Arabian Journal for Science and Engineering (2020) 45:10459–10470
https://doi.org/10.1007/s13369-020-04802-1

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Phishing Website Detection: Forest by Penalizing Attributes Algorithm and Its Enhanced Variations

Yazan Ahmad Alsariera[1] · Adeyemo Victor Elijah[2] · Abdullateef O. Balogun[3,4]

Received: 27 March 2020 / Accepted: 14 July 2020 / Published online: 21 July 2020

Abstract
The damaging effect of phishing is traumatizing as attackers or hackers execute theft of sensitive information from users subtly for inappropriate or unauthorized usage. In the light of curbing phishing, blacklisting of websites proved ineffective as the deployment of phishing websites are rampantly increasing and often short-lived. Hence, machine learning (ML) methods are seen as viable measures and used to develop deplorable models that can detect a phishing website. ML methods are fast gaining attention and acceptance in detecting phishing websites as they can cope with the dynamism of phishing websites and attackers. However, ML methods still suffer some shortcomings in terms of low detection accuracy, high false alarm rate (FAR) and induced bias of developed ML solutions. In addition, with the evolving nature of phishing attacks, there is a continuing imperative need for novel and effective ML-based methods for detecting phishing websites. This study proposed 3 meta-learner models based on Forest Penalizing Attributes (ForestPA) algorithm. ForestPA uses a weight assignment and weight increment strategy to build highly efficient decision trees by exploiting the prowess of all attributes (non-class inclusive) in a given dataset. From the experimental results, the proposed meta-learners (ForestPA-PWDM, Bagged-ForestPA-PWDM, and Adab-ForestPA-PWDM) are highly efficient with the least accuracy of 96.26%, 0.004 FAR, and 0.994 ROC value. Further, with the superiority of the proposed models over other existing methods, we recommend the development and adoption of meta-learners based on ForestPA for phishing website detection and other cybersecurity attacks.

Keywords Phishing · Meta-learners · Machine learning · Cybersecurity

1 Introduction

With respect to the advancement in information technology (IT), there has been an increasing number of digital services provided through the Internet ranging from financial services to gaming applications [1]. Financial services, social media, and online gaming applications are the top engaging digital web-based services with a huge and increasing audience. The enormous amount of users of these digital services signifies its success and acceptability in modern society [2]. With a high level of acceptability, there are cybersecurity issues such as privacy disclosure, identity theft, and phishing that comes with these digital services via the Internet [3].

In recent times, fake websites are being developed and hosted by criminals to steal sensitive information such as credit card information, passwords, and usernames from unsuspecting users for illegal activities or transactions. This is referred to as a form of a phishing attack [4]. It is a critical cybersecurity issue plaguing cyberspace with a severely

✉ Yazan Ahmad Alsariera
yazan.ahmad@nbu.edu.sa

Adeyemo Victor Elijah
v.e.adeyemo@gmail.com

Abdullateef O. Balogun
balogun.ao1@unilorin.edu.ng

[1] Department of Computer Science, Faculty of Science, Northern Border University, 73222 Arar, Kingdom of Saudi Arabia

[2] School of Built Environment, Engineering and Computing, Leeds Beckett University, Headingley Campus, Leeds LS6 3QS, United Kingdom

[3] Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, 32610 Perak, Malaysia

[4] Department of Computer Science, University of Ilorin, PMB 1515 Ilorin, Nigeria

damaging after effect on the Internet users and businesses [5, 6].

Vrbančič et al. [7] described phishing as an extensive fraud that occurs when a malicious website acts, looks, and feels almost identical to a legitimate website, bearing in mind that the utmost goal is the obtainment of victim's sensitive data. Over the years, efforts and countermeasures are put in place in detecting phishing websites such as having a blacklist repository, educating digital users about cybersecurity, Google PageRank method, and even developing machine learning (ML) models [8]. There are three categories of phishing detection mechanisms according to [9], namely (1) machine learning; (2) heuristic; and (3) list-based methods.

However, phishing website is highly becoming capable of avoiding detection due to the evolving nature of conducting phishing attacks by attackers as there are ways of evading these conventional countermeasures [10, 11]. In the case of ML methods, which is the focus of this study, there have been models with relatively high detection accuracy. However, they often suffer from high false alarm rate (FAR) aside seeking the improved performance of phishing methods [2, 12].

Given the unrelenting efforts of attackers to conduct subtle digital activities through social engineering that often lead to stealing of private information, identity theft, financial loss, and customers' inability to trust previously attacked organizations such as bank and e-commerce ventures [13], and the growing prowess of phishing websites to evade detection [10], the need for a robust, up-to-date Phishing Websites Detection Models (PWDMs) is imminent to effectively and efficiently prevail in categorizing legitimate website from a phishing website and thereby abate this nefarious activity.

Hence, this study proposed novel ForestPA-based meta-learning models for detection of phishing websites. ForestPA uses a weight assignment and weight increment strategy to build highly efficient decision trees by exploiting the prowess of all attributes (non-class inclusive) in a given dataset.

Specifically, the following are contributions of this study to the body of knowledge:

(1) The use of more recent and comprehensive featured phishing website data as input data, i.e. the UCI phishing website dataset, for the development of PWDMs.
(2) Implementation of ForestPA algorithm for detecting both legitimate and phishing websites;
(3) Implementation of Bagging and Boosting Meta-learners for improving ForestPA performance; and
(4) An empirical comparison of the proposed PWDMs with existing state-of-the-art phishing methods.

More so, it is the intention of this study to answer the following research questions:

(1) How effective is the ForestPA algorithm implementation for detecting phishing and legitimate websites?
(2) How effective are the Meta-learners (Bagged-ForestPA and Boosted-ForestPA) in detecting phishing and legitimate websites?
(3) How well is the performance of the proposed PWDMs compared with existing state-of-the-art methods?

