Available online at www.equijost.com

Equity Journal of Science and Technology, 2022 9(2): 5 - 11

ISSN 2354-1814; E-ISSN 2683-5961

EQUIJOST *An Official Publication of Kebbi State University of Science and Technology, Aliero, Nigeria*

# Multi-level security–base intrusion detection scheme for mitigating phishing attack in cloud computing environment

*Haruna A. Enefola, Alimi O. Maruf, Danlami Gabi
Department of Computer Science, Kebbi State University of Science and Technology, Aliero, Nigeria
*Corresponding Author's email: aminu4friends@gmail.com

**Abstract**

The majority of computing devices in the 21st century is deployed in the cloud, providing access to information and online services. However, phishing attacks continue to pose a risk to the confidentiality, integrity, and availability of information, exposing legitimate users to potential threats. Thus, despite previous efforts to combat phishing attacks and improve detection models, there is still room for further performance improvement. To address this issue, the study proposed a multi-level phishing attack detection scheme for the cloud environment. The experimental results via simulation show our proposed multi-level model for phishing detection can achieve 0.9997 each for accuracy and F1-measure, and 0.9998 for specificity and precision each, as well as 0.9996 and 0.0003 respectively. A comparison between the proposed approach and the benchmarked algorithm shows that the proposed approach performs significantly better. This indicates that the proposed model can be used by computing and application industries to prevent phishing attacks, and it demonstrates the effectiveness of integrating multiple models for research analysis and experimentation in the academic domain.

**Keywords:** Precision, Cloud Computing, Accuracy, False Positive Rate, Recall

## 1. Introduction

Cloud Computing is a new computing paradigm through which services are provided to customers using pays-as-you-go model without customer purchasing the underlying physical Infrastructure [1]. The paradigm has been rapidly developed along with the trend of IT services. It is efficient and cost economical for cloud consumers to use computing resources as much as they desired from the Cloud providers [2]. Services offered by cloud computing include 1) Infrastructure as a Service (IaaS), where the user has control over complete virtual machines [3], such as server, storage, network and virtualization. 2) Platform as a Service (PaaS), where the user can deploy user-created applications in cloud if the provider supports the languages, APIs, and tools used for creating application [3] like Google App Engine, and Microsoft's Azure. 3) Software as a Service (SaaS) which enables users to execute provider's applications [3]. All these services are provided via the internet to enable uninterrupted availability. This research chose SaaS layer (where customers build or interact with software applications) of the computing paradigm to carry out its research, based on the fact of the layer's importance and at the same time its vulnerability to phishing attacks, involvement of multiple stakeholders (cloud service providers, application developers, and end-users), and criticality for business operations while providing access to essential software applications. However, there are several challenges that were of great concerns to SaaS layer of the cloud computing environment. These include, scheduling, control, and fault tolerance, quality of service and security [4]. This research addresses the concern of security as one of the major challenges at the SaaS of the cloud paradigm with focus on phishing attacks as a common security threat.

Phishing is the act of attempting to acquire sensitive information such as usernames, passwords, and credit card details by masquerading as a trustworthy entity in an electronic communication [5]. Phishing attacks use an e-mail messages and websites designed to look as if they came from known and legitimate organizations, in order to deceive people in giving out their personal, financial or other sensitive information. These growing menaces have posed lots of damages to cyberspace thereby causing valuable data and financial losses to both users and companies [6]. The United States Computer Emergency Readiness Team (CERT) gathered security details from various agencies, which stated that there were 107,655 incidents in 2014, with 43,889 information exploits from federal agencies [7].

The traditional network security measures like firewall are gradually been overpowered by hackers on the cloud computing environment due to their limitation and inability to stop many outsider attacks [8]. Therefore, several researchers [2, 9-11, 13] have proposed different forms of IDS to detect different forms of attacks in cloud computing environment.

Thus, researchers and industries worked to better understand and improve on techniques that can be employed to checkmate associated threats as it related to phishing attacks, similarly, the intrusion of cloud computing environment continues to be a serious challenge due to the inability of existing techniques to address the sophisticated cyber-attacks threats [14]. Despite all the strict security policies and instruments, more organization and individual are increasingly been faced with the threats of a wide range of phishing attacks against their systems security.

However, in view of the limitations of the existing techniques such as setback characterized by performance deficiency relating to poor accuracy and false positive rate [15-17], hence, the need for a model that can render performance enhancement in terms of accuracy as well as false positive rate for phishing detection is of significant. Therefore, this paper proposes a multilevel security-based intrusion detection scheme to mitigate the impact of phishing attack in cloud computing environment.

