SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V10I4P114

Volume 10 Issue 4, 141-151, April 2023
© 2023 Seventh Sense Research Group®

*Original Article*

# Improved Random Forest Algorithm using Chicken Swarm Optimization for Phishing Website Classification Model

C. Rajeswary[1], M. Thirumaran[2]

[1,2]*Department of Computer Science and Engineering, Puducherry Technological University, Puduchery, India.*

[1]*Corresponding Author : rajeswary.c@pec.edu*

Received: 20 February 2023 Revised: 25 March 2023 Accepted: 15 April 2023 Published: 30 April 2023

***Abstract*** *- Phishing is a type of online fraud which enables attackers to trick individuals into giving away confidential data like login credentials or financial data. A phishing website utilizes a URL that is comparable to a reasonable website to trick users into thinking it is legitimate, or it may comprise suspicious links or forms which has been developed for collecting sensitive data from users. Machine learning (ML) can be utilized to categorise websites as phishing or legitimate to protect users from falling victim to these attacks. There are several approaches to using machine learning for phishing website classification. This article focuses on the design of Chicken Swarm Optimization with Improved Random Forest for Phishing Website Classification (CSOIRF-PWC) technique. The CSOIRF-PWC technique aims to discriminate the legitimate and phishing websites accurately. To execute this, the presented CSOIRF-PWC approach initially performs the data normalization process. Next, the classification of phishing websites takes place using the IRF classifier. For improving the classification performance of the RF classifier, the parameter tuning process is performed through the CSO algorithm, which supports attaining improved classification performance. The simulation values of the CSOIRF-PWC methodology are investigated on two datasets, and the outputs are investigated under diverse measures. The comprehensive comparative outcomes emphasized the enhanced performance of the CSOIRF-PWC system over other methodologies in terms of several measures.*

***Keywords*** *- Phishing websites, Classification models, Random forest, Chicken swarm optimization, Machine learning, Security.*

## 1. Introduction

Cyberspace usage is increasing since it serves a crucial role in today's business and commercial activities, offering many online services which simplify our lives [1]. For example, online banking through the web has now been popular since several users are using it. The omnipresent nature of the internet for sharing data has certainly brought multiple attacks. Phishing and replay, denial of service, pharming, and masquerading are some of the notable attacks [2]. Phishing can succinctly be described as suspicious and fraudulent practices that include disseminating or sending several e-mails claiming to originate from reliable companies or individuals to appeal to the objective of revealing classified private data. Phishing assaults are becoming the main concern due to a rise in their numbers [3]. It is a broadly used, destructive, and effective assault in which hackers try to track users to reveal delicate data, like their debit card details and passwords [5].

A common phishing attack method includes phishing websites [6], in which the hacker traps users from accessing fake websites by copying the appearances and names of legal sites like Amazon, eBay, and Facebook. It is problematic for an individual to distinguish phishing sites from normal sites from phishing sites as it is similar. Sometimes, users do not verify the whole URL of the website, and also, once the user enters a phishing site, the hacker could access personal and delicate data [7]. Heuristic detection technology was presented for identifying phishing sites through deriving features of many third-party services and web pages; among third-party service aspects are WHOIS information, website ranking, and networking traffic recognition to solve problems with blacklist methods [8].

Machine learning (ML) related phishing website detection leverages ML techniques for detecting manually derived phishing site URL features. The efficacy of identification will be enhanced through this approach. This semi-automatic approach needs professionals to derive URL aspects physically [9].

It needs to update URL features because of the recurrent variations in the URL frameworks, demanding high maintenance costs and professional operation [10, 11]. Several authors have modelled ML-related solutions for preventing phishing attacks [12].

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

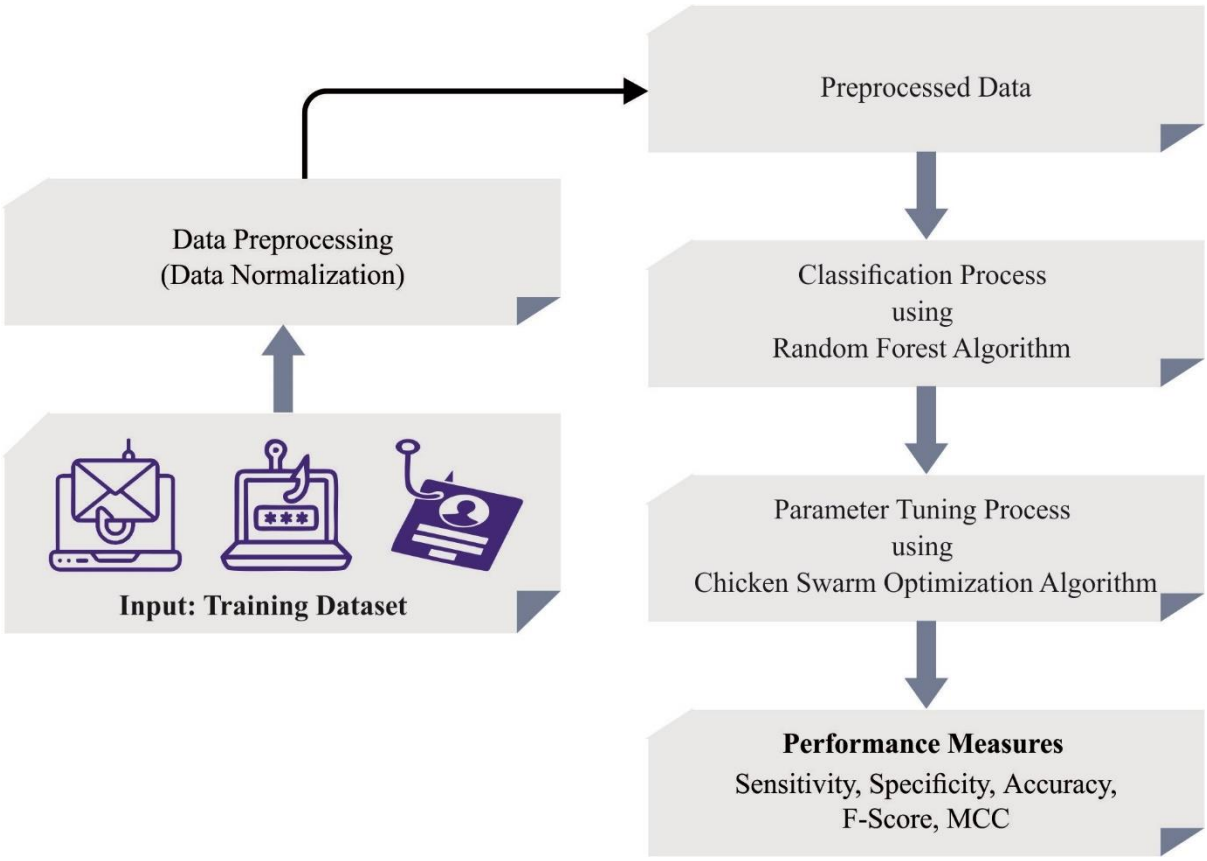


**Fig. 1 Workflow of CSOIRF-PWC model**

Still, many ML methods suffer from several problems, like the intervention of third parties information, high false-positive rates, and high response time [13].