The remaining Sections of this paper include Related Works which presents the critical review of existing related methods that are currently published. The 'Method' section provides details about the phishing website datasets, the implemented algorithms, this study experimental framework, and the performance evaluation metrics for testing the proposed phishing website detection model. The 'Experimental Results' section reports the performance results of all developed and tested models in a step-wise manner (i.e. ForestPA-PWDM, Bagged-ForestPA-PWDM, Boosted-ForestPA-PWDM). The reported performance of each model (as espoused in the 'Method' section) was discussed individually while some figures (i.e. summarized visualization) being provided. The 'Discussion' section comprehensively discussed the performance of the proposed methods of this study. More so, a comprehensive comparative analysis of the methods developed and evaluated in this study against existing methods (as reviewed in the 'Related Work' section) was presented. Lastly, the 'Conclusion' section brings this study to an end by providing the answers to the highlighted questions of this research. In addition, the 'Conclusion' section identifies the future works of this study.

2 Related Works

The critical review of some related existing studies based on phishing attack is essential in order to establish and amplify the importance and significance of this study. Zamir et al. [14] presented a machine learning-based method for detecting phishing website using the same dataset as this study. The study conducted various experiments using information gain, gain ratio, relief-F, principal component analysis (PCA), and recursive feature elimination feature selection algorithms. Also, it made use of support vector machine (SVM), Naïve Bayes, Random Forest, K-nearest neighbour (kNN), Bagging and Neural Network (NN) machine learning algorithms. These algorithms were used for finding optimal features, developing individual machine learning models, and were also combined using in two (2) different Stacking methods vis-à-vis (RF + NN + Bagging) and (kNN + RF + Bagging). The performances of all developed

models were evaluated using accuracy, recall and precision metrics. Conclusively, the research's proposed method (i.e. PCA feature selection and Stacking (RF + NN + Bagging)) produced the best accuracy of 97.4%

The research conducted by Abdulrahaman et al. [2] presented PWDMs based on Random Forest with Wrapper Feature Selection Method. The study presented a decision table PWDM of 93.24% accuracy with 0.75 false alarm rate (FAR), a sequential minimal optimization (SMO) for support vector classifier PWDM of 93.81% accuracy with 0.066 FPR, a Naïve Bayes (NB) PWDM of 92.98% accuracy with 0.076 FPR, and the proposed wrapper-based Random Forest PWDM of 97.259% accuracy with 0.03 FPR. The main limitation of the study is the performance comparison of an ensemble method model against single classifier models. Random Forest is an ensemble of decision tree, and it is expected to produce superior model against single classifiers. In addition, some of the models had high FPR values.

Ali and Ahmed [3] study hybridized Deep Neural Networks (DNN) and genetic-based feature selection with weighting methods. The DNN model achieved an accuracy of 88.77% with True Positive Rate of 85.83%. The proposed DNN with GA based on feature weighting had 91.13% accuracy and 90.79% TPR. The limitation of the developed model of this study is that the accuracy and TPR scores are relatively low which indicate a very high FPR value.

The research work of Zabihimayvan and Doran [9] presented a method for detecting phishing websites using machine learning-based strategy. Importantly to the publication is the usage of fuzzy rough set (FRS) algorithm for executing an efficient feature selection process as a data preprocessing method for enhancing phishing website detection models. The resulting features were used to generate a subset of the original phishing website dataset and served as input into three machine learning algorithms (1) Multi-layered Perceptron (MLP); (2) Random Forest; and (3) SMO. The performance of FRS was evaluated using the F-Measure metric while it was compared against other feature selection algorithms namely (1) Information Gain (IG); (2) Correlated Feature set (CFS); and (3) a hybridized decision tree and the Wrapper method (DW). The experimentation of these methods was conducted on three benchmark phishing datasets and further tested on 14,000 website samples. The best variation of FRS model (FRS algorithm used in conjunction with the Random Forest classification method) achieved 95% F-measure value.

Ferreira et al. [4] implemented the (MLP for developing PWDM. Their proposed PWDM produced a reported accuracy of 87.61%. In the same vein, Vrbančič et al. [7] used swarm intelligence approach (an evolutionary algorithm) for finding optimal parameter settings of Deep Learning Neural Network (TDLBA). The proposed model was fitted and evaluated on the UCI phishing datasets. TDLBA produced an accuracy of 96.5%. The performance of the developed model of this study was evaluated using only the accuracy measure. This limited the study as accuracy is neither the appropriate nor the only measure of evaluating a classification model whose data are highly imbalance.

Subasi et al. [13] in their study implemented PWMDs based on Artificial Neural Network (ANN), Classification and Regression Trees (CART), and Rotation Forest (RoF), respectively. From their experimental results, ANN had an accuracy of 96.91% with AUC-ROC score of 0.995, CART had 95.79% accuracy with AUC-ROC score of 0.981, and RoF had 96.79% accuracy and AUC-ROC score of 0.994. Although the models of this study are relatively high performing with terms of accuracy and AUC-ROC; however, the FPR of the models were not reported which will determine the viability of the models.

Summarily, from existing studies, various ML, DL, evolutionary algorithms, and feature selection techniques have been applied to develop viable PWDMs. However, the problem of high FAR still persists. In addition, the application of meta-learners for classification tasks has been proven to be effective as it reduces variance and bias in classification processes [15]. Consequently, this study proposes novel meta-learners (ForestPA-PWDM, Bagged-ForestPA-PWDM, and Adab-ForestPA-PWDM) based on ForestPA for detecting phishing websites.