The contribution of this paper is as follow:

i. We deployed a framework for a multi-level security-based intrusion detection model for phishing attack in cloud computing environment with better performance over existing framework.
ii. We evaluate the performance of the designed model through relevant state of the art performance metrics.

## 2. Related Works

Modi *et al.,* [18] inferred that multiple hosts in the network can be secured from attackers by utilizing a few properly deployed NIDS. If run in stealth mode, the location of NIDS can be hidden from attacker. The NIDS is unable to perform analysis if traffic is encrypted [19], in cloud environment, the attacks on hypervisor or VMs are detected by positioning NIDS at the cloud server which interacts with external network, however, it cannot detect attacks inside a virtual network contained by hypervisor. Cloud provider is responsible for installing NIDS in cloud [18], the mechanism of detection is to compare the present behaviour to an already observed behaviour in real time. NDIS is positioned next to the firewall and the network, and in the cloud, it is placed in the cloud server that deals with external networks [20]. As an example of NIDS, the work proposed in [21] that is a signature-based IDS to detect the DDoS. The IDS is positioned in virtual switch to log the traffic and drop the packets that match a known signature. Another solution is Cloud Intrusion Detection Service (CIDS) [20], proposed using Snort that is a signature-based IDS that deployed at the network level. This solution is intended to work as on demand cloud service that any host desired to benefit from this service should be subscribe to the CIDS.

Balogun *et al.,* [15] proposed an Optimized Decision Forest (ODF) model to address the existing trending challenges associated with machine learning algorithms in terms of performance rating as well as the dynamics found in checkmating phishing attack detection, the proposed ODF was claimed to have achieved an optimal performance for phishing detection compared to existing techniques deployed to arrest phishing attack, the following performance score was recorded 98.37%, 0.999, 0.98, 0.967 and 0.016 respectively for accuracy, AUC, F-Measure MCC and False Positive Rate (FPR) respectively, nevertheless, performance can be improved upon to enhance detection of phishing attack.

A new ensemble model for phishing detection was developed to counter phishing attack, the model was further compared with other existing machine learning algorithms that have been used to analysis phishing detection, algorithms such as Random Forest (RF), Support Vector, Machine, Naive Bayes, JRip, C4.5, PART, and KNN, the proposed model was said to have outperformed these machine learning models in terms of accuracy for phishing detection with a score of 98.24, expect for RF which achieved an accuracy score of 98.36, however, evaluation of other relevant state of the art performance metric will give a clear view of performance evaluation [16].

Saravan and Subramanian [22] proposed a framework for detection of phishing websites based Genetic Algorithm (GA) and Adaptive Resonance Theory Mapping (ARTMAP), the model was acclaimed to have achieved a success performance rating that outperformed existing methods used in websites phishing attack detection, the following performance metric of accuracy, sensitivity, specificity, error rate, and detection time was recorded; 95.27, 92.78, 91.28, 4.73 and 4.93 respectively, however, based on contemporary research available, the performance of the model can be better improved.

The study carried out to address the challenge of optimality in terms of phishing attack prediction was experimented by [23], of which an optimal detection model that is hinged on a meta-heuristic population technique and Sine Cosine Algorithm (SCA) based K-NN was proposed, the proposed model was compared against Decision Tree (DT), and Naive Bayes, which was claimed to have outperformed both of the aforementioned models in term of accuracy, F-measure, TPR, FPR and Mean Absolute Error (MAE), with the following scores; 97.18, 0.97, 0.97, 0.03, and 0.0323 respectively, however, with less than 3000 instances, more instances for model train can breed a robust detection system.

Subasi and Kremic [24] presented an intelligent phishing detection framework to address the issues existing in phishing attack detection, accuracy, F-measure and Area under Receiver Operation Characteristic (ROC) was utilized to evaluate the strength of the system, more so, SVM, kNN, ANN, RF, CART, C4.5, RepTree, were further deployed forming and ensemble with Multiboosting and Adaboost for a comparative analysis, of which a combination of Adaboost and SVM was recorded to have outperformed the other machine learning algorithms, with an accuracy of 97.61%, 0.976 F-measure and 0.996 ROC Area, however, time complexity is at excess, meanwhile, performance can still be improved upon.

