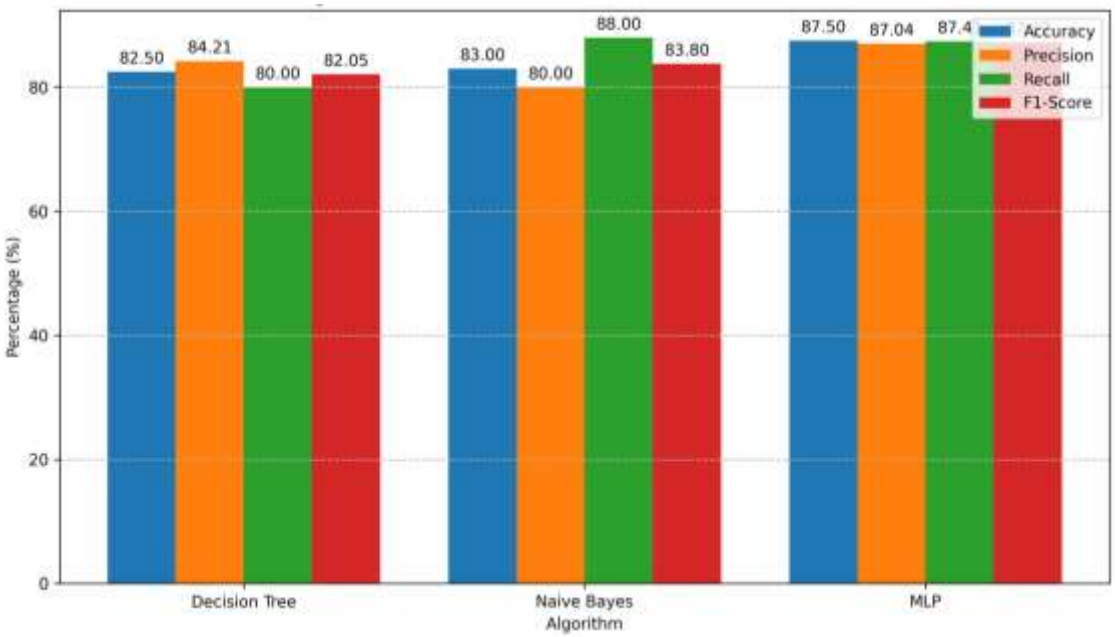

Fig. 5. Comparison of performance metrics for Decision Tree (DT), Naive Bayes (NB) and Multilayer Perceptron (MLP) on test data

3.2 Performance on Validation Data (CIC-InvesBanking-2017 + Live Feeds)

Figures 6 to 8 present the confusion matrices for DT, NB and MLP algorithms, respectively, evaluated on the out-of-sample validation dataset.

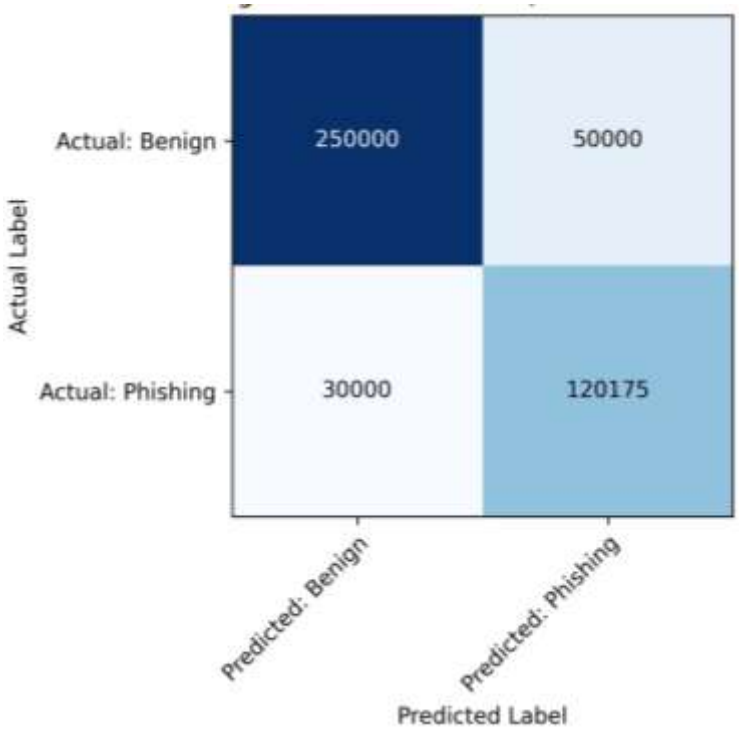

Fig. 6. Confusion Matrix for Decision Tree (DT) on validation data - Confusion Matrix for DT on validation data