3 Method

3.1 Dataset

There are existing standard datasets for conducting ML experiments for the development of phishing website detection models. Although, some researches chose to crawl the internet and compile a list of legitimate and phishing websites. In this study, we make use of the standard phishing website dataset created by [16]. The dataset is made available on the UCI data repository (https://archive.ics.uci.edu/ml/machine-learning-databases/00327/) for the sole purpose of developing ML-based phishing website detection models. The dataset contained comprehensive features cutting across four (4) different categories [11]. The categories of which the engineered and extracted features belong to are: (1) Address Bar-based features (2) Abnormal-Based features, (3) HTML and JavaScript-based features, and lastly (4) Domain-based features. These categories produce ranging numbers of independent features, more so, the availability of statistical reports on a URL from the reputable organization was made to into feature. The details of the dataset used by this study for experimentation are provided in Table 1.

As depicted in Table 1, the dataset consists of 31 attributes, of which only one (1) is the class variable (label).

Table 1 Description of Studied Phishing Website Dataset

Dataset Description		
Total number of attributes	31	
No. of independent variables	30	
No. of class variables	1	
Details of the class variable	Name: Result	
	−1 == Legitimate	1 == Phishing
	4898	6157
Total number of instances	11,055	

With a total of 11,055 instances, the dataset distribution is between two class labels vis–a–vis "-1" representing the legitimate website instances and "1" representing the phishing website instances. The total number of phishing website instances constitute the majority but does not totally dominate the data distribution as the legitimate website instances are over 44% of the data. Table 2 presents the attributes of the phishing website dataset.

3.2 Implemented Algorithms

This study proposes and implements three (3) novel phishing website detection models (PWDM) using the datasets discussed in the previous sub-section. The ForestPA algorithm was implemented and improved version of the same algorithm was carried out using meta-learning methods. The enhanced ForestPA via meta-learners is solely on improving the performance of ForestPA. Thus, three proposed phishing detection models were developed by implementing these algorithms vis–a–vis: (1) ForestPA, (2) AdaBoost, and (3) Bagging algorithms.

As described by Zhou et al. [18], the ForestPA algorithm promotes strong diversity by taking into consideration weight-related concerns which include but not limited to weight assignment strategy and weight increment strategy. It is a method that usually builds a set of highly accurate decision trees having exploited the strength that lies in all non-class attributes available in the given dataset. ForestPA had been previously used for developing IDS, a core feature in Network security, in a research carried out [17] and was also used with other heuristic techniques as carried out by [18].

Algorithmically, ForestPA randomly updates the weights of attributes that appear in the latest tree within a Weight-Range (WR) which is defined as

$$\mathrm{W\,R}^{\lambda} = \begin{cases} \left[0.0000, \mathrm{e}^{-\frac{1}{\lambda}}\right], \lambda = 1 \\ \left[\mathrm{e}^{-\frac{1}{\lambda-1}} + \rho, \mathrm{e}^{-\frac{1}{\lambda}}\right], \lambda > 1 \end{cases} \tag{1}$$

Table 2 Phishing Website Data Attributes

No.	Attributes
1	Having_IP_Address
2	URL_Length
3	Shortining_service
4	Having_At_Symbol
5	Double_slash_redirecting
6	Prefix_Suffix
7	Having_Sub_Domain
8	SSLfinal_State
9	Domain_registration_length
10	Favicon
11	Port
12	HTTPS_token
13	Request_URL
14	URL_of_Anchor
15	Links_in_tags
16	SFH
17	Submitting_to_email
18	Abnormal_URL
19	Redirect
20	On_mouseover
21	RightClick
22	popUpWindow
23	Iframe
24	age_of_domain
25	DNSRecord
26	web_traffic
27	Page_Rank
28	Google Index
29	Links_pointing_to_page
30	Statistical_report
31	Class

where $\succ$ represents the attribute level and ρ ensures that WR for the different levels is non-overlapping. In the light of addressing the negative effect of keeping weights that are absent in the latest tree, ForestPA implements a method of systemic increment of weights of the attribute that has not been tested in the subsequent trees. For example, an attribute A_i is tested at level ρ of the T_{j-1} − th tree with η height and its weight is w_i. Thus, calculating the weight increment value σ_i of A_i is:

$$\sigma_{\mathrm{i}} = \frac{1.0 - \mathrm{w_i}}{(\eta + 1) - \lambda} \tag{2}$$

As such, the ForestPA is a viable method for producing reliable and robust ML models. Hence, ForestPA was used in this study to build a phishing detection model (ForestPA-PWDM).

In addition, AdaBoost which is a meta-learner method sequentially applies weak single classifier to training the re-weighted training data. As revealed by [19], AdaBoost executes a majority vote at the end of its training phase for making its final decision having integrated all the weak hypotheses developed by the weak single classifiers into one and final hypothesis. Originally, AdaBoost was developed for binary classification purposes and thus provides the justification for the selection of the algorithm for detecting a phishing website.

Algorithm 1. The AdaBoost.M1 Algorithm

Input: Training set $S = \{x_i, y_i\}, i = 1 \ldots m, y_i \in Y, Y = \{c_1, c_2, \ldots, c_k\}, c_k$is the class label;

The number of Iterations T;

Weak classifier I.

