IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 19, 2020, accepted July 29, 2020, date of publication August 3, 2020, date of current version August 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013699

AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites

YAZAN AHMAD ALSARIERA[1], VICTOR ELIJAH ADEYEMO[2], ABDULLATEEF OLUWAGBEMIGA BALOGUN[3,4], (Member, IEEE), AND AMMAR KAREEM ALAZZAWI[3]

[1]Department of Computer Science, Faculty of Science, Northern Border University, Arar 73222, Saudi Arabia
[2]School of Built Environment, Engineering, and Computing, Leeds Beckett University, Leeds LS6 3QS, U.K.
[3]Department of Computer and Information Sciences, Faculty of Science and IT, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[4]Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin, Ilorin 1515, Nigeria

Corresponding author: Yazan Ahmad Alsariera (yazan.ahmad@nbu.edu.sa)

ABSTRACT Phishing is a type of social web-engineering attack in cyberspace where criminals steal valuable data or information from insensitive or uninformed users of the internet. Existing countermeasures in the form of anti-phishing software and computational methods for detecting phishing activities have proven to be effective. However, new methods are deployed by hackers to thwart these countermeasures. Due to the evolving nature of phishing attacks, the need for novel and efficient countermeasures becomes crucial as the effect of phishing attacks are often fatal and disastrous. Artificial Intelligence (AI) schemes have been the cornerstone of modern countermeasures used for mitigating phishing attacks. AI-based phishing countermeasures or methods possess their shortcomings particularly the high false alarm rate and the inability to interpret how most phishing methods perform their function. This study proposed four (4) meta-learner models (AdaBoost-Extra Tree (ABET), Bagging –Extra tree (BET), Rotation Forest – Extra Tree (RoFBET) and LogitBoost-Extra Tree (LBET)) developed using the extra-tree base classifier. The proposed AI-based meta-learners were fitted on phishing website datasets (currently with the newest features) and their performances were evaluated. The models achieved a detection accuracy not lower than 97% with a drastically low false-positive rate of not more 0.028. In addition, the proposed models outperform existing ML-based models in phishing attack detection. Hence, we recommend the adoption of meta-learners when building phishing attack detection models.

INDEX TERMS Artificial intelligence (AI), cyber security, extra trees, phishing, phishing website detection, meta – learners.

I. INTRODUCTION

Cybersecurity refers to the management and development of technologies, tools, and techniques required for the protection of data, devices, and information [1]. It covers various aspects of computer and network security including Intrusion Detection System (IDS), Anti-virus, Phishing etc. Phishing is a nefarious cyber-attack plaguing the digital world with a direct impact on the physical world. Phishing is now a well-known subject-matter and the effect of successfully conducted phishing attacks is known to be disastrous. Hence, phishing disastrous outcomes emphasized the imperative need for developing effective and efficient solutions or methods to curb it [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Mercaldo.

It is a known fact that the occurrence of phishing attacks is no longer limited to SMS, pop-ups, and e-mails but also spreads to QR codes and spoof mobile applications [3]. A number of latest phishing techniques are hosted or linked to websites [4]. Consequently, phishing detection solutions are being produced which are broadly categorized into (i) list based; (ii) heuristics; and (iii) machine learning (ML) methods [5]. The evolving nature of phishing attacks requires viable and improved methods for its detection as there is no silver bullet for phishing elimination [6], [7]. However, the machine learning solutions (which is the premise for this study) proved to handle dynamic phishing attacks better than other methods. Through the review of literature, most of the existing ML methods suffer from various limitations such as high false alarm rate, low detection rate, and the inability of single classifiers and some hybridized methods

to produce highly effective and efficient phishing website detection solutions [8]–[13]. On this note, this study proposes novel meta-learner models based on the extra-tree algorithm. As a result, the key contributions of this study to the body of knowledge as related to solving the highlighted issues are:

1. The usage of a comprehensively featured and more recent phishing website data.
2. Proposal, development and implementation of Ada-Boost-Extra Tree (ABET), Bagging –Extra tree (BET), Rotation Forest – Extra Tree (RoFBET) and Logit-Boost-Extra Tree (LBET) phishing website detection methods.
3. Comparative analysis of recent ML phishing website detection methods against the proposed methods of this study.