Khan *et al.*, [17] performed a comprehensive analysis of different selected machine models, with the aim to establish the most effective machine learning algorithms for phishing detection through performance metric evaluation, the employed machine learning algorithms are DT, SVM, RF, NB, kNN and ANN, it was established that RF outperformed other models in terms of accuracy, specificity, precision, recall and F1 score with 92.94%, 94.70%, 89.41%, 89.41% and 89.41% respectively, however, performance can be improved on if model is properly managed or tuned.

Moruff *et al.*, [25] in the research targeted at the comparative analysis of some selected machine learning algorithms to classify phishing URLs, evaluated the following models logistic regression, Gaussian naive bayes, DT and kNN, the result of the analysis revealed that DT achieved the optimal score of 1%, 0.99, 1, 0.99 for accuracy, precision, recall and F1-measure respectively, outperforming other models, however, it is clear that with the score achieved in accuracy and recall, it is an indication that research in this field is concluded in perfection and absolute, this in research is not acceptable, of which the challenge may have arisen from lack of proper dataset cleaning and pre-processing.

The proposed scheme is significant because it addresses the limitations of traditional security measures as seen in reviewed literature, majorly performance setback, also the proposed scheme provides a more comprehensive approach to detecting and mitigating phishing attacks in the cloud environment. The scheme utilizes multiple levels models to increase the accuracy of the detection and minimize false positives. Furthermore, the scheme is designed to be scalable and adaptable to different cloud computing environments, making it a practical solution for organizations of varying sizes and needs.

**Table 2.1**: Summary of the related literatures

| Reference | Methods | Problem | Objective | Strength and Weaknesses |
|---|---|---|---|---|
| [15] | Optimized Decision Forest | Phishing attack detection | Optimal performance for phishing detection | Though performance enhancement was achieved, FPR is on the high side |
| [16] | New ensemble model | To counter phishing attack | Improve on accuracy | Other art of the art perform metric was neglected |
| [17] | Comprehensive analysis | Phishing detection | Establish the most effective machine learning algorithms | RF effectiveness was established; however, model optimization was a drawback |
| [22] | Genetic Algorithm and Adaptive Resonance Theory Mapping | Detection of phishing websites | Performance improvement | Optimal performance was achieved, nevertheless, need for enhancement still exist |
| [23] | Meta-heuristic population technique and Sine Cosine Algorithm based K-NN | Phishing attack prediction | Performance enhancement | Though optimal result was obtained, however, dataset set for training was very small |
| [24] | Intelligent phishing detection framework | Phishing attack detection | Evaluate the strength of the proposed system | Challenge of time complexity is at excess |
| [25] | Comparative analysis | Classification of phishing URLs | Establish model with optimal performance | Performance score result, indicates that research is concluded in this domain, which is not acceptable, however, challenge maybe linked unprocessed data format |

## 3. Materials and methods

The experiment was run on a cloud environment the google collaboration python platform, with NVIDIA K80/T4 Graphical Processor Unit (GPU), 12GB GPU memory, 12GB Random Memory Access (RAM), and 358GB disk space. Other materials used in this research, include publicly available dataset from a known research database repository that is used majorly for academic studies often utilized in existing studies, and also used in analysis by baseline article for this research, the phishing dataset from kaggle, [26] which consist of 10000 instances and 48 features of phishing attack, inclusive of both the legitimate instances of a balanced ratio.

Basic method of data cleaning known as dropping of index numbering series was employed for efficient model and ensuring that relevant features are deployed for candidate model training, while normalization (data scaling) based Min-Max technique was also employed, thus, scaling the dataset in a bound range of [1] in order to improve model performance, convergence, and robustness to outliers.

Figure 3.1, depicts the design framework for a multi-level security-based intrusion detection model for phishing attack.