1 Initializing weights distribution of $D_1(i) = 1/m$

2 For t = 1 to T

3 Train classifier $I(S, D_t)$, get a weak hypothesis

$h_t = X \rightarrow \{c_1, c_2, \ldots, c_k\}$

4 Compute the error rate of $h_t, \varepsilon_t \leftarrow \sum_{i=1}^{m} D_t(i)[y_i \neq h_t(x_i)]$

5 If ε_t> 0.5 then

6 $T \leftarrow t - 1$

7 Continue

8 End if

9 Set $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

10 For $i = 1\ to\ m$

11 Update weight $D_{t+1})i) = D_t(i)\beta_t^{1-[y_i \neq h_t(x_i)]}$

12 End for i

13 End for t

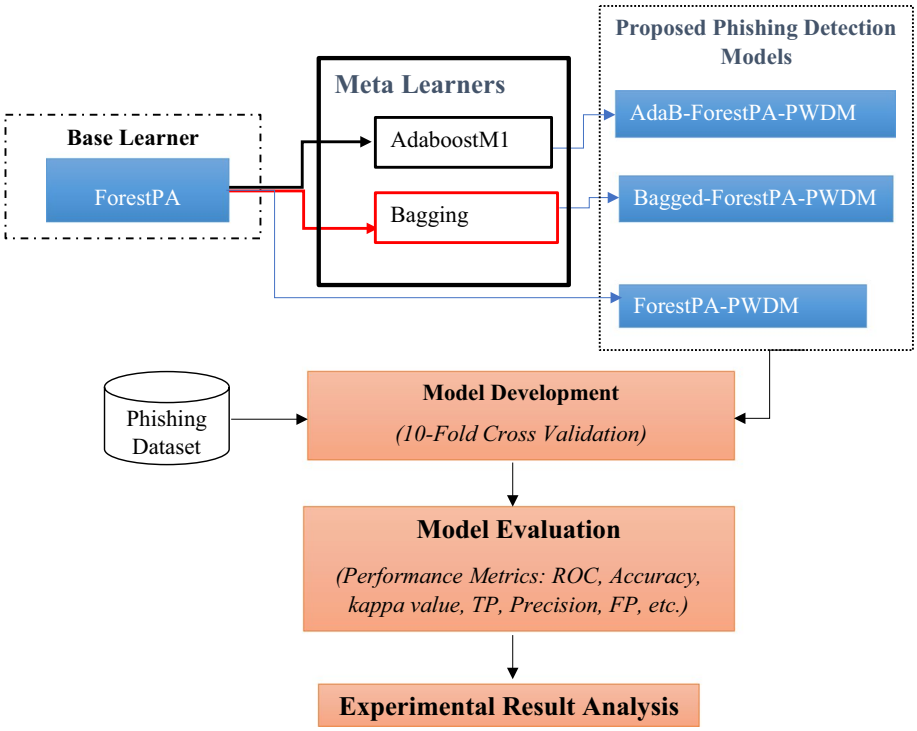
Output: the final hypothesis

$$H(x) = \arg max\left(\sum_{t=1}^{T} \ln\left(\frac{1}{\beta_t}\right)[Y \neq h_t(X)]\right)$$

In this study, an extended version of AdaBoost meta-leaner (AdaBoost.M1) was considered as used by [20]. AdaBoost.M1 algorithm, as outlined in Algorithm 1, was used in this study to develop an enhanced variation of the ForestPA model (AdaB-ForestPA-PWDM).

The bagging meta-learner method is a method whose base-learners, during its training phase, learn from the original dataset by extracting different subsets from the original dataset for fitting different models [21]. Bagging meta-learner ensures that the variance of each developed model is being reduced while keeping the bias of the same models from increasing by applying aggregation technique on all the developed models. According to [22], bagging meta-leaner executes a random resampling of the original dataset, develops multiple base classifiers by fitting models on the resampled subsets and then aggregates the models into a single model for the sole purpose of making predictions. The Bagging meta-leaner is presented in Algorithm 2.

Fig. 1 Experimental Framework



Algorithm 1. The Bagging Algorithm

Input: training set S,
Inducer I,
integer T (number of bootstrap samples).

1. for $i = 1$ to T {
2. S' = bootstrap sample from S (i.i.d. sample with replacement)
3. $C_i = I(S')$
4. }
5. $C^*(x) = \arg\max \sum_{i:C_i(x)=y} 1$ (the most frequently predicted label y)

Output: classifier C^*

Accordingly, this study proposes an enhanced ForestPA based on bagging meta-learner for phishing website detection model (Bagged-ForestPA-PWDM). Bagged-ForestPA-PWDM creates multiple ForestPA models on random subsets of the selected dataset and then aggregates the same models to produce a final model for the detection of phishing websites.

3.3 Experimental Framework

Using the three (3) different machine learning algorithms discussed above, three predictive models were developed after fitting the algorithms on the aforementioned datasets. Since it is known that model development is the next stage after the dataset and algorithm selection process and method

identification phases, the N-fold cross-validation model development method was implemented in this study.

In this phase, the proposed PWDMs are trained and evaluated accordingly as presented in Fig. 1. The proposed models were trained and tested using *N*-fold cross-validation method (in this case, $N=10$). *N*-fold cross-validation simply divides a given dataset into *N* partitions, trains with $N-1$ partitions of the data, and then tests the ensuing model with *N*th partition. This process is iterative is repeated for *N* times until all parts of the data are being used for both training and test. At the end of the iteration, the models are aggregated and evaluated mostly using weighted or average metric values.

According to the experimental framework (See Fig. 1), the proposed PWDMs are implemented in the *Phishing Detection Models* module. The ensuing models are evaluated based on the *Model Development* module. Tenfold cross-validation technique is used for fitting each proposed PWDMs on the phishing data accordingly. The performances of the ensuing models on the test data were assessed using selected evaluation metrics. Conclusively, a comparative performance analysis of the developed PWDMs model is being carried as well as a comparison with existing state-of-the-art methods.

The proposed PWDMs models were implemented using the WEKA Data mining tool. The respective parameters settings of the proposed models are presented in Table 3.