The remaining part of this article is organized as follows. Section II discusses the review of related works. Section III illustrates the research methodology. Section IV presents the experimental settings, results analysis and discussion, and comparative analysis with existing methods. Lastly, Section V draws conclusions and indicate future works.

II. RELATED WORKS

In this section, the phishing attack will be discussed in the context of cybersecurity. In addition, a comprehensive review of existing methods deployed for the detection of phishing activities is discussed.

As revealed by Ferreira *et al.* [14], phishing was described as a technique for perpetrating online fraud by criminals through the usage of the Internet. Criminals sought to steal personal information, security credentials, and even bank details and password fraudulently by employing phishing techniques. In a simple statement, the term phishing can be explained with the analogy of fishing i.e. phishing is the act wherein internet criminals go to ''fish'' personal information as well as financial information of victims who ''took the baited hook'' that was released by a phisher (''fisherman'').

Nowadays, phishing is considered a fast-growing threat in cybersecurity and thus countermeasures are being developed to detecting phishing activities. Phishing is carried out via email and or websites that contain malicious content for stealing information from an uninformed or inattentive user of the Internet. As a result of the fast-growing and evolving nature of phishing, existing countermeasures are being developed in three different approaches vis-à-vis education, legal and technical approaches [3].

This study is based on a technical (i.e. ML) approach for detecting phishing websites and thus, existing technical countermeasures as related to this research works are being reviewed.

As presented by Subasi *et al.* [2] research work, random forest (an ensemble of decision tree classifiers) algorithm was used to develop an intelligent model for phishing website detection. It was evaluated using ROC curve, accuracy, and f-measure metrics [4]. The developed method was compared against models developed using k-NN, SVM, ANN, Rotation Forest, C4.5, CART and NB algorithms (which are known to be single classifiers and can be used as base learners for ensemble methods). The Random Forest model yielded the best model as expected with an accuracy of 97.36%, f-measure score of 0.974 and AUC value of 0.996. The implementation was carried out using the phishing website dataset sourced from the UCI repository. The limitation of the study is comparing the performance of an ensemble method – Random Forest Model, against single classifiers which are known to produce less effective models when compared against ensemble methods. More so, the chance for improving the method performance exist.

The research conducted by Alqahtani [3] produced a novel method for detecting phishing websites. This method was referred to as Phishing Websites Classification using Association Classification (PWCAC). PWCAC was novel as it makes use of the association rule induction technique to categorize whether a website is genuine or a phishing website. Using the phishing website dataset developed by [15], The PWCAC algorithm was used to develop a phishing website detection model and then its performance was evaluated. The PWCAC model achieved an accuracy of 95.20% and an F-measure score of 95.11%. The performance of the novel PWCAC model outperformed the like of C4.5, RIPPER, CBA, MAC models as reported in their work while the improvement of the PWCAC performance is required.

The research work of Yang *et al.* [4] presented a novel method for phishing website detection which was referred to as Dynamic Category Decision Algorithm (DCDA) having proposed and used multidimensional feature phishing detection (MFPD) approach. The research work made use of a deep learning implementation particularly the CNN (convolutional neural network)- Long Short Tern Memory (LSTM) algorithm to pre-process data and extract local features that are correlated as well as context-dependency in order to classify a website as legitimate or phishing. After which the classification result of the CNN-LSTM model was added as an attribute of an existing multi-dimensional feature dataset and supplied as input to an XGBoost classification algorithm for fitting a final model for detecting a phishing website.