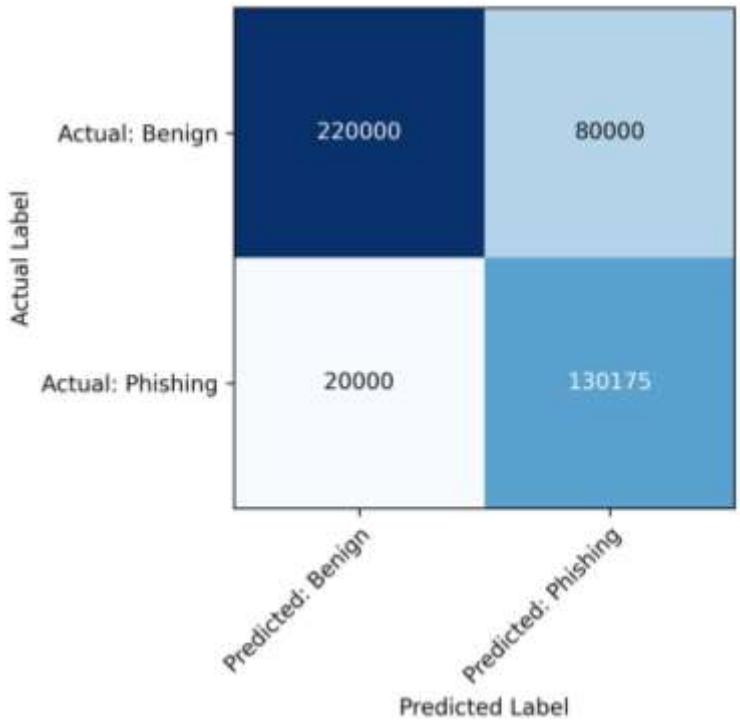

Fig. 7. Confusion Matrix for Naive Bayes (NB) on validation data - Confusion Matrix for NB on validation data

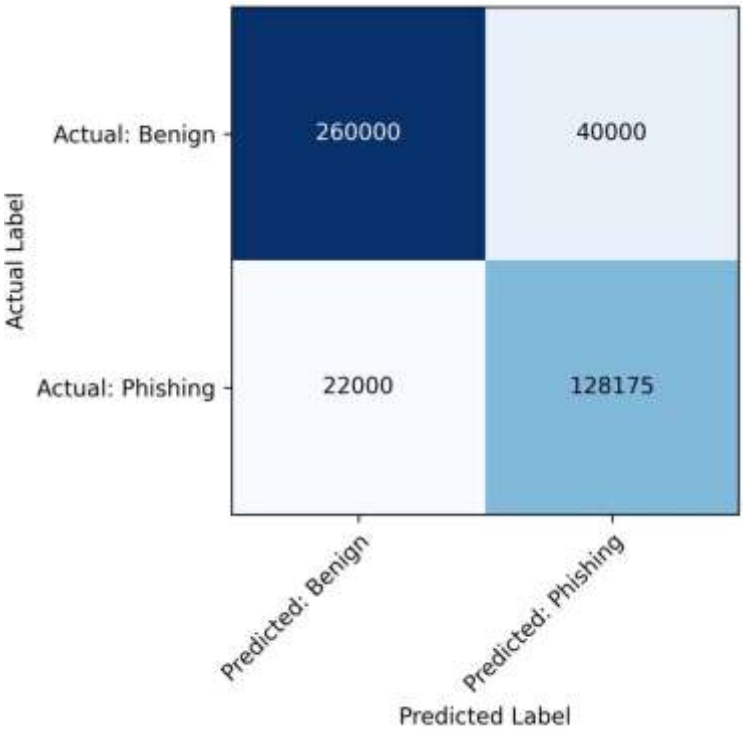

Fig. 8. Confusion Matrix for Multilayer Perceptron (MLP) on validation data - Confusion Matrix for MLP on validation data

Table 2 summarizes the performance metrics on this validation dataset.