3.4 Performance Evaluation Metrics

Following the model development process stage, the developed models are evaluated. As such, the performances of models were evaluated using popular evaluation metric for this kind of study

This section presents the performance evaluation metrics used for measuring the efficacy of the proposed PWDMs in this study. In accordance with existing and related studies, accuracy, TP-rate, FP-rate, Precision (P), Recall (R), F-Measure, ROC and cohen's Kappa values were used for evaluating the performances of PWDMS [20, 23, 24]. The mathematical formulas for each metric are described as follows:

(1) Accuracy: is the percentage of all correctly classified phishing websites.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

(2) Recall: is the total number of phishing websites that are correctly classified.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

(3) Precision: is the number of predicted phishing websites that are actually phishing websites.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

(4) F-measure: is the weighted harmonic mean of the precision and recall of the test. The best value will be at 1 and worst at 0 value.

$$F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

(5) Cohen's Kappa: is a chance-corrected measure calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than 0 means that the classifier is doing better than chance

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{7}$$

More so, the confusion matrix [25] was also used for evaluating the performances of the PWDMs as shown in Table 4. Also, the inherent metrics obtained through the confusion matrix were also used such as the true positive rate (TP rate) and the False Positive Rate (FP Rate). The confusion matrix is presented below

(6) True Positive (TP) rate: refers to the rate at which actual phishing website instances are correctly classified as that phishing website.

Table 3 Parameter setting of the proposed PWDMs

PWDMs	Parameter settings
ForestPA-PWDM	Batchsize = 100; numberofTrees = 10; seed = 1; simpleCartMinimumRecords = 2, simpleCartPrunninFolds = 2
AdaB-ForestPA-PWDM	AdaBoost.M1: batchsize = 100; classifier = ForestPA; numIterations = 10; seed = 1; weightedThreshold = 100 ForestPA: batchsize = 100; numberofTrees = 10; seed = 1; simpleCartMinimumRecords = 2, simpleCartPrunninFolds = 2
Bagged-ForestPA-PWDM	Bagging.M1: bagsizePercent = 100; batchsize = 100; classifier = ForestPA; numExectionslots = 1;; numIterations = 10; seed = 1. ForestPA: batchsize = 100; numberofTrees = 10; seed = 1; simpleCartMinimumRecords = 2, simpleCartPrunninFolds = 2

Table 4 Confusion Matrix

		Predicted class	
		Legitimate website	Phishing website
Actual class	Legitimate website	True negative (TN)	False positive (FP)
	Phishing Website	False negative (FN)	True positive (TP)

$$TP = \frac{TP}{TP + FN} \tag{8}$$

(7) False Positive (FP) rate: is the value of the incorrectly classified legitimate websites as a phishing website

$$FP = \frac{FP}{FP + TN} \tag{9}$$

(8) Receiver Operating Characteristic (ROC) Curve: is not susceptible to the majority class bias and does not ignore the minority class during its evaluation. It plots the FP rate on the X-axis and plots the TP rate on the Y-axis.

4 Experimental Results

Having implemented the proposed framework of this research, the results are being reported for each developed model starting from ForestPA-PWDM to both of its enhanced variations (i.e., AdaB-ForestPA-PWDM and Bagged-ForestPA-PWDM). Reporting the results for the ForestPA model, Tables 5 and 6 present the performance scores of the model and its corresponding confusion matrix, respectively.

From Table 5, it is seen that the ForestPA-PWDM produced an accuracy of 96.26% while having a TP rate of 0.973 and an FP rate of 0.050. The model did highly better than chance with a kappa score of 0.92. Also, the F-measure score of 0.967, having scored a recall value of 0.973 and a precision of 0.961, and a ROC score of 0.994 strongly indicates that the ForestPA-PWDM possess highly strong predictive prowess for determining both the majority (phishing) and minority (legitimate) class without bias. More so, the confusion matrix (as shown in Table 6) revealed that 5989 of the 6159 phishing websites were correctly classified as well as 4653 of 4898 legitimate websites were also correctly classified by ForestPA-PWDM.

Table 6 ForestPA-PWDM Confusion Matrix

		Predicted class	
		Legitimate website	Phishing website
Actual class	Legitimate website	4653	245
	Phishing website	168	5989

As previously mentioned, improving the performance of ForestPA-PWDM was sought after in this study. The result of Bagged-ForestPA-PWDM implementation is being discussed and presented in Tables 7 and 8, respectively. Once again, Bagged-ForestPA-PWDM is the implementation of the Bagging meta-learner algorithm which made use of ForestPA as its base learner.

From Table 7, it is seen that the Bagged-ForestPA-PWDM produced an accuracy of 96.58% while having a TP rate of 0.978 and ann FP rate of 0.049. The model did highly better than chance with a kappa score of 0.93. Also, the F-measure score of 0.97, having scored a recall value of 0.978 and precision of 0.962, and a ROC score of 0.995 strongly indicates that the Bagged-ForestPA-PWDM possess the stronger predictive capability for correctly classifying both the majority (phishing) and minority (legitimate) class without bias. More so, Table 8 illustrates the confusion matrix of the model which had 6019 of the 6159 phishing websites correctly classified as well as 4658 of 4898 legitimate websites being correctly classified by Bagged-ForestPA-PWDM.