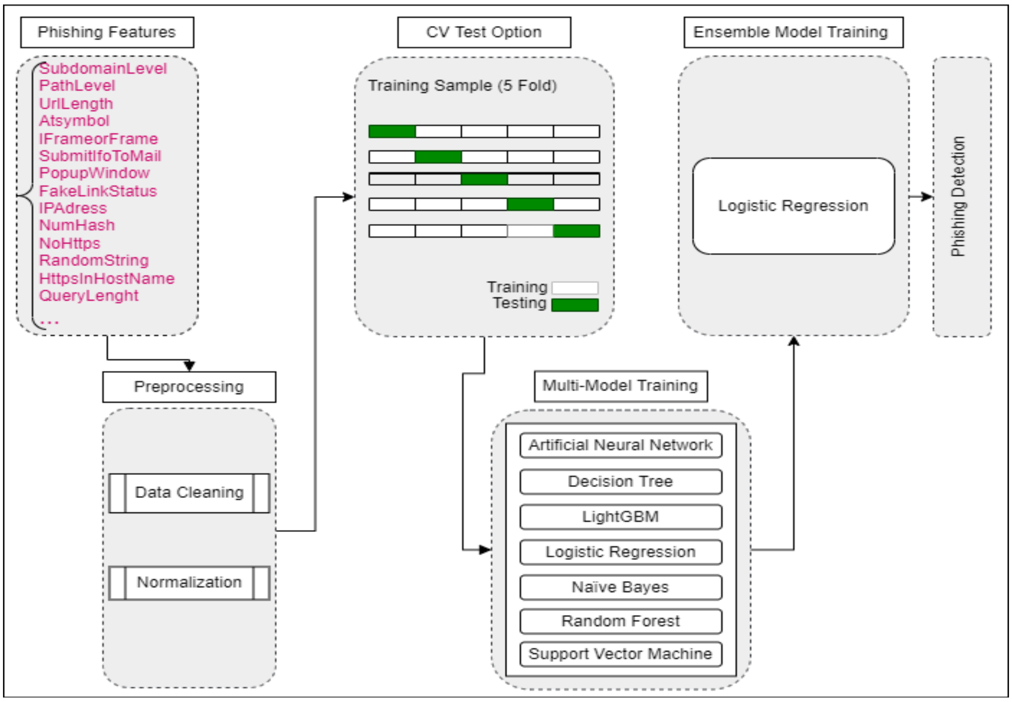

Figure 3.1: Multi-level model for phishing attack detection framework

The cross-validation (CV) test option phase, which entails the training, testing ratio of the dataset for proposed model building, was adopted, with a CV value of 5. The CV was chosen based on its capacity to construct enhanced model for experiments based on literatures.

The CV data is use for the multi-model training and validation, the multi-models form part of what is known as base learners (composition of multi-model), which in this case are Artificial Neural Network (ANN), Decision Tree (DT), Light Gradient Boost Machine (LightGBM), Logistics Regression (LR), Random Forest (RF) and support Vector Machine (SVM). Each train and makes predictions based on the complete supplied training set. The predictions from each serve as features to be fed into what is known as the meta-learner/ensemble model or stacking model (Logistic Regression), for training with a goal to discovering the best way to combine the output of the base learners and learn any existing pattern of misclassification and then generates the final model for prediction of phishing attack in cloud environment.

## 4. Results and Discussions

### *4.1 Result Presentation Based on Single Model Performance*

Figure 4.1 gives a graphical accuracy performance representation based on each model deployed for experimental analysis.

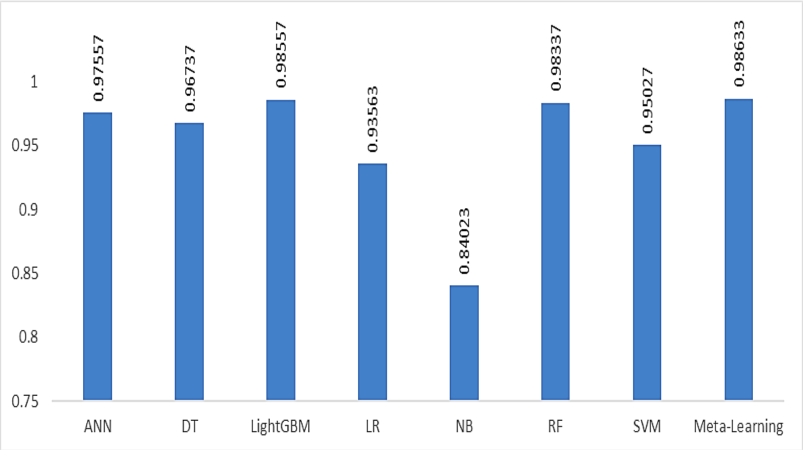

Figure 4.1: Accuracy performance of each model at training

### *4.2 Performance analysis of proposed multi-level model*

Figure 4.2 present the performance of the proposed multi-level model for phishing attack in cloud environment. The experimental analysis encompasses the performance evaluation of relevant performance metric in the domain of this research, such as accuracy, F1-measure, specificity, precision, recall and FPR.

Accuracy and F1-measure scores obtained from the experimental analysis of the proposed multi-level model for phishing detection at final stage of model validation is 0.9997 each, this reveals that the proposed model have the sufficient capacity to be able to detect accuracy phishing attack in cloud environment, this novel model proves the strength of multi-model engagement for research analysis to provide an enhanced performance to checkmating threat attack such as phishing in cloud environment.