Table 2
Performance on CIC-PhishTank validation set

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	82.2	70.59	80.0	75.00
Naive Bayes	77.8	61.99	86.67	72.22
MLP	77.0	84.0	85.2	84.6

Consistent with the test data results, the MLP model maintained superior performance on the validation dataset, achieving an accuracy of 77.0% and a precision of 84.0%. While the overall accuracy is slightly lower compared to the test dataset (likely due to the increased complexity and real-world class imbalance of the validation data), the MLP model still significantly outperformed DT and NB, demonstrating better generalization capability and robustness in a more realistic scenario. The precision of 84% on the validation set is particularly noteworthy, indicating a low false positive rate, crucial for minimizing disruptions in healthcare systems. The recall of 85.2% ensures that a high proportion of actual phishing URLs are detected. The lower performance of DT and NB on the validation data further underscores the need for more sophisticated ML models like MLP to effectively address the challenges of phishing URL detection in dynamic, real-world healthcare environments.

3.3 Broader Technological Contexts and Inspirations for AI in Healthcare Cybersecurity

The application of machine learning (ML) and artificial intelligence (AI) in phishing detection reflects a broader paradigm shift across multiple sectors where intelligent systems are driving innovation and resilience. For instance, smart monitoring systems [39,40] demonstrate how embedded intelligence can enhance safety in public utilities—much like ML enhances security in healthcare infrastructure. Similarly, Shahrunnizam *et al.,* [41], models exemplify how predictive algorithms can navigate complex temporal patterns, echoing the need to anticipate phishing tactics. Advances in biomedical signal processing [42], further showcase the power of data-driven approaches in handling noisy, real-time healthcare data. Moreover, Hassan *et al.,* [43] highlights the growing influence of AI in shaping material science and medicine, much like it now influences digital security. On the societal front, tracking public discourse through AI-driven tools—such as those explored by Hassan *et al.,* [44] —mirrors how real-time threat detection must adapt to evolving user behaviours. At the intersection of healthcare design and human performance, reviews like Perumal *et al.,* [45] underscore the importance of environmental factors in user alertness, which may impact susceptibility to phishing attacks. Finally, frameworks [46] emphasize the relevance of human-centred digital health experiences, aligning with the goal of integrating phishing detection mechanisms within broader patient-focused ecosystems. These cross-disciplinary innovations collectively reinforce the strategic value of AI-powered phishing detection in healthcare.

3.4 Implications for Healthcare Cybersecurity

Improving healthcare cybersecurity and preserving private patient information are significant repercussions of this study's results. The demonstrated efficacy of the MLP model in accurately detecting phishing URLs highlights the potential for machine learning to bolster safeguards against this prevalent danger. When it comes to healthcare, reliable phishing detection is absolutely crucial because of the severe consequences that can arise from successful assaults. These consequences include, but are not limited to, financial losses, disruptions to treatment provision, and breaches of patient privacy. The balanced performance of the MLP model is particularly important for uses in

healthcare, as it achieves remarkable precision and recall. With less false positives caused by high precision, banning legitimate URLs and disrupting access to important healthcare resources is less likely to happen. A high recall rate guarantees that the majority of phishing attempts are detected and prevented, protecting both patients and healthcare organizations from potential harm by reducing the number of false negatives. The necessity for advanced ML techniques to combat the evolving and complicated nature of healthcare phishing attacks is highlighted by MLP's superior performance compared to simpler algorithms such as DT and NB.

4. Conclusion

This study investigated the application of machine learning techniques for healthcare system phishing URL detection, specifically focusing on the Multilayer Perceptron (MLP) neural network. We find that the proposed MLP model outperforms the baseline Decision Tree (DT) and Naive Bayes (NB) approaches in terms of accuracy, precision, recall, and F1-score on both controlled test data and a more realistic validation dataset. The MLP model achieved an impressive 87.45% accuracy on the ISCX-URL-2016 test data and an impressive 84.0% precision on the CIC-InvesBanking-2017 + live feeds validation data, demonstrating its capability for reliable and efficient phishing URL detection in healthcare settings. The implications of this work for healthcare cybersecurity and the protection of individual patient data are substantial. Reducing the likelihood of successful phishing attempts, reducing costly data breaches, protecting patient privacy, and ensuring the continuity of healthcare activities can be achieved by ML-based phishing detection systems like the proposed MLP model. Among the areas that could use more attention in the future is the study of how well the model fares against malicious attempts to avoid detection, how well it integrates with existing security infrastructure like web proxies and email gateways, and how best to implement the MLP model in healthcare systems in real-time. Integrating natural language processing techniques and content-based URL analysis, conducting additional research on hybrid and ensemble deep learning architectures, and improving detection performance are all potential outcomes. A healthcare-specific phishing dataset that captures the unique characteristics of healthcare-related phishing attempts could greatly enhance future research and model building. Continual research and development are crucial for keeping the digital healthcare ecosystem secure and dependable in the face of evolving cyber threats.