Having implemented their algorithms, Yang *et al.* [4] conducted their research experiments on real-life data collected from the Internet via the *PhishTank* website (for phishing website) and *dmoztools.net* (for legitimate websites). The developed model performance was evaluated using accuracy, false-positive rate, false-negative rate, and cost. The MFPD approach of developing XGBoost yielded an accuracy of 98.99%, a false positive rate of 0.59, a false negative rate of 1.43 and cost 1.4. It outperformed the traditional XGBoost method, the CNN-LSTM method and some other existing method that was compared with it. The limitation of the research carried out by Yang *et al.* [4] lies in its reported high false-positive rate.

The research work conducted by Mohammad *et al.* [9] made use of a self-structuring neural network on the UCI

phishing website dataset. The model was developed via training, validation, and test split method. Thus, the best-produced model of the research work had 92.48% accuracy on the test set, 91.12% accuracy of the validation set, MSE score of 0.0280 and a learning rate of 0.5799. Clearly, the accuracy of the model produced by Mohammad *et al.* [9] study is too low considering the dire effect of successful phishing attacks and thus requires major improvement which this study set out to do.

A Content-Based Associative Classification method for phishing detection was presented by Dedakia & Mistry [11]. The research work improved upon the Multi-Label Class Associative Classification (MCAC) algorithm by considering the content-based features. Thus, the research work mainly focused on extracting new features which are spelling error, pop-ups window usage, copied website, right-click disabled, and form usage with submit button. With the additional features, the improved MCAC was implemented and evaluated yielding an accuracy of 94.29%. Also, the accuracy of the MCAC is comparatively low and required further improvement.

A hybrid phishing detection method was developed by Ali & Ahmed [12]. This method is a hybridization of an evolutionary algorithm and a deep neural network. The implemented evolutionary algorithm in their research work was a genetic- algorithm (GA) technique which was used to find the highly informative features from the original feature sets. Also, the authors opted for a fully connected feedforward neural network applied by H_20 having justified their selection by stating that it performs better on tabular (transactional) data than CNN or RNN algorithms which are great only on image and sequential data. The performance of the implemented model was evaluated using the following metrics: accuracy, sensitivity (i.e. TPR), specificity (TNR) and geometric mean (GM). The UCI dataset [15] was used and the resulting model performance was compared against C4.5, kNN, SVM, back-propagation neural network and the Naïve Bayes (NB) classifier. The model developed by Ali and Ahmed [12] produced an accuracy of 88.77%, a sensitivity of 85.82%, a specificity of 93.34%, and a GM of 89.50% which outperformed other compared models. As it is clearly indicated by the reported results, the hybridization of the evolutionary algorithm and the deep neural network had low detection rate as its major shortcoming. With sensitivity and accuracy values lower than 90% significantly points out the limit to which the method can detect phishing website.

An improved hybrid of Back-Propagation neural network (BPNN) and dual feature evaluation for detecting phishing websites were presented by Zhu *et al.* [16]. The research work presented the DF.GWO-BPNN by using the grey wolf algorithm to optimize the neural network constructed by the BP and then make use of the dual feature method to evaluate the improved BPNN results. The DF.GWO-BPNN model was also evaluated and compared with existing methods (such as BPNN, SVM, PSO-BPNN, etc.) using accuracy, total false-negative rate, root mean square error and forward sample recognition rate (P.Acc) metrics. The proposed DF.GWO-BPNN produced an accuracy of 98.78%, a total false-negative rate of 0.65%, P.Acc rate of 98.70% and an RMSE value of 36.59 which is the best model presented by the research work.