The performance scores obtained from the experimental analysis of the proposed multi-level model for phishing detection is 0.9998 for as it specificity and precision each, indicating that the enhanced capability of the model is correct and precise to be able to identify phishing attacks in cloud environment, with a score above 99%, this is worth noting in the domain of machine learning as well as malware classification and detection performance strength.

To further establish the strength the proposed model against existing baseline literature, Table 4.1, depicts the comparative analysis of the proposed multi-level model for phishing detection in cloud environment (MLMPDC) with existing approaches from baseline articles.

The comparative analysis from Table 4.1 proved that the proposed MLMPDC model outperformed the baseline existing articles techniques based on the performance metrics of accuracy, F1 Measure, precision, recall and FPR as against the studies by [15-17].

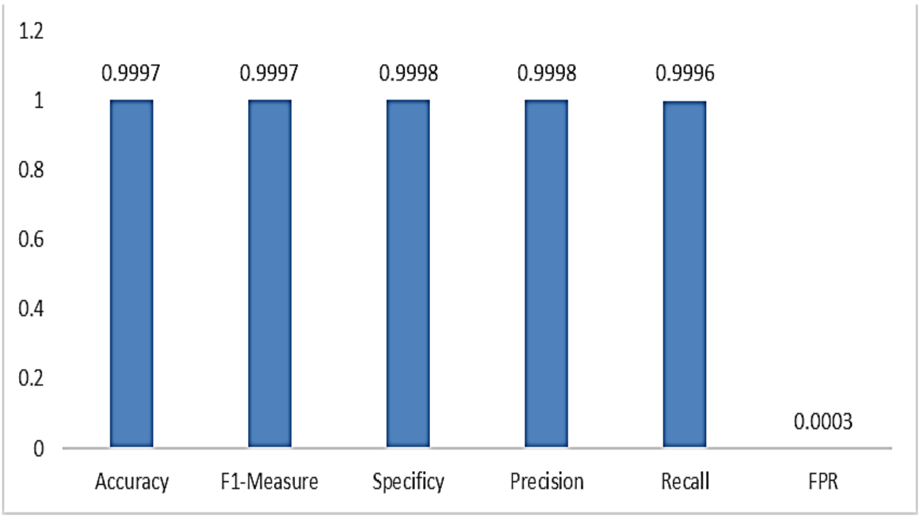

Figure 4.2: Performance evaluation of proposed model

**Table 4.1:** Comparative analysis of proposed model against baseline articles

| | Technique | Accuracy | F1-Measure | Specificity | Precision | Recall | FPR |
|---|---|---|---|---|---|---|---|
| **Proposed Model** | MLMPDC | 0.9997 | 0.9997 | 0.9998 | 0.9998 | 0.9996 | 0.0003 |
| **[15]** | ODF | 0.9837 | 0.98 | - | - | - | 0.016 |
| **[16]** | Ensemble | 0.9836 | - | - | - | - | |
| **[17]** | RF | 0.9294 | 0.8941 | 0.947 | 0.8941 | 0.8941 | |

**N.B:** (-) means the metric value is not reported in the reference.

## 5. Conclusions

Phishing attack continues to pretense severe threats to the security of cloud computing environment because of its strong reliance on Internet. With regards to the massive amount of traffic data and the dynamic nature of cyber-threats, most of the traditional detection models are faced with performance drawback, hence, this research proposed a robust and efficient multi-level model for phishing attack detection in cloud environment for solving the problem of classification error in the existing models. To build the new multi-level model, experimental procedure analysis was performed to examine and identify the appropriate capability functions of selected machine learning models. The multi-models were deployed as based model and LR was integrated as meta-model to effectively detect phishing attack in cloud environment. The results show that the proposed model outperformed the traditional selected models and the benchmarked approaches for the detection of phishing attack in cloud computing environment. The superior performance of the proposed model is because of the optimal function of integrating multi-model for detection of phishing attack to achieve improved detection performance. Based on the finding of this research, it can be concluded that the proposed model is a robust, efficient and intelligent technique capable of detecting phishing attack in cloud environment. In addition, with the capable technique of handling problem of performance deficiency as experienced in existing approaches by baseline articles for phishing detection. Furthermore, in future study, it is recommended that other new evolutionary based algorithms should be used to automatically tune parameter for training models on phishing detection datasets.