Lastly, the result of AdaB-ForestPA-PWDM implementation is being discussed. Once again, AdaB-ForestPA-PWDM

Table 5 ForestPA-PWDM evaluation scores

Evaluation metric	Score
Accuracy (%)	96.2641
Kappa	0.9242
TP rate	0.973
FP rate	0.050
Precision	0.961
Recall	0.973
F-measure	0.967
ROC	0.994

Table 7 Bagged-ForestPA-PWDM evaluation scores

Evaluation metric	Score
Accuracy (%)	96.581
Kappa	0.9306
TP Rate	0.978
FP Rate	0.049
Precision	0.962
Recall	0.978
F-measure	0.970
ROC	0.995

Table 8 Bagged-ForestPA-PWDM Confusion Matrix

		Predicted class	
		Legitimate website	Phishing website
Actual class	Legitimate website	4658	240
	Phishing website	138	6019

Table 9 AdaB-ForestPA-PWDM evaluation scores

Evaluation metric	Score
Accuracy (%)	97.4029
Kappa	0.9473
TP rate	0.981
FP rate	0.035
Precision	0.973
Recall	0.974
F-measure	0.974
ROC	0.996

Table 10 AdaB-ForestPA-PWDM Confusion Matrix

Actual class	Predicted Class	
	Legitimate website	Phishing website
Legitimate website	4729	169
Phishing Website	118	6039

is the implementation of the AdaBoost.M1 algorithm that used ForestPA as its base learner. The evaluation scores of the model are being presented in Tables 9 and 10, respectively.

The AdaB-ForestPA-PWDM showed an excellent predictive strength with an accuracy score of 97.40% and a ROC score of 0.996. These scores indicate the massive categorization prowess of the model with respect to both classes without bias. The kappa score of 0.9473 also signifies that the predictive strength of this model was not made out of chance but of intensified learning of the input data by the model. The TP rate of 0.981 showed the great strength of the model in detecting the phishing website, likewise, the FP rate of 0.035 reflects the ability of the model to drastically abate the problem of false notification of legitimate website as a phishing website. Also, the f-measure value of 0.974 (having produced a precision score of 0.973 and recall score of 0.974) supports the high predictive capability of AdaB-ForestPA-PWDM to ascertain if a website is ether legitimate or phishing.

5 Discussion

In this section, the reported results of the results will be discussed. The discussion will compare the reported performance of this study among themselves and against existing methods was reviewed in the related work section. Table 11 provides a tabular comparative analysis of this study and the existing methods.

As seen, Table 11 presents a tabular comparative analysis of the developed PWDMs of the study as well as with other reviewed related existing methods. Having implemented the ForestPA algorithm (i.e. both as a single classifier and with two (2) enhanced variations (1) Bagging and (2) Boosting methods), as the proposed PWDMs which were evaluated and reported, it is, therefore, necessary to discuss the results. Beforehand, it is important to comparatively analyse and discuss the PWDMs of this study. While the ForestPA-PWDM does not produce a sub-standard model on its own, without gainsaying the two (2) implemented meta-learner approaches produced better models when evaluated across all the performance evaluation metrics as depicted in Figs. 2 and 3.

It is noteworthy to highlight that in the light of this study, machine learning algorithms are highly competent in ascertaining whether a website is either legitimate or phishing. Through this study, it is evident that the simple implementation of appropriate machine learning algorithms for a defined problem is better than implementing complex and or hybridized algorithms. Often time, the implementation of deep learning for finding solutions to some problem is inappropriate—as the case of this phishing website detection, where simple machine learning algorithms will outperform deep learning methods (as seen in Table 11) because deep learning methods mainly performs on big data with multi-dimensions and tens of thousands of instances. Also, the development of complex model through hybridization of various feature selection technique and various stand-alone machine learning algorithm by [14] produced an accuracy of 97% (values of other performance metrics were not reported) which extremely competes with of 97.404% accuracy. However, the computational cost of [14] high performing method puts it at loss against the AdaB-ForestPA-PWDM. More so, the problem of parametrization of each of the four (4) algorithms used by [14] model is yet another detriment as compared to this study's AdaB-ForestPA-PWDM.

The simple implementation of ForestPA for PWDM (ForestPA-PWDM) produced an accuracy of 96.26%, 0.963 TPR, 0.04 FPR, and ROC of 0.994 outperformed various existing phishing website detection methods that implemented deep learning such as the study of [4] of 87.61% and the DNN implementation of [3] with 88.77% accuracy with TPR of 0.858, and also the [3] DNN with GA-based features

Table 11 Comparative Analysis of Existing Phishing Website Detection Model

Methods	Author	Accuracy (%)	TP Rate	FP Rate	F-measure	ROC curve
ForestPA-PWDM	*	**96.264**	**0.963**	**0.04**	**0.967**	**0.994**
Bagged-ForestPA-PWDM	*	**96.581**	**0.966**	**0.037**	**0.970**	**0.995**
AdaB-ForestPA-PWDM	*	**97.404**	**0.974**	**0.028**	**0.974**	**0.996**
PCA + NN + RF + bagging	[14]	97.4	–	–	–	–
Random forest and wrapper feature selection	[2]	97.295	0.973	0.03		0.996
DNN	[3]	88.77	85.83	–		–
DNN with GA-based features weighting	[3]	91.13	90.79	–		–
FRS + random forest algorithm	[9]	–	–	–	0.95	–
Decision table	[2]	93.243	0.932	0.75		0.979
SMO	[2]	93.804	0.938	0.066		0.936
Naïve Bayes	[2]	92.981	0.93	0.076		0.981
Logistics regression	[7]	94.01	–	–		–
TDLBA/TDLHBA	[7]	96.5	–	–		–
ANN–MLP	[4]	87.61%	–	–		–
ANN	[13]	96.91	–	–		0.995
CART	[13]	95.79	–	–		0.981
Rotation forest	[13]	96.79	–	–		0.994