This article focuses on the design of Chicken Swarm Optimization with Improved Random Forest for Phishing Website Classification (CSOIRF-PWC) technique. The presented CSOIRF-PWC technique initially performs a data normalization process. Next, the classification of phishing websites takes place using the IRF classifier. To improve the classification performance of the RF classifier, the tuning process is performed using the CSO technique, which supports attaining improved classification performance. The simulation values of the CSOIRF-PWC method are investigated on two datasets, and the outputs are investigated under diverse measures.

## 2. Related Works

The authors [15] presented an effective Hybrid DLcentric Phishing Detection System utilizing the MCS-DNN method. Initially, pre-processed can be accomplished on input data sets to ameliorate its quality. After, feature selection (FS) and clustering were executed to diminish the processing period and elevate the accuracy through CM-WOA and CoK-means, correspondingly. The features which were selected at the time of FS are given into the MCS-DNN technique that categorizes the phishing and legal websites. Finally, the K-fold Cross Validations were used to estimate the accuracy of the suggested system. Al-Sarem et al. [16] offer an optimizer stacking ensemble system for phishing site recognition. The optimization was executed through a GA for tuning the parameters of numerous ensemble ML techniques, which include GradientBoost, RF, LightGBM X, GBoost, Bagging, and AdaBoost. The optimized techniques were then ranked, and the optimal 3 methods were chosen as base classifiers of a stacking ensembling technique.

Aljofey et al. [17] modelled a novel technique for solving the anti-phishing issue. The novel features of this method were signified by URL character order without phishing previous data, many textual contents, and hyperlink information of webpage, which is compiled and given to training XGBoost technique. In [18], an intellectual phishing website detection framework is presented. The author leverages distinct ML methods for classifying websites as phishing or legitimate. Various classifier techniques are used for applying an accurate intellectual phishing website detection framework. Taha [20] modelled an effective

ensemble learning technique for phishing website recognition related to soft weight voting to enrich the recognition of phishing sites. A base classifier comprising 4 heterogeneous ML techniques was used for classifying the sites as phishing sites.

Stobbs et al. [21] explored the outcome that distinct features and optimization methods have on the accuracy of intellectual phishing recognition utilizing Techniques. This study looks at both hyperparameter optimization and the FS technique. For the tuning process, either TPE (Tree-structured Parzen Estimator) or GA have been examined, with optimal selection being method dependent. For FS, GA, MFO (Moth Flame Optimization), and PSO have been utilized, with PSO working better with the RF method. Yu [23] presented a hybrid method that combined the benefits of the DL neural network of ML, DBN, and the technique of SVM. Deep features were derived by the quick classifications of the DL method. The resultant feature vectors integrating with URL numerical aspects and web page code aspects, and webpage text attributes were given into the SVM method for classifying purposes.

## 3. The Proposed Model

In this article, we have established a novel CSOIRF-PWC technique for differentiating legitimate and phishing websites. The presented CSOIRF-PWC technique comprises data normalization, IRF-based phishing website classification, and a CSO-based parameter tuning process. Fig. 1 demonstrates the workflow of the CSOIRF-PWC algorithm.

### *3.1. Data Normalization*

Primarily, the presented CSOIRF-PWC technique performed the data normalization process. It is utilized min-max-based normalized from the presented technique that transforms a data value dv to dv' from the limit (min_$new$_value to max_$new$_value), as defined in (1).

$$dv = \frac{dv - dv\_\min}{dv\_\max - dv\_\min}(\max\_new\_value - \min\_new\_value) \quad (1)$$

In Eq. (1), the range of entirely transformed elements is represented by min_$new$_value to max_$new_value$. During this work, it is employed max_$new_value$ 1 and-max_$new_value$ to 0 and $min\_new\_value$ to 1. These transformed elements are then exploited as input data to the FS system.

### *3.2. Phishing Website Classification using RF Method*

In this work, phishing websites are classified using the IRF classifier. RF integrates many individual decision trees (DTs) [24]. CART is widely applied as a DT within RF due to the simplification and nonparametric behaviours. Every DT depends on random bootstraps data. Assume training dataset $TD = \{(X_1, Y_1), (X_2, Y_2), (X_N, Y_N)\}$ contains $N$ observations for the classification problem, $X_i$ refers to the input vector owing $M$ feature as $X_i = (x_{i1}, x_{i2}, x_{iM}), Y_i$ indicates the resultant scalar, the procedure of establishing the RF classification technique is discussed in Algorithm 1. Fig. 2 depicts the framework of RF. The primary objective of the RF training phase is to create several de-correlated $DTs$. To minimize modification related to the classification, an overlapping sample solution termed 'bagging' is implemented in the $RF$. It especially removes observation with replacement to produce non-dependent bootstrap samples from trained data. Next, every DT is trained from distinct bootstrap samples, which leads to improved tree diversity. In addition, to restrain the relationships amongst $DTs$, the better split of every node can be attained by randomly choosing $m$ subset feature rather than $M$ feature. Moreover, by applying dissimilar node features and bootstrap samples, noise immunity of RF is enhanced by averaging de-correlated $DTs$.

**Algorithm 1:** Pseudocode of RF Classifier

RF Training Procedure
For $j = 1$ to J: (J refers to the count of $DTs$)
Develop bootstrap sample $BS_j$ with $N$ observations from $TD$;
Fit a tree $DT_j$ depends on its $BS_j$:
Start dividing a node with every observation of $BS_j$.
Recursively repeating the subsequent procedures on all unsplit nodes:
Arbitrarily select $m$ features $(m < M)$ from $M$ candidates: $m \leftarrow M$
Find the split solution with the optimal set impurity amongst each feasible split of $m$ features in Process $i$.
Divide this node into 2 sub-nodes dependent upon the gained split solution from Process $ii$.
Gain the well-trained RF using ensembling every base DT learner $h_j(.)$.
end procedure
RF classification procedure
For a novel observation $X_{new}$, the output $RF(X_{new})$ of RF is predicted by:

$$RF(X_{new}) = \arg\max_Y \sum_{j=1}^{J} I\,(\tilde{h}_j(X_{new}) = Y)$$

where $\tilde{h}(X_{new})$ refers to the $jth$ DT's predictive outcome with $X_{new}$ as inputs. $I$ was a zero-one judgement with $I(\tilde{h}(X_{new}) = Y) = 1$. $\arg\max_Y$ outcomes the class with maximal counting number in every $DT$.