The research work of Vrban i *et al.* [17] presented a swarm intelligence approach for setting the parameters of a deep learning neural network. The research work proposed and implemented two methods based on hybrid and modified bat algorithm which are members of the swarm intelligence family inspired by the character exhibited by micro-bats. The methods implemented were referred to as TLDBA/TLDHBA, whose responsibility lies in finding optimal parameters for the deep learning neural network. Also, the deep learning NN implemented was a feed-forward NN with two hidden layers that were fully-connected. The dataset used is the UCI machine learning repository phishing website dataset [15]. The performance of the implemented model was compared against Naïve Bayes (NB), Random Tree (RT), Logistic Regression (LR), and J48 classifier among others. The implementation of the proposed model in the study achieved a minimum accuracy of 94.4% and a maximum accuracy of 96.9% which was outperformed by the RT model implemented in the study as the RT model achieved a minimum accuracy of 96.9% and a maximum accuracy of 97.1%. As reported by their study [17], the major weakness of this study is that the implemented RT method – a single classifier model, in the study already produced a more usable model than the proposed TLDBA/TLDHBA methods of 94.4% accuracy, both in terms of minimum and maximum accuracy.

A Deep Belief Network (DBN) was also implemented to detect phishing websites by Verma *et al.* [10]. The DBN model extracts deep hierarchical representation from the given dataset by using Restricted Boltzmann machines (RBM) to develop its model. Finally, the model was fine-tuned by supervised gradient descent (i.e. a logistic regression classifier) in order to classify the input based on the last hidden layer output. The performance of the developed model was evaluated using the accuracy metric and it was able to achieve a 94.426% accuracy. Although, the performance of the model was compared against J48 and even Random Forest algorithm implementations and it outperformed both methods, the accuracy produced by the DBN method is comparatively low even when compared to accuracies of some existing methods.

The published work of Zabihimayvan and Doran [5] used fuzzy rough set feature selection method to enhance the performance of three ML algorithms: (1) Multiperceptron (ii) RandomForest and (iii) SMO algorithms for developing phishing website development model. The performance of the developed models was measured using F-measure score. The best model (i.e. RandomForest – a homogeneous ensemble) maximized an F-measure score of 95% as reported.

The recently published work of Zamir *et al.* [18] presents a framework for phishing websites detection using some stacking approaches involving various ML algorithms and also

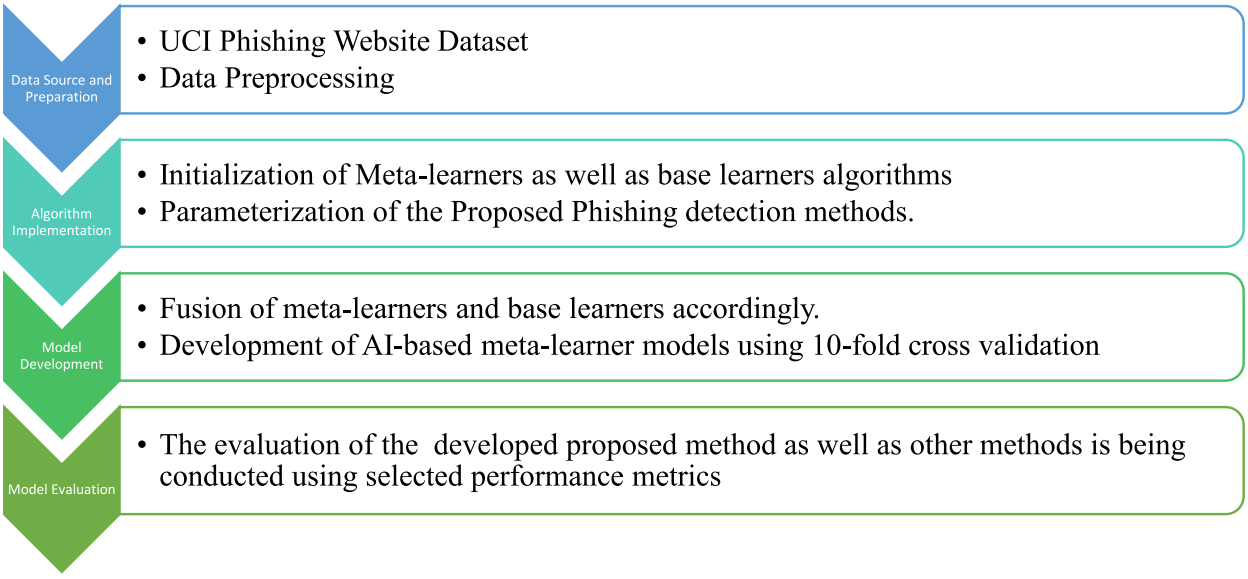

FIGURE 1. Overall research methodology.