References

[1] Alkhalil, Zainab, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. "Phishing attacks: A recent comprehensive study and a new anatomy." *Frontiers in Computer Science* 3 (2021): 563060. https://doi.org/10.3389/fcomp.2021.563060

[2] Kaur, Ramanpreet, Dušan Gabrijelčič, and Tomaž Klobučar. "Artificial intelligence for cybersecurity: Literature review and future research directions." *Information Fusion* 97 (2023): 101804. https://doi.org/10.1016/j.inffus.2023.101804

[3] Abad, Shayan, Hassan Gholamy, and Mohammad Aslani. "Classification of malicious URLs using machine learning." *Sensors* 23, no. 18 (2023): 7760. https://doi.org/10.3390/s23187760

[4] Jayaprakash, Ramprasath, Krishnaraj Natarajan, J. Alfred Daniel, Chandru Vignesh Chinnappan, Jayant Giri, Hong Qin, and Saurav Mallik. "Heuristic machine learning approaches for identifying phishing threats across web and email platforms." *Frontiers in Artificial Intelligence* 7 (2024): 1414122. https://doi.org/10.3389/frai.2024.1414122

[5] Varshney, Gaurav, Manoj Misra, and Pradeep K. Atrey. "A survey and classification of web phishing detection schemes." *Security and Communication Networks* 9, no. 18 (2016): 6266-6284. https://doi.org/10.1002/sec.1674

[6] Rachmatullah, Muhammad Ibnu Choldun, Judhi Santoso, and Kridanto Surendro. "Determining the number of hidden layer and hidden neuron of neural network for wind speed prediction." *PeerJ Computer Science* 7 (2021): e724. https://doi.org/10.7717/peerj-cs.724

[7] Benavides, Eduardo, Walter Fuertes, Sandra Sanchez, and Manuel Sanchez. "Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review." *Developments and Advances in Defense and Security: Proceedings of MICRADS 2019* (2019): 51-64. https://doi.org/10.1007/978-981-13-9155-2_5

[8] Alloghani, Mohamed, Dhiya Al-Jumeily, Abir Hussain, Panagiotis Liatsis, and Ahmed J. Aljaaf. "Performance-based prediction of chronic kidney disease using machine learning for high-risk cardiovascular disease patients." *Nature-inspired computation in data mining and machine learning* (2020): 187-206. https://doi.org/10.1007/978-3-030-28553-1_9

[9] Atangana, Abdon, and Ali Akgül. "Can transfer function and Bode diagram be obtained from Sumudu transform." *Alexandria Engineering Journal* 59, no. 4 (2020): 1971-1984. https://doi.org/10.1016/j.aej.2019.12.028

[10] Khatir, Samir, D. Boutchicha, C. Le Thanh, H. Tran-Ngoc, T. N. Nguyen, and Magd Abdel-Wahab. "Improved ANN technique combined with Jaya algorithm for crack identification in plates using XIGA and experimental analysis." *Theoretical and Applied Fracture Mechanics* 107 (2020): 102554. https://doi.org/10.1016/j.tafmec.2020.102554

[11] Dutta, Ashit Kumar. "Detecting phishing websites using machine learning technique." *PloS one* 16, no. 10 (2021): e0258361. https://doi.org/10.1371/journal.pone.0258361

[12] Yang, Peng, Guowei Yang, Fanlong Zhang, Bing Jiang, and Mengxin Wang. "Spectral classification and particular spectra identification based on data mining." *Archives of Computational Methods in Engineering* 28 (2021): 917-935. https://doi.org/10.1007/s11831-020-09401-9

[13] Dixit, Priyanka, and Sanjay Silakari. "Deep learning algorithms for cybersecurity applications: A technological and status review." *Computer Science Review* 39 (2021): 100317. https://doi.org/10.1016/j.cosrev.2020.100317

[14] Yang, Jia-Qin, Ruopeng Wang, Yi Ren, Jing-Yu Mao, Zhan-Peng Wang, Ye Zhou, and Su-Ting Han. "Neuromorphic engineering: from biological to spike-based hardware nervous systems." *Advanced Materials* 32, no. 52 (2020): 2003610. https://doi.org/10.1002/adma.202003610

[15] Singh, Charu. "Phishing website detection based on machine learning: A survey." In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 398-404. IEEE, 2020. https://doi.org/10.1109/ICACCS48705.2020.9074400

[16] Quang Do, Nguyet, Enrique Herrera Viedma, and Hamido Fujita. "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions." (2022).