end procedure

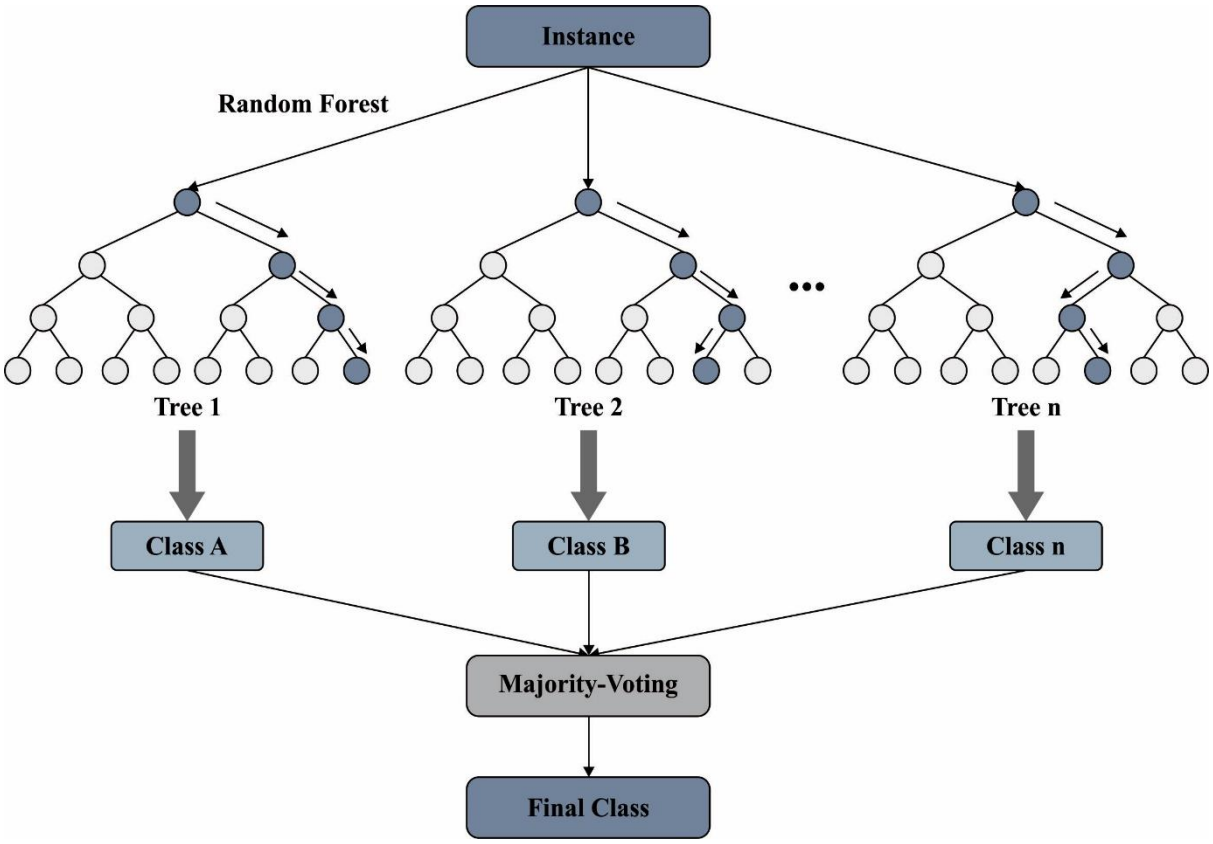


**Fig. 2 Structure of RF classifier**

Furthermore, for every DT within $RF$, due to the bagging solution, few training datasets are frequently applied as a bootstrap sample, leading to some other observations not being selected to fitting these $DT$. This observation is called an Out-of-Bag (OOB) sample. Generally, nearly one-third of TD composes an OOB sample and wouldn't be utilized during the process of RF training. As a result, every time DT is trained, the OOB sample is used for evaluating the performance of the classification of $DT$. In that way, RF is capable of achieving unbiased estimation without applying an external dataset. The OOB prediction with the $E_{OOB}$ The generalized error of RF is attained for classifying battery product property. It is noteworthy that the last. $E_{OOB}$ can be evaluated by the error rate of OOB prediction instead of averaging every DT's OOB error. Consequently, a confusion matrix for the battery manufacturing classification is produced, whereas a class-wise error can also be attained for all the classes.

### *3.3. CSO-based Parameter Tuning Process*

To increase the classification performance of the RF classifier, the tuning process is accomplished by using the CSO algorithm [25]. In general, the parameters involved in RF, such as the maximum depth of a CART (dmax), the minimal count of samples in a node (Ns, min), and the number of CARTs (NT ), etc., need to be tuned for accomplishing optimum performance of the RF classifier. The possible solution is to optimally adjust the RF parameters via the CSO system like that the classifier accuracy gets improved. CSO algorithm is a new metaheuristic approach initially coined in 2014 that is stimulated by chickens' natural behaviours during the foraging procedure.

The computation performance of the algorithm can be successfully enhanced by establishing both level relationships and level mechanisms amongst the individual population. In this study, the individual population can be divided into chick, rooster, and hen. Together with the chick following the mother hen, the rooster leads the population to move randomly, and the hen follows the roosters. During the iterative process, the upgrade of level mechanism, the competition, and mutual learning betwixt individual populations, are continuously progressing in search of the optimum global location. In the $D$-dimensional solution space, N represents the chicken count. Where $X_i^t (i \in [1, N])$ indicates the location of $i^{th}$ individuals in $t^{th}$ iteration. $N_r$ is the number of individual roosters, where the subgroup with the better location in the individual population, and the location can be updated as follows:

$$X_i^{t+1} = X_i^t * \left(1 + randn(0, \sigma^2)\right) \quad (2)$$

$$\sigma^2 = \begin{cases} 1 & f_i \le f_k \\ \exp\left(\frac{(fk - f_i)}{|\gamma_i + \varepsilon|}\right) & otherwise \end{cases} \quad (3)$$

Where $randn(0, \sigma^2)$ stands for the random Gaussian with a mean value of 0 and a standard deviation of $\sigma^2, \varepsilon$ indicates the minimal constant, $k$ indicates the random value $(k \in [1, N_r], k \neq i), f$ shows the fitness value of individuals. $N_h$is the hen counts, every hen will update the location with the rooster as a target and compete with other chicks. The updated location can be given as:

$$X_i^{t+1} = X_i^t + s_1 * r(X_{r1}^t - X_i^t) + s_2 * r(X_{r2}^t - X_i^t) \quad (4)$$

$$s_1 = \exp\left(\frac{f_i - f_{r1}}{abs(f) + \varepsilon}\right) \quad (5)$$

$$s_2 = \exp(f_{r2} - f_i) \quad (6)$$

Now $X_{r1}^t$ signifies the location of the rooster, followed by $i^{th}$ hens, $X_{r2}^t$ indicates the random selection of rooster or hen location. $X_{r1}^t \neq X_{r2}^t$