diverse feature selection methods. The study experimented using the following feature selection technique: gain ratio, information gain, recursive feature elimination (RFE), principal component analysis and Relief-F. The machine learning algorithm implemented by the study includes RandomForest, support vector machine, bagging, neural network (NN), k-nearest neighbour and Naïve Bayes). These algorithms were hybridized following two different stacking methods for improvising the classifiers performance accuracy. The most performing ML implementation of the study (i.e. Stacked RF + NN + Bagging) was fitted on the most informative subset of the original created by the RFE feature selection method which yielded an accuracy of 97.4%.

The review of these relevant existing methods further established the identified problems aimed to be solved by this study. One of the problems is the inability of single classifier methods to highly detect evolving phishing websites. This became obvious through the usual outperformance of most single classifiers models by either ensemble methods or hybridized algorithms. Thus, it led to this study of finding a better method (i.e. AI meta-learners) to detect phishing effectively. As seen through review of existing methods, various methods vis-à-vis the deep learning methods, hybridized algorithms, and single classifier approach for detecting phishing websites mostly produce models that are of comparatively low accuracy while having relatively high false-positive rate. Thus, it becomes more expedient that this study is carried out for the proposal and implementation of AI meta-learners in order to produce models that effectively detect phishing websites with comparatively high accuracy while achieving a relatively low false positive and negative rates.

III. METHODOLOGY

In this section, the discussion of the overall research methodology and the experimental framework is made. Then, the description of the proposed ABET, RoFET, BET and LBET phishing website detection models, as well as a description of the dataset features, is also discussed. More so, the parameter settings of the phishing websites detection methods (including both meta-learners and base learners) were discussed briefly.

A. OVERALL RESEARCH METHODOLOGY

This study considers the phishing detection problem as an AI-based classification problem [13] wherein the result of the decision making phase leads to detecting if a given website is either a legitimate or a phishing website. Thus, consideration of the AI meta-learner algorithms as the basis for developing a credible and viable phishing website detection models to combat phishing threats and its evolving nature was made. These proposed AI meta-learner approach will serve as a solution to the identified problems with the existing methods that were reviewed. The selected meta-learners which are Bagging, AdaBoost, Rotation Forest, and LogitBoost, as well as one base learner (i.e. the Extra-tree algorithm), were the selected algorithms to be used in this study.

The overall methodology is broken into four (4) step-wise modules as depicted in Figure 1. The first module (data source and preparation) involves the obtainment and preparation of datasets for experimental purposes. The phishing website dataset is used in this study as it has been vastly used in existing studies (See Section II). It was developed by Mohammad *et al.* [15] and available on UCI and Kaggle databases. The phishing website dataset contains 11,055 instances, 30 independent features, and one (1) class attributes having two labels (''−1'' for a phishing website and ''1'' for a legitimate website).

The 30 independent features (See [15] for full details of features) are broadly distributed and categorized in four divisions namely:

i. HTML and JavaScript-based features (having 5 of the 30 features).
ii. Abnormal based features (having 6 of the 30 features).
iii. Domain-based features (having 7 of the 30 features).
iv. Address-Bar based features (having 12 of the 30 features).