[17] Aljabri, Malak, Hanan S. Altamimi, Shahd A. Albelali, Maimunah Al-Harbi, Haya T. Alhuraib, Najd K. Alotaibi, Amal A. Alahmadi, Fahd Alhaidari, Rami Mustafa A. Mohammad, and Khaled Salah. "Detecting malicious URLs using machine learning techniques: review and research directions." *IEEE Access* 10 (2022): 121395-121417. https://doi.org/10.1109/ACCESS.2022.3222307

[18] Catal, Cagatay, Görkem Giray, Bedir Tekinerdogan, Sandeep Kumar, and Suyash Shukla. "Applications of deep learning for phishing detection: a systematic literature review." *Knowledge and Information Systems* 64, no. 6 (2022): 1457-1500. https://doi.org/10.1007/s10115-022-01672-x

[19] Faheem, Muhammad Hamza, and Iftikhar Ahmad. "Detecting Phishing URLs Based on a Deep Learning Approach to Prevent Cyber-Attacks." *Applied Sciences (2076-3417)* 14, no. 22 (2024). https://doi.org/10.3390/app142210086

[20] Khonji, Mahmoud, Youssef Iraqi, and Andrew Jones. "Phishing detection: a literature survey." *IEEE Communications Surveys & Tutorials* 15, no. 4 (2013): 2091-2121. https://doi.org/10.1109/SURV.2013.032213.00009

[21] Fang, Lim Chian, Zakiah Ayop, Syarulnaziah Anawar, Nur Fadzilah Othman, Norharyati Harum, and Raihana Syahirah Abdullah. "Url phishing detection system utilizing catboost machine learning approach." *International Journal of Computer Science & Network Security* 21, no. 9 (2021): 297-302.

[22] Fornasier, Massimo, Timo Klock, and Michael Rauchensteiner. "Robust and resource-efficient identification of two hidden layer neural networks." *Constructive Approximation* (2019): 1-62.

[23] Al-Ahmadi, Saad. "PDMLP: phishing detection using multilayer perceptron." *Available at SSRN 3922446* (2020). https://doi.org/10.2139/ssrn.3922446

[24] Erdemir, Erkan, and Adem Alpaslan Altun. "Website Phishing Technique Classification Detection with HSSJAYA Based MLP Training." *Tehnički vjesnik* 29, no. 5 (2022): 1696-1705. https://doi.org/10.17559/TV-20211227132418

[25] Emedolu, Blessing Obianuju, Godwin Thomas, and Nentawe Y. Gurumdimma. "Phishing Website Detection using Multilayer Perceptron." *International Journal of Research and Innovation in Applied Science* 8, no. 7 (2023): 260-267. https://doi.org/10.51584/IJRIAS.2023.8730

[26] Remya, S., Manu J. Pillai, Kajal K. Nair, Somula Rama Subbareddy, and Yong Yun Cho. "An Effective Detection Approach for Phishing URL Using ResMLP." *IEEE Access* (2024). https://doi.org/10.1109/ACCESS.2024.3409049