$N_c$ represent the number of chicks where the chick moves with the mother hen, and the location can be upgraded as follows:

$$X_i^{t+1} = X_i^t + FL * (X_m^t - X_i^t) \quad (7)$$

**Algorithm 2:** Pseudocode of CSO algorithm

Initiate
The proportion of different chickens and the number of populations parameters are set;
Randomly initialize the population, $t = 0$;
While $(t < G_{\max})$
If $t = I \| mod(t, G) == 0$
The population can be divided into hen, chick, and rooster subgroups, and the relationship of every subgroup is established;
End if
for $i = \mathrm{I} : N_r$
The rooster update is a random exploration that can be performed using Eq. (2);
End for
for $i = \mathrm{I} : N_h$
The hen moves with the rooster and competes with other hens; the update can be performed using Eq. (4);
End for
for $i = \mathrm{I} : N_c$
The chick moves with the mother, and the update can be performed using Eq. (7);
End for
Calculate the new solution;
Record the better individual and the fitness value.
End while
Where $X_m^t$ is mother hen position, $FL(FL \in [0,2])$ is the following co-efficient.

The CSO manner grows a fitness function (FF) for accomplishing better classification results. It expresses a positive integer for signifying a good efficacy of candidate solutions. During this vase, the minimization of the classifier error rate can measure that FF is expressed in Eq. (8).

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \quad (8)$$

## 4. Experimental Validation

In this segment, the result analysis of the CSOIRF-PWC approach is investigated utilizing a dataset comprising 47210 samples, as depicted in Table 1. The phishing websites from PhishTank (https://phishtank.org/) and Legal websites have been accumulated from ALEXA (https://www.alexa.com/)

**Table 1. Dataset details**

| Source | Class | No. of Instances |
|---|---|---|
| ALEXA | Legitimate URL | 24719 |
| PhishTank | Phishing URLS | 22491 |
| **Total Number of Instances** | | **47210** |

The confusion matrix of the CSOIRF-PWC approach on the phishing website classification process is illustrated in Fig. 3. The outputs represent that the CSOIRF-PWC technique has properly discriminated against legitimate and phishing URLs.

In Table 2 and Fig. 4, the phishing website classification outputs are provided with 70:30 of TRS and TSS. The experimental validation represented that the CSOIRF-PWC approach effectually recognizes legitimate and phishing URLs. For instance, on 70% of TRS, the CSOIRF-PWC technique gains average $accu_y$ of 94.47%, $sens_y$ of 94.47%, $spec_y$ of 94.47%, $F_{score}$of 94.49%, and MCC of 88.98%. On the other hand, on 30% of TSS, the CSOIRF-PWC system obtains average $accu_y$ of 94.27%, $sens_y$ of 94.27%, $spec_y$ of 94.27%, $F_{score}$of 94.29%, and MCC of 88.58%.

**Table 2. Phishing website classifier output of CSOIRF-PWC system on 70:30 of TRS/TSS**

| Class | Accuracy | Sensitivity | Specificity | F-Score | MCC |
|---|---|---|---|---|---|
| **Training Phase (70%)** | | | | | |
| Legitimate | 95.28 | 95.28 | 93.65 | 94.81 | 88.98 |
| Phishing | 93.65 | 93.65 | 95.28 | 94.17 | 88.98 |
| **Average** | **94.47** | **94.47** | **94.47** | **94.49** | **88.98** |
| **Testing Phase (30%)** | | | | | |
| Legitimate | 94.91 | 94.91 | 93.63 | 94.51 | 88.58 |
| Phishing | 93.63 | 93.63 | 94.91 | 94.06 | 88.58 |
| **Average** | **94.27** | **94.27** | **94.27** | **94.29** | **88.58** |