**References**

1. Hajimirzaei B, Navimipour NJ. Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *ICT Express.* 2018. Available from: https://doi.org/10.1016/j.icte.2018.01.014
2. Zhang F, Fu X, Yahyapour R. C base: a new paradigm for fast virtual machine migration across data centers. In: *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*. 2017. p. 284-293
3. Kumar PR, Raj PH, Jelciana P. Exploring data security issues and solutions in cloud computing. *Procedia Computer Science.* 2018; 125(9):691–697. Available from: https://doi.org/10.1016/j.procs.2017.12.089
4. Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of intrusion detection systems techniques, datasets and challenges. *Cybersecurity.* 2019; 2(1): 1-22.
5. Alkhalil Z, Hewage C, Nawaf L, Khan I. Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science*. 2021; 9(3):560 -063.
6. Li Y, Yang Z, Chen X, Yuan H, Liu W. A stacking model using URL and HTML features for phishing webpage detection. *Future Generation Computer Systems.* 2019; 94:27–39. Available from: https://doi.org/10.1016/j.future.2018.11.004
7. Gupta BB, Arachchilage NA, Psannis KE. Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems.* 2018; 67(2): 247-267.
8. Thomas J. Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. *International Journal of Business Management*. 2018; 12(3): 1-23.
9. Fiermonte, M. The threat of social engineering to networked systems [Dissertation]. Utica College; 2019
10. Lallie HS, Shepherd LA, Nurse JR, Erola A, Epiphaniou G, Maple C, Bellekens X. Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers & Security*. 2021; 105: 102248.
11. Rao RS, Pais AR. Detection of phishing websites using an efficient feature-based machine learning framework. *Neural Computing and Applications*. 2019; 31(8):3851-3873.
12. Khraisat A, Gondal I, Vamplew P, Kamruzzaman, J. Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*. 2019; 2(1): 1-22.
13. Baker T, Asim M, Tawfik H, Aldawsari B, Buyya R. An energy-aware service composition algorithm for multiple cloud-based IoT applications. *Journal of Network and Computer Applications.* 2017; 89: 96-108.
14. Wang W, Ren L, Chen L, Ding Y. Intrusion detection and security calculation in industrial cloud storage based on an improved dynamic immune algorithm. *Information Sciences.* 2018. Available from: https://doi.org/10.1016/j.ins.2018.06.072
15. Balogun AO, Mojeed HA, Adewole KS, Akintola AG, Salihu SA, Bajeh AO, Jimoh RG. Optimized decision forest for website phishing detection. In: *Proceedings of the Computational Methods in Systems and Software*. Springer, Cham.; 2021
16. Prince MSM, Hasan A, Shah FM. A new ensemble model for phishing detection based on hybrid cumulative feature selection. In: *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. 2021. p.7-12.
17. Khan SA, Khan W, Hussain A. Phishing attacks and websites classification using machine learning and multiple datasets (a comparative analysis). In: *International Conference on Intelligent Computing*. Springer, Cham. 2021.p.301 -313
18. Modi C, Patel D, Borisaniya B, Patel H, Patel, A, Rajarajan M. A survey of intrusion detection techniques in cloud. *Journal of network and computer applications.* 2013; 36(1):42-57.
19. Bace, RG, Mell P. Intrusion detection systems. 2001.
20. Zouhair C, Abghour N, Moussaid K, El Omri A, Rida M. A review of intrusion detection systems in cloud computing. *Security and Privacy in Smart Sensor Networks.* 2018; 253-83.
21. Bakshi A, Dujodwala YB. Securing cloud from ddos attacks using intrusion detection system in virtual machine. In: *2010 Second International Conference on Communication Software and Networks* 26 Feb 2010, IEEE; 2010.p.260-264).

22. Saravanan P, Subramanian S. A framework for detecting phishing websites using GA based feature selection and ARTMAP based website classification. *Procedia Computer Science*. 2020; 171:1083-1092.
23. Moorthy RS, Pabitha P. Optimal detection of phising attack using SCA based K-NN. *Procedia Computer Science*. 2020; 171:1716-1725.
24. Subasi A, Kremic E. Comparison of adaboost with multiboosting for phishing website detection. *Procedia Computer Science*. 2020; 168: 272-278.
25. Moruff OA, Maruf AO, Tosho A. Performance analysis of selected machine learning algorithms for the classification of phishing URLs. *Journal of Computer Science and Control Systems*. 2020; 13(2):16-19.
26. Shashwat T. *Phishing Dataset for Machine Learning*. 2021. Available from: https://www.kaggle.com/shashwatwork/phishing-dataset-for-machine-learning