The symbol (*) and bold typeface represents the method implemented by this study

The symbol (-) indicates that the data are unavailable

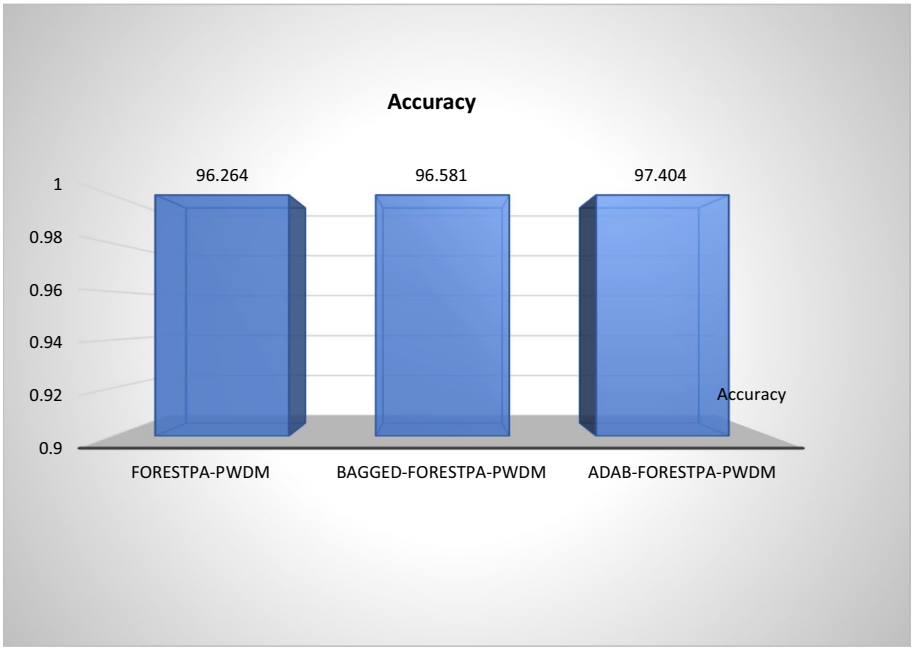


Fig. 2 Accuracies of the developed PWDMs of this study

weighting implementation which had 91.13% accuracy with 0.908 TPR. This evidence reinforced the high predictive capability of machine learning and the often-inappropriate usage of a deep learning algorithm. Also, ForestPA-PWDM outperformed quite a number of machine learning implementation such as the CART PWDM produced by [13] with an accuracy of 95.79% with 0.981 ROC, [2] decision table of 93.24% with 0.75 FPR, SMO of 93.804% with 0.936ROC, Naïve Bayes of 92.98% accuracy with 0.76FPR, and [7] logistic regression PWDMs of 94.01% accuracy. This as evidence strengthens the superior predictive prowess of ForestPA-PWDM over existing machine learning phishing website detection methods and also provides answer to the first and fourth research questions of this study.

In addition, the application of FRS feature selection and Random Forest classification algorithm by [9] produced an

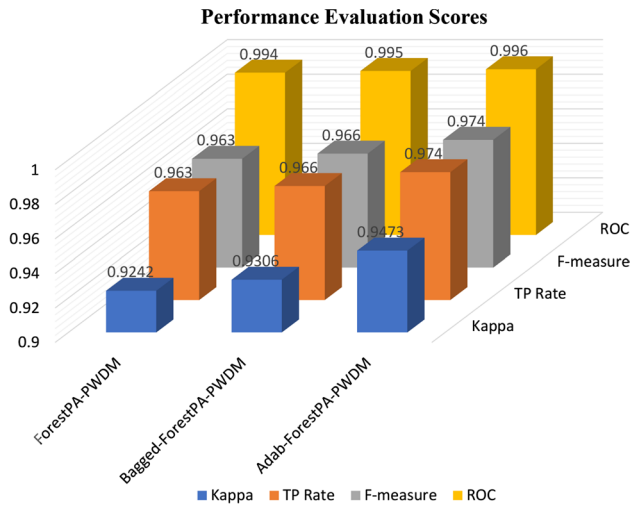


Fig. 3 Pictorial representation of some performance evaluate scores of all models

F-measure value of 95 (i.e. 0.95) which is outperformed by the simple ForestPA-PWDM implementation as well as its enhanced variations. While other existing methods such as [13] Rotation Forest PWDM of 96.79% accuracy with 0.994 ROC, [7] TDLBA/TDLHBA of 96.5% accuracy and [2] Wrapper-based Random Forest PWDM of 97.25% outperformed out ForestPA-PWDM of this study, the improved PWDMs, particularly the AdaB-ForestPA-PWDM, outperformed existing methods having produced an accuracy of 97.404%, TPR of 0.974, FPR of 0.028, and ROC curve value of 0.996. Also, this as evidence provides an answer to the second and third research questions of this study.

6 Conclusion

Paying full attention to the results and discussion sections, this research work revealed answers to several research questions. In response to the first question, the ForestPA algorithm was implemented and used to fit the ForestPA-PWDM. The result of which was able to detect phishing and legitimate website with a ROC curve value of 0.994, the accuracy of 96.26% and FPR of 0.04. This indicates that the ForestPA algorithm effectively detects either website types with very high accuracy with a bias to the majority class and with very little false alarm rate.

Answering the second research question, it was discovered that bagging meta-learner improves ForestPA and was also effective in detecting legitimate and phishing websites. The implementation of the Bagging meta-learner method by using ForestPA as base-learning produced the Bagged-ForestPA-PWDM whose performance did better than the ForestPA-PWDM. The effectiveness of the Bagged-ForestPA-PWDM is seen having produced a better accuracy of 96.581%, TPR of 0.966, PR of 0.037, and ROC of 0.995—all better than the ForestPA-PWDM performance.

In response to the third research question, this study revealed that as good as both the ForestPA-PWDM and Bagged-ForestPA-PWDM can be, the Boosting Meta-learner method surely provides superior performance. For the purpose of detecting phishing and legitimate websites, the Boosting meta-learner method (AdaB-ForestPA-PWDM) improved upon the ForestPA implementation by increasing its accuracy to 97.404%, TPR to 0.974, and ROC curve to 0.9966 while also further reducing the FPR to 0.028 which means that in real-time, false alarm notifications are next to zero.