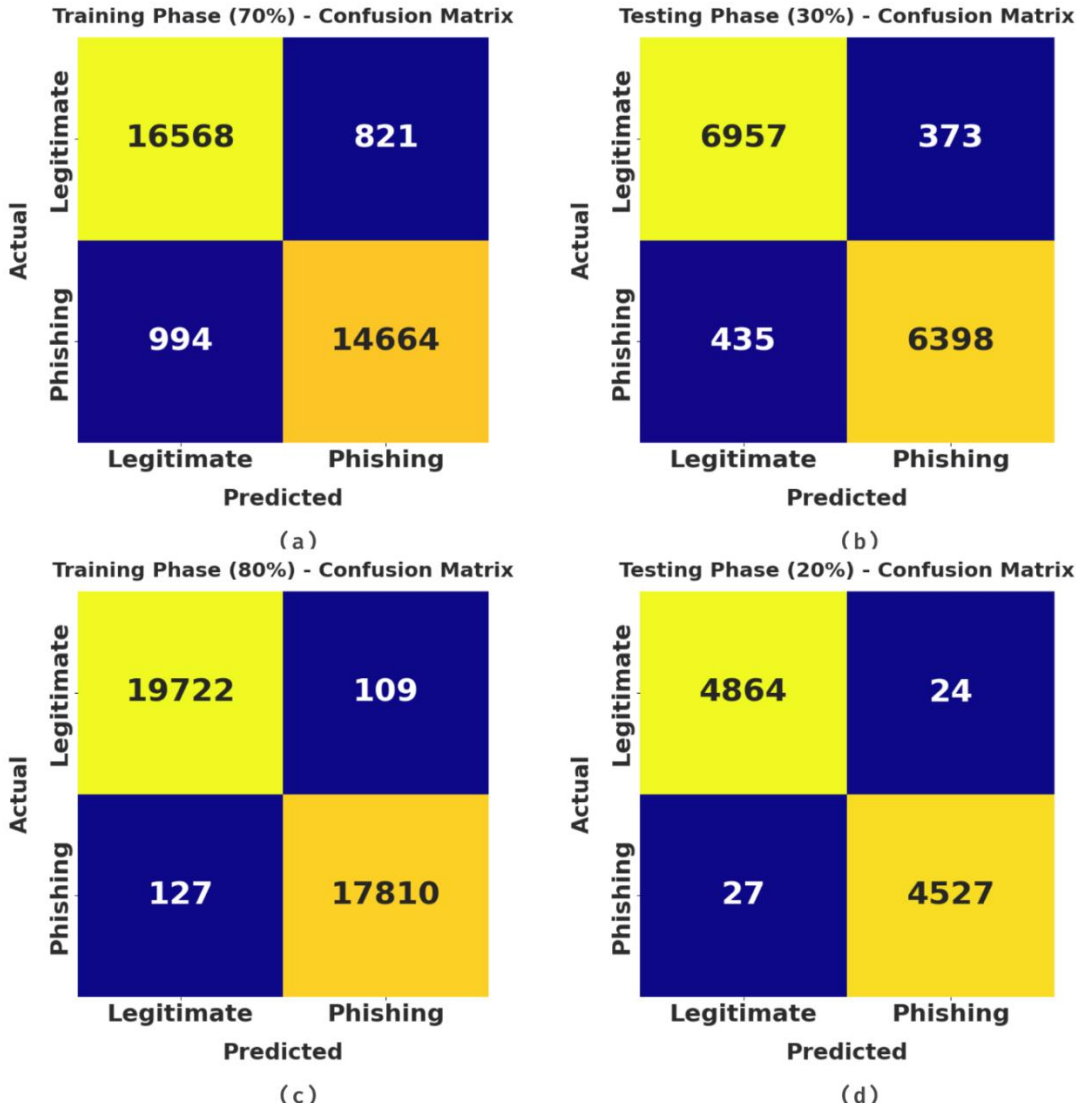


**Fig. 3 Confusion matrices of CSOIRF-PWC algorithm (a-b) TRS/TSS of 70:30 and (c-d) TRS/TSS of 80:20**

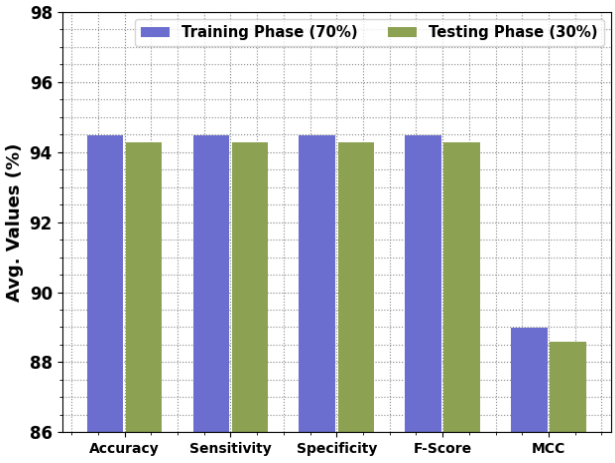


**Fig. 4 Average outcome of CSOIRF-PWC model on 70:30 of TRS/TSS**

In Table 3, the phishing website classification outcomes are given with 80:20 of TRS and TSS. The experimental outcome signified that the CSOIRF-PWC method effectively recognizes legitimate and phishing URLs.

**Table 3. Phishing website classifier output of CSOIRF-PWC system on 80:20 of TRS/TSS**

| Class | Accuracy | Sensitivity | Specificity | F-Score | MCC |
|---|---|---|---|---|---|
| **Training Phase (80%)** | | | | | |
| Legitimate | 99.45 | 99.45 | 99.29 | 99.41 | 98.75 |
| Phishing | 99.29 | 99.29 | 99.45 | 99.34 | 98.75 |
| **Average** | **99.37** | **99.37** | **99.37** | **99.37** | **98.75** |
| **Testing Phase (20%)** | | | | | |
| Legitimate | 99.51 | 99.51 | 99.41 | 99.48 | 98.92 |
| Phishing | 99.41 | 99.41 | 99.51 | 99.44 | 98.92 |
| **Average** | **99.46** | **99.46** | **99.46** | **99.46** | **98.92** |

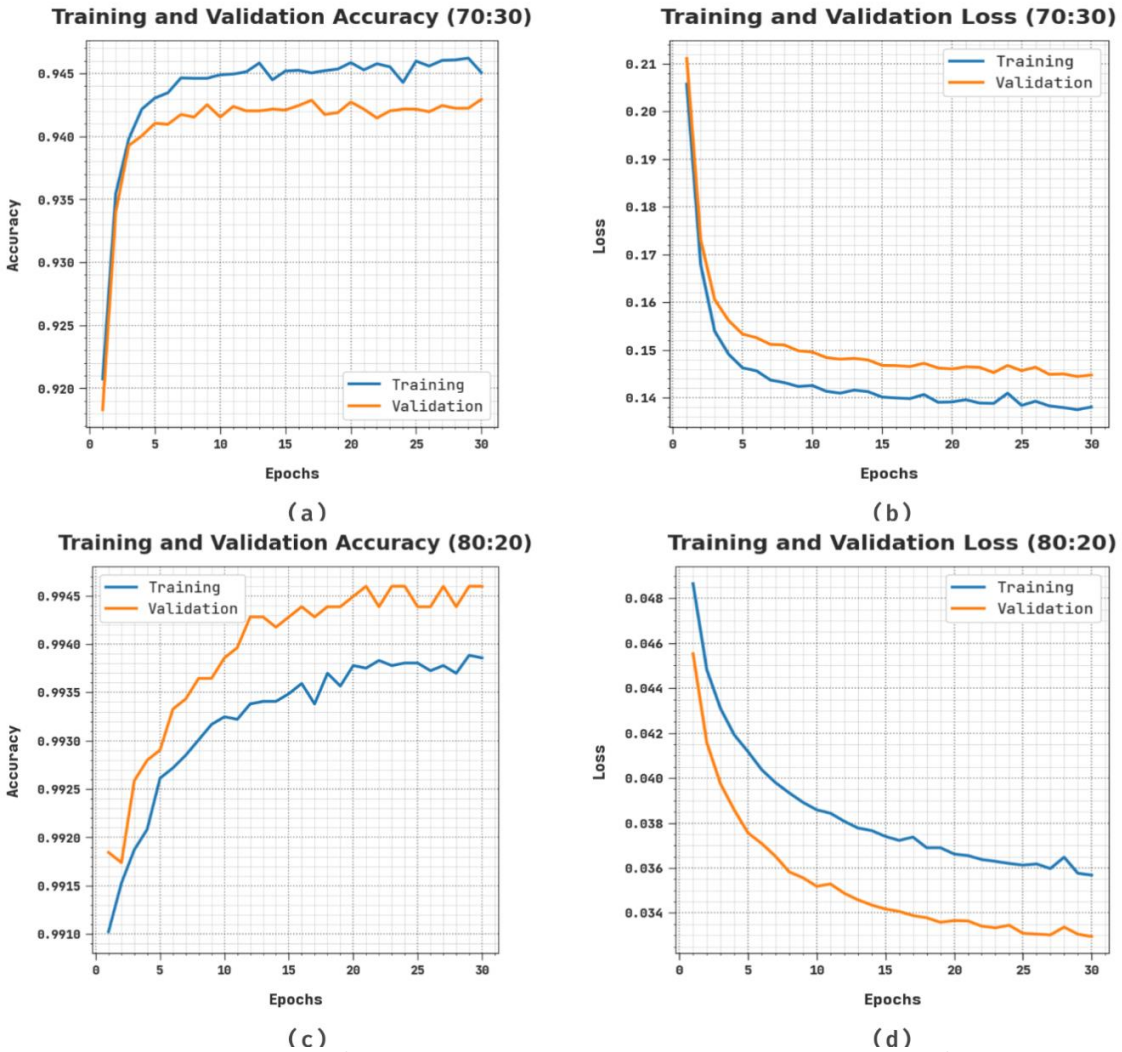


**Fig. 5 (a-c) Training and validation of accuracy on 70:30 and 80:20 and (b-d) Training and validation of loss on 70:30 and 80:20**