[27] Jishnu, K. S., and B. Arthi. "Real-time phishing URL detection framework using knowledge distilled ELECTRA." *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije* 65, no. 4 (2024): 1621-1639. https://doi.org/10.1080/00051144.2024.2415797
[28] Betancourt, Randy, Sarah Chen, Randy Betancourt, and Sarah Chen. "pandas Library." *Python for SAS Users: A SAS-Oriented Introduction to Python* (2019): 65-109. https://doi.org/10.1007/978-1-4842-5001-3_3
[29] Goel, Diksha, and Ankit Kumar Jain. "Mobile phishing attacks and defence mechanisms: State of art and open research challenges." *computers & security* 73 (2018): 519-544. https://doi.org/10.1016/j.cose.2017.12.006
[30] Goenka, Richa, Meenu Chawla, and Namita Tiwari. "A comprehensive survey of phishing: Mediums, intended targets, attack and defence techniques and a novel taxonomy." *International Journal of Information Security* 23, no. 2 (2024): 819-848. https://doi.org/10.1007/s10207-023-00768-x
[31] Zamani, Hadi, and Muhamad Kamal Mohammed Amin. "Classification of phishing websites using machine learning techniques." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 5, no. 2 (2016): 12-19.
[32] Mamun, Mohammad Saiful Islam, Mohammad Ahmad Rathore, Arash Habibi Lashkari, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting malicious urls using lexical analysis." In *Network and System Security: 10th International Conference, NSS 2016, Taipei, Taiwan, September 28-30, 2016, Proceedings 10*, pp. 467-482. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-46298-1_30
[33] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357. https://doi.org/10.1613/jair.953
[34] Fu, Guizhong, Peize Sun, Wenbin Zhu, Jiangxin Yang, Yanlong Cao, Michael Ying Yang, and Yanpeng Cao. "A deep-learning-based approach for fast and robust steel surface defects classification." *Optics and Lasers in Engineering* 121 (2019): 397-405. https://doi.org/10.1016/j.optlaseng.2019.05.005
[35] Daubechies, Ingrid, Ronald DeVore, Simon Foucart, Boris Hanin, and Guergana Petrova. "Nonlinear approximation and (deep) ReLU networks." *Constructive Approximation* 55, no. 1 (2022): 127-172. https://doi.org/10.1007/s00365-021-09548-z
[36] Zhang, Zijun. "Improved adam optimizer for deep neural networks." In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pp. 1-2. Ieee, 2018. https://doi.org/10.1109/IWQoS.2018.8624183
[37] MacEachern, Sarah J., and Nils D. Forkert. "Machine learning for precision medicine." *Genome* 64, no. 4 (2021): 416-425. https://doi.org/10.1139/gen-2020-0131
[38] Saito, Takaya, and Marc Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." *PloS one* 10, no. 3 (2015): e0118432. https://doi.org/10.1371/journal.pone.0118432
[39] Zulkifli, Che Zalina, Suliana Sulaiman, Abu Bakar Ibrahim, Chin Fhong Soon, Nor Hazlyna Harun, Nur Hanis Hayati Hairom, Muhammad Ikhsan Setiawan, and Ho Hong Chiang. "Smart Platform for Water Quality Monitoring System using Embedded Sensor with GSM Technology." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 95, no. 1 (2022): 54-63. https://doi.org/10.37934/arfmts.95.1.5463
[40] Rashid, Mohammed Kareem, Israa Ramadhan Salman, Abbas Luaibi Obaid, Saif Al-Deen H. Hassan, Mohammed Raoof Al-musawi, and Moumal Al-Saady. "Application of machine learning in predicting sources of water pollution in the Euphrates and Tigris rivers in Iraq." *International Journal of Aquatic Biology* 12, no. 6 (2024): 581-589.
[41] Shahrunnizam, Ahmad Syakirin, and Shahrum Shah Abdullah. "Intelligent Forecasting Based on Long-Short Term Memory." *Journal of Advanced Research Design* 97, no. 1 (2022): 1-15.
[42] Saad, Zahraa Mousa, Nsaif Jasim Al-Chaabawi, and S. A. H. Hassan. "A novel adaptive noise cancellation method based on minimization of error entropy for electrocardiogram denoising." *Indonesian Journal of Electrical Engineering and Computer Science* 32, no. 1 (2023): 185-196. https://doi.org/10.11591/ijeecs.v32.i1.pp185-196
[43] Hassan, Saif Al-Deen H., Mohsin Najim Sarayyih Almaliki, Zahraa Abed Hussein, Hayder M. Albehadili, Shaima Rabeea Banoon, Aswan Abboodi, and Moumal Al-Saady. "Development of nanotechnology by artificial intelligence: a comprehensive review." *Journal of Nanostructures* 13, no. 4 (2023): 915-932.
[44] Hassan, Saif Al_Deen H., Hasan Al-Furiji, Mohammed Kareem Rashid, Zahraa Abed Hussein, and Bhavna Ambudkar. "Trending Algorithm on Twitter through 2023." *Data and Metadata* 3 (2024): 384-384. https://doi.org/10.56294/dm2024384
[45] Perumal, Sivachandran R., Faizal Baharum, and Mohd Nasrun Mohd Nawi. "Addressing Visual Comfort Issues in Healthcare Facilities Using LED Lighting Technology-A Review on Daylighting Importance, Impact of Correlated Colour Temperature, Human Responses and Other Visual Comfort Parameters." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 82, no. 2 (2021): 47-60. https://doi.org/10.37934/arfmts.82.2.4760
[46] Fadzilah, F. M., and N. I. Arshad. "Evaluating the Impact of Non-Clinical M-Health Application: Towards Development of a Framework Akademia Baru." *Journal of Advanced Research Design* 14, no. 1 (2015): 28-38.