Lastly, the answer to the fourth research question is extensively provided in the discussion section. Concisely, the phishing website detection models of this study comparatively outperformed various existing methods. With the ForestPA-PWDM outperforming more than half of the existing methods, the AdaB-ForestPA-PWDM classically outperformed all existing methods. Thus, the development and deployment of the developed PWDMs of this study as software for real-time detection of attack are considered as an important future work. More so, the hybridization of the methods with high performing feature selection method is considerable future work.

References

1. Yang, P.; Zhao, G.; Zeng, P.: Phishing website detection based on multidimensional features driven by deep learning. IEEE Access **7**, 15196–15209 (2019)
2. Abdulrahaman, M.D.; Alhassan, J.K.; Adebayo, O.S.; Ojeniyi, J.A.; Olalere, M.: Phishing attack detection based on random forest with wrapper feature selection method. Int. J. Inf. Process. Commun **7**(2), 209–224 (2019)
3. Ali, W.; Ahmed, A.A.: Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting. IET Inf. Secur. **13**(6), 659–699 (2019)
4. Ferreira, R.P.; et al.: Artificial neural network for websites classification with phishing characteristics. Soc. Netw. **7**, 97–109 (2018)
5. Wei, B.; et al.: A deep-learning-driven light-weight phishing detection sensor. Sensors **19**(19), 4258 (2019)
6. Soon, G.K.; Chiang, L.C.; On, C.K.; Rusli, N.M.; Fun, T.S.: Comparison of ensemble simple feedforward neural network and deep learning neural network on phishing detection. Lect. Notes Electr. Eng. **603**, 595–604 (2020)
7. Vrbančič, G.; Fister, I. Jr.; Podgorelec, V. (2018) Swarm intelligence approaches for parameter setting of deep learning neural network : case study on phishing websites classification. In: Int. Conf. Web Intell. Min. Semant.,
8. Gajera, K.; Jangid M.; Mehta, P.; Mittal, J. A novel approach to detect phishing attack using artificial neural networks combined

with pharming detection. In: Proc. 3rd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2019, pp. 196–200 (2019).
9. Zabihimayvan, M.; Doran, D. Fuzzy rough set feature selection to enhance phishing attack detection. In: IEEE Int. Conf. Fuzzy Syst., vol. 2019-June (2019)
10. Zhu, E.; Liu, D.; Ye, C.; Liu, F.; Li, X.; Sun, H. Effective phishing website detection based on improved BP neural network and dual feature evaluation. In: IEEE Intl Conf Parallel Distrib. Process. with Appl. Ubiquitous Comput. Commun. Big Data Cloud Comput. Soc. Comput. Networking, Sustain. Comput. Commun., pp. 759–765, (2018).
11. Singh, C.; Smt. Meenu, Phishing website detection based on machine learning: a survey. In: 6th International Conference on Advanced Computing & Communication Systems (ICACCS), 2020, pp. 398–404.
12. Mohammad, R.M.; Thabtah, F.; McCluskey, L.: Predicting phishing websites based on self-structuring neural network. Neural Comput. Appl. **25**(2), 443–458 (2014)
13. Subasi, A.; Molah, E.; Almkallawi, F.; Chaudhery, T. J.; Intelligent phishing website detection using random forest classifier. In: Int. Conf. Electr. Comput. Technol. Appl., vol. IEEE, pp. 1–5 (2017)
14. Zamir, A.; et al.: Phishing web site detection using diverse machine learning algorithms. Electron. Libr. **38**(1), 65–80 (2020)
15. Mazini, M.; Shirazi, B.; Mahdavi, I.: Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms. J. King Saud Univ. Comput. Inf. Sci. **31**(4), 541–553 (2019)
16. Mohammad, R. M.; Thabtah, F.; McCluskey, L. An assessment of features related to phishing websites using an automated technique. 2012 IEEE, 2012. In: International Conference for Internet Technology and Secured Transactions., 2012, pp. 492–497.
17. Panigrahi, R.; Borah, S.: Dual-stage intrusion detection for class imbalance scenarios. Comput. Fraud Secur. **2019**(12), 12–19 (2019)
18. Zhou, Y.; Cheng, G.; Jiang, S.; Dai, M. An efficient intrusion detection system based on feature selection and ensemble classifier," arXiv Prepr. arXiv1904.01352., 2019.
19. Sun, B.; Chen, S.; Wang, J.; Chen, H.: A robust multi-class AdaBoost algorithm for mislabeled noisy data. Knowledge Based Syst. **102**, 87–102 (2016)
20. Haixiang, G.; Yijing, L.; Yanan, L.; Xiao, L.; Jinling, L.: BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification. Eng. Appl. Artif. Intell. **49**(October), 176–193 (2016)
21. Collell, G.; Prelec, D.; Patil, K.R.: A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data. Neurocomputing **275**, 330–340 (2018)
22. Lee, S.J.; Xu, Z.; Li, T.; Yang, Y.: A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making. J. Biomed. Inform. **78**, 144–155 (2018)
23. David, J.; Thomas, C.: Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic. Comput. Secur. **82**, 284 (2019)
24. Balogun, A.O.; Oladele, R.O.; Mojeed, H.A.; Amin-balogun, B.; Adeyemo, V.E.; Aro, T.O.: Performance analysis of selected clustering techniques for software defects prediction. African J. Comput. ICT **12**(2), 30–42 (2019)
25. Niyaz, Q.; Sun, W.; Javaid, A. Y.; Alam, M A deep learning approach for network intrusion detection system. In: Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS) (2016)