For sample, on 80% of TRS, the CSOIRF-PWC method reaches average $accu_y$ of 94.37%, $sens_y$ of 99.37%, $spec_y$of 99.37%, $F_{score}$of 99.37%, and MCC of 98.75%. Conversely, on 20% of TSS, the CSOIRF-PWC system obtains average $accu_y$ of 99.46%, $sens_y$ of 99.46%, $spec_y$ of 99.46%, $F_{score}$of 99.46%, and MCC of 98.92%.

Fig. 5 grants the accuracy and loss graph investigation of the CSOIRF-PWC model on 70:30 and 80:20 of TRS/TSS. The outcomes revealed that the accuracy value inclines to rise and the loss value inclines to reduce with maximum epoch count. It is experimental that the TLOS is lower and VACY is higher.

Fig. 6 establishes the classifier outcome of the CSOIRF-PWC technique 70:30 and 80:20 of TRS/TSS. Fig. 6a-6c reveals the PR examination of the CSOIRF-PWC model. The figures stated that the CSOIRF-PWC technique had acquired maximal PR performance under all classes. Lastly, Fig. 6b-6d illustrates the ROC study of the CSOIRF-PWC technique. The figure described that the CSOIRF-PWC technique had given an outcome in proficient outcomes with maximum ROC values under various classes.

Table 4 reports the overall comparative results of the CSOIRF-PWC technique [27-30].

**Table 4. Relative result of CSOIRF-PWC model with current systems**

| Methods | Accuracy | Sensitivity | Specificity | F-Score |
|---|---|---|---|---|
| CSOIRF-PWC | 99.46 | 99.46 | 99.46 | 99.46 |
| RF Model | 98.56 | 95.75 | 98.82 | 97.68 |
| LR Model | 98.80 | 98.70 | 98.53 | 98.08 |
| SVM Model | 96.17 | 97.75 | 95.09 | 95.40 |
| MLP Model | 95.70 | 95.51 | 98.51 | 98.85 |
| CNN-BERT Model | 96.50 | 97.15 | 97.43 | 98.73 |
| DNN Model | 95.87 | 96.87 | 95.49 | 97.77 |
| CNN Model | 95.45 | 95.21 | 95.79 | 97.17 |