In the second module, the selected AI algorithms (both meta-learners and base learners) were initialized with appropriated parameters in order to develop the proposed and other meta-learners' phishing detection models. Essentially, the 'number of iterations' parameter for all implemented meta-learners was set to 100. The third module saw the fusion of meta-learners and base-learners accordingly for the purpose of experimentation. Various metal-learners and base-learners were combined to fit AI models on the datasets whose results were then passed for evaluation. More importantly, the development of the AI-based meta-learner models was conducted using the N-fold cross-validation technique [19]. In this research work, N was set to 10. Thus, the model development process, considering 10-fold cross-validation, underwent the rigorous process of partitioning the datasets into ten (10) equal groups and then train on nine (9) of the partitioned data while testing on the remaining one (the tenth part). This process was iterated 10times and the test data were varied accordingly until all parts of the data are disjointly used for training and testing of the model.

The fourth module involves the evaluation of the developed model. Since the model was developed using a 10-fold cross-validation technique, the evaluation of the models' performances was based on a weighted average of the models over the 10 iterated folds of the cross-validation. The performances of the models' were evaluated using ROC, Accuracy, False Positive (FP) and F-measure as these metrics are widely used for evaluation of AI-based classification models and it is widely used to evaluate the research works that are closely related to this work as seen in Section II.

B. EXPERIMENTAL FLOW CHART

Having discussed the overall research methodology and its flow, the experimental flowchart which defines the interaction of the processes is being designed and depicted in Figure 2.

In Figure 2, the first component is the algorithms box which houses four (4) meta-learners vis-à-vis AdaBoost.M1 [20], 21], LogitBoost [22], [23], Bagging [24], [25] and Rotation Forest [26], [27] algorithms and the base -learner which is the Extra Trees [28]–[30] algorithm. The second component contained the proposed methods namely the AdaBoost.M1 and Extra trees, Bagging and Extra-trees, Rotation Forest and Extra trees, and lastly LogitBoost and Extra tree methods.

Each of these methods will be implemented and use to develop their respective models using the phishing website dataset – a total of four (4) distinct AI-based meta-learner methods for phishing detection. The development of these distinct models is a result of fitting the methods on the phishing website dataset using the 10-fold cross-validation technique as discussed in the previous sub-section.

At the completion of the development of each of the models, evaluation of the same was conducted using relevant performance metrics and thereafter the results of the models were analyzed. Analysis of the models' performances was carried out in order to compare them among themselves and with other existing methods or framework as reviewed in Section II.

C. PROPOSED ABET, RoFET, BET AND LBET PHISHING WEBSITE DETECTION METHODS

In this research work, the proposed phishing website detection methods are referred to as ABET, RoFET, BET, and LBET. ABET is the method that combines the AdaBoost.M1 meta-learner and the Extra-tree algorithm. ABET is a boosted Extra-tree, iterated 100 times over a sub-sampled dataset extracted from the original phishing website dataset. In the same vein, RoFET is an ensemble of extra-tree classifiers fitted on a transformed dataset via principal component filter while ensuring high accuracy and the reduction of bias.

BET is a bagged Extra-tree with 150 iterations. In other words, BET is a meta-learner that aggregates the results of 150 extra-tree over a bootstrapped dataset. LBET is a fusion of the LogitBoost meta-learner and the Extra-Tree algorithm. LBET in simple terms is the best fitted logistic regression of Extra-tree algorithm that handles both noise and outliers inherent in the given dataset.

1) ALGORITHMS

In this sub-section, the algorithms of the proposed models – ABET, RoFET, BET, and LBET Phishing Website Detection Models, will be discussed.

a: ABET ALGORITHM

This is the implemented ABET algorithm used for developing the ABET method. It is outlined in Algorithm 1.

b: RoFET ALGORITHM

The algorithm for developing the RoFET phishing website detection method is being outlined in Algorithm 2.

c: BET ALGORITHM

The algorithm outlined in Algorithm 3 is the algorithm used to fitting the Bagging Extra-tree phishing website detection model.

d: LBET ALGORITHM

The algorithm for the LBET meta-learner as used to detecting phishing website in this research work is being outlined in Algorithm 4.

D. PERFORMANCE EVALUATION METHODS

Since the type of classification carried out in this research work is known as binary classification (i.e. class attribute with