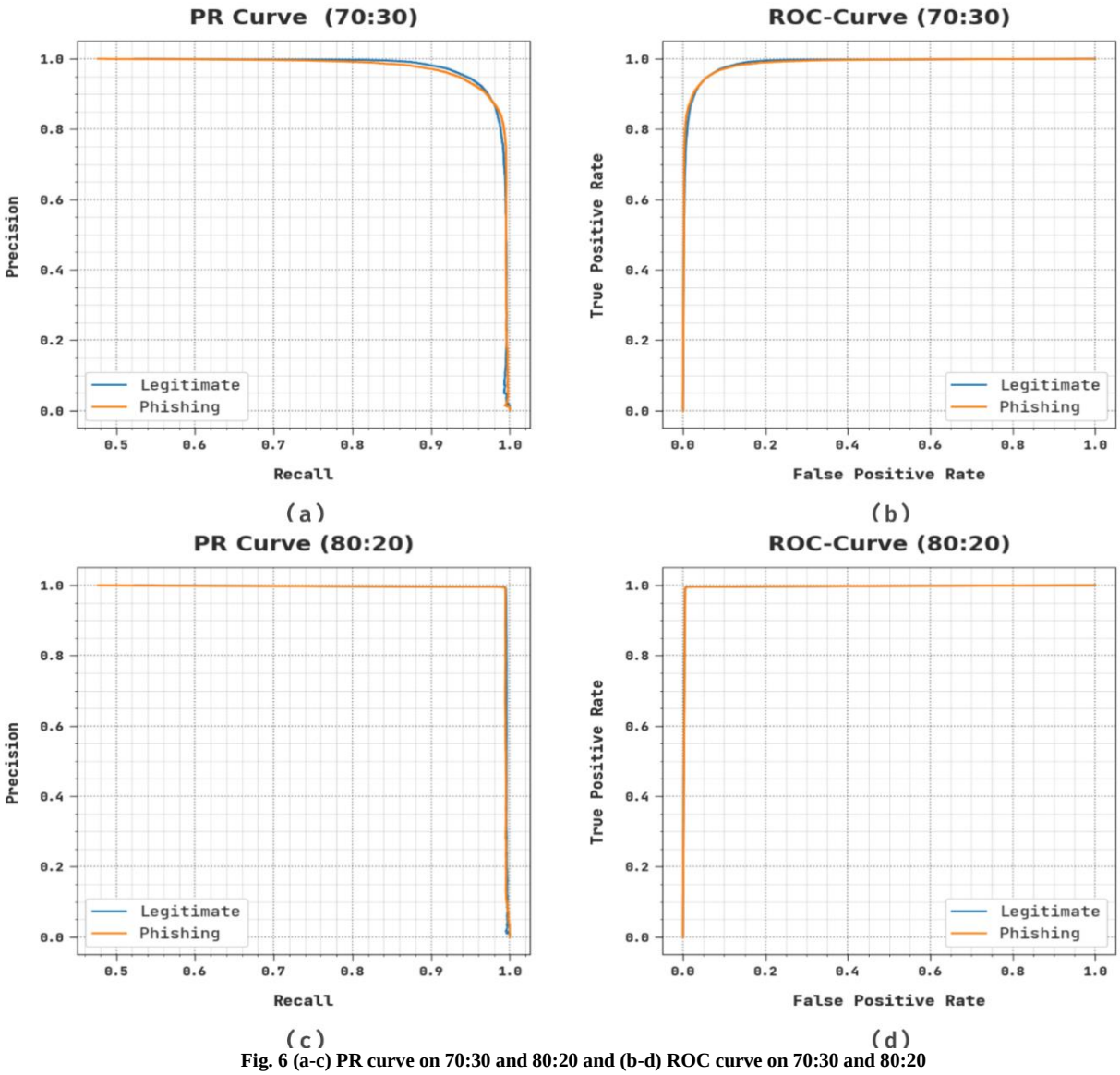


**Fig. 6 (a-c) PR curve on 70:30 and 80:20 and (b-d) ROC curve on 70:30 and 80:20**

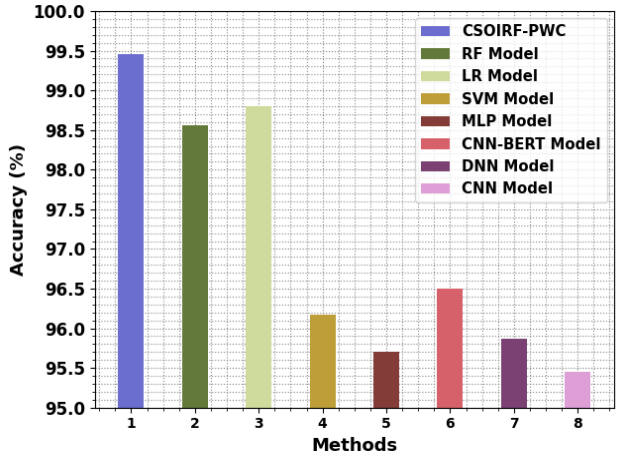


**Fig. 7 $Accu_y$ the outcome of the CSOIRF-PWC model with current approaches**

Fig. 7 exhibits a relative study of the CSOIRF-PWC technique with current methods in terms of $accu_y$. The outcomes indicate that the MLP, DNN, and CNN models obtain minimally $accu_y$ of 95.70%, 95.87%, and 95.45% respectively. Followed by the SVM and CNN-BERT models attain certainly improved $accu_y$ of 96.17% and 96.50%, respectively. Meanwhile, the RF and LR models resulted in reasonable $accu_y$ of 98.56% and 98.80%, respectively. But the CSOIRF-PWC technique reports maximum outcomes with $accu_y$ of 99.46%.

Fig. 8 displays a comparative investigation of the CSOIRF-PWC approach with current methods in terms of $sens_y$. The outcomes designate that the MLP, DNN, and CNN techniques acquire lesser $accu_y$ of 95.51%, 96.87%, and 95.21% correspondingly. Afterwards, the SVM and

CNN-BERT systems certainly accomplish better $sens_y$ of 97.75% and 97.15% correspondingly. In the meantime, the RF and LR models resulted in reasonable $sens_y$ of 95.75% and 98.70% correspondingly. Finally, the CSOIRF-PWC technique reports maximal outcomes with $sens_y$ of 99.46%.

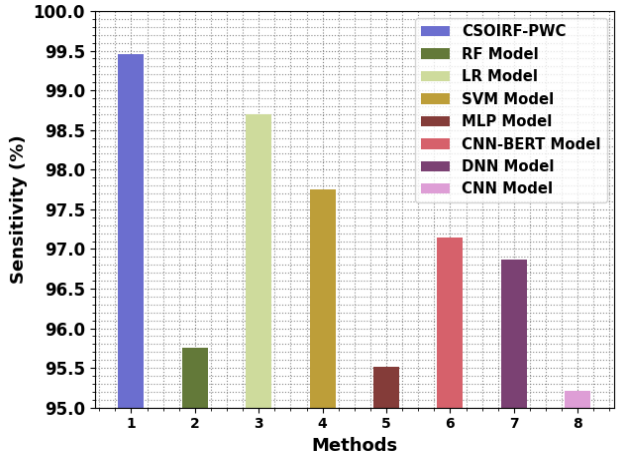


**Fig. 8 $Sens_y$ the outcome of the CSOIRF-PWC model with current systems**

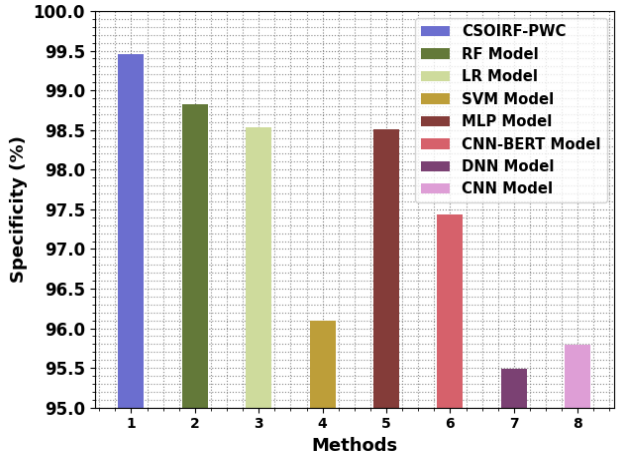


**Fig. 9 $Spec_y$ the outcome of the CSOIRF-PWC model with current systems**

Fig. 9 shows a relative analysis of the CSOIRF-PWC technique with current methods for $spec_y$. The outcomes implied that the MLP, DNN, and CNN models obtain minimally $spec_y$ of 98.51%, 95.49%, and 95.79% correspondingly. Next, the SVM and CNN-BERT methods certainly attain improved $spec_y$ of 95.09% and 97.43%, correspondingly. In addition, the RF and LR techniques resulted in reasonable $spec_y$ of 98.82% and 98.53%, correspondingly. Finally, the CSOIRF-PWC technique reports higher outcomes with $spec_y$ of 99.46%.

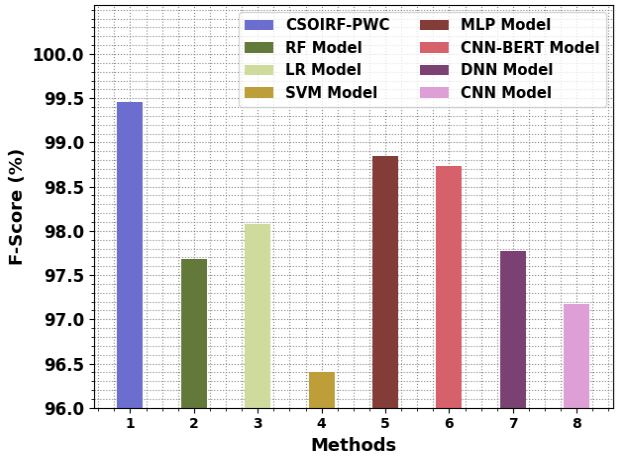


**Fig. 10 $F_{score}$ the outcome of the CSOIRF-PWC model with current systems**

Fig. 10 reveals a comparative study of the CSOIRF-PWC algorithm with current techniques in terms of $F_{score}$. The outcomes indicate that the MLP, DNN, and CNN methodologies gain minimal $F_{score}$ of 98.85%, 97.77%, and 97.17% correspondingly. In addition, the SVM and CNN-BERT approaches gain certainly enhanced $F_{score}$ of 95.40% and 98.73% correspondingly. Besides, the RF and LR models resulted in reasonable. $F_{score}$ of 97.68% and 98.08%, correspondingly. Eventually, the CSOIRF-PWC technique reports maximum outputs with $F_{score}$ of 99.46%. These outputs guaranteed the enhancement of the CSOIRF-PWC method over other techniques.

## 5. Conclusion

In this study, we have established a novel CSOIRF-PWC technique for differentiating legitimate and phishing websites. Initially, the presented CSOIRF-PWC technique performed the data normalization process. Next, the classification of phishing websites takes place using the IRF classifier. For improving the classification performance of the RF classifier, the tuning process is accomplished by using the CSO approach, which supports attaining improved classification performance. The simulation values of the CSOIRF-PWC method are investigated on two datasets, and the outcomes are reviewed under different measures. The comprehensive comparative outcomes emphasized the enhanced performance of the CSOIRF-PWC method over other techniques in terms of several measures. Thus, the presented CSOIRF-PWC method can be implemented for the automated classification of phishing websites. In the coming days, the achievement of the CSOIRF-PWC technique can be advanced by FS models.

## References


[1] Ammar Almomani, "Phishing Website Detection with Semantic Features Based on Machine Learning Classifiers: A Comparative Study," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1-24. [CrossRef] [Google Scholar] [Publisher Link]

[2] Lizhen Tang, and Qusay H. Mahmoud, "A Survey of Machine Learning-Based Solutions for Phishing Website Detection," *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 672-694, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Zeyad Ghaleb Al-Mekhlafi et al., "Phishing Websites Detection by Using Optimized Stacking Ensemble Model," *Computer Systems Science and Engineering*, vol. 41, no. 1, pp. 109-125, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Aditya Kharat et al., "Implementation of Defence Schemes for Phishing Attacks on Mobile Devices," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 6, pp. 28-34, 2019. [CrossRef] [Publisher Link]

[5] Ammar Odeh, Ismail Keshta, and Eman Abdelfattah, "Machine Learningtechniquesfor Detection of Website Phishing: A Review for Promises and Challenges," *In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0813-0818, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yazan A. Alsariera, "Intelligent Tree-Based Ensemble Approaches for Phishing Website Detection," *Journal of Engineering Science and Technology*, vol. 17, no. 1, pp. 563-582, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Charu Singh, and Meenu, "Phishing Website Detection Based on Machine Learning: A Survey," *In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 398-404, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Arriane Livara, and Rowell Hernande, "An Empirical Analysis of Machine Learning Techniques In Phishing E-Mail Detection," *In 2022 International Conference for Advancement In Technology (ICONAT)*, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] R. Selvaraj et al., "Optimized Machine Learning for CHD Detection Using 3D CNN-Based Segmentation, Transfer Learning and Adagrad Optimization," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 3, pp. 20-34, 2023. [CrossRef] [Publisher Link]

[10] Rajesh Singh et al., "Internet of Wild Things with the Integration of Vision Technology and Lora Network," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 3, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] A N Shilpa, and C S Veena, "A Hybrid Compressive Sensing Network for ROI-Based Medical Image Recovery," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 3, pp. 8-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mohammad A. Alsharaiah et al., "A New Phishing-Website Detection Framework Using Ensemble Classification and Clustering," *International Journal of Data and Network Science*, vol. 7, no. 2, pp. 857-864, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shatha Ghareeb et al., "Analysis of Feature Selection and Phishing Website Classification Using Machine Learning," *In 2023 15th International Conference on Developments in Esystems Engineering (DeSe)*, pp. 178-183, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Dr. Ananthi Sheshasaayee, and V.Vidyapriya, "Reorganisation of Adaptive Websites Using Web Usage Mining Techniques," *International Journal of Computer & Organization Trends*, vol. 4, no. 3, pp. 53-58, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[15] Anitha, J., and Kalaiarasu, M., "A New Hybrid Deep Learning-Based Phishing Detection System Using MCS-DNN Classifier," *Neural Computing and Applications*, vol. 34, no. 8, pp. 5867-5882. [CrossRef] [Google Scholar] [Publisher Link]

[16] Mohammed Al-Sarem et al., "An Optimized Stacking Ensemble Model for Phishing Websites Detection," *Electronics*, vol. 10, no. 11, p. 1285. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ali Aljofey et al., "An Effective Detection Approach for Phishing Websites Using URL and HTML Features," *Scientific Reports*, vol. 12, no. 1, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Abdulhamit Subasi, Emir Kremic, "Comparison of Adaboost with Multiboosting for Phishing Website Detection," *Procedia Computer Science*, vol. 168, pp. 272-278, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Peravali Kavya, "An Efficient Machine Learning Based Algorithm for Preventing Phishing Websites," *SSRG International Journal of Computer Science and Engineering*, vol. 5, no. 12, pp. 10-13, 2018. [CrossRef] [Publisher Link]

[20] Altyeb Taha, "Intelligent Ensemble Learning Approach for Phishing Website Detection Based on Weighted Soft Voting," *Mathematics*, vol. 9, no. 21, p. 2799, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jordan Stobbs, Biju Issac, and Seibu Mary Jacob, "Phishing Web Page Detection Using Optimised Machine Learning," *In 2020 IEEE 19th International Conference on Trust, Security and Privacy In Computing and Communications (TrustCom)* , pp. 483-490, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jebakumar Immanuel et al., "*The Effectiveness of Security Images In Internet Banking," International Journal of P2P Network Trends and Technology*, vol. 7, no. 3, pp. 6-9, 2017. [Google Scholar] [Publisher Link]

[23] Xuqiao Yu et al., "Phishing Websites Detection Based on Hybrid Model of Deep Belief Network and Support Vector Machine," *In IOP Conference Series: Earth and Environmental Science* , vol. 602, no. 1, p. 012001, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Kailong Liu, "Feature Analyses and Modeling of Lithium-Ion Battery Manufacturing Based on Random Forest Classification," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 6, pp.2944-2955. [CrossRef] [Google Scholar] [Publisher Link]

[25] Zhi-Feng Liu et al., "Prediction Short-Term Photovoltaic Power Using Improved Chicken Swarm Optimizer-Extreme Learning Machine Model," *Journal of Cleaner Production*, vol. 248, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ameerabanu .B, and Niaz Ahamed .V .M, "Authenticated Framework for Security Based on Imperceptible Captcha," *SSRG International Journal of Computer Science and Engineering*, vol. 3, no. 5, pp. 38-42, 2016. [CrossRef] [Publisher Link]

[27] Rundong Yang et al., "Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning," *Sensors*, vol. 21, no. 24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Ali Aljofey et al., "An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL," *Electronics*, vol. 9, no. 9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[29] Muna Elsadig et al., "Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction," *Electronics*, vol. 11, p. 3647, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mohammad Almseidin et al., "Cyber-Phishing Website Detection Using Fuzzy Rule Interpolation," *Cryptography*, vol. 6, no. 2, p. 24, 2022. [CrossRef] [Google Scholar] [Publisher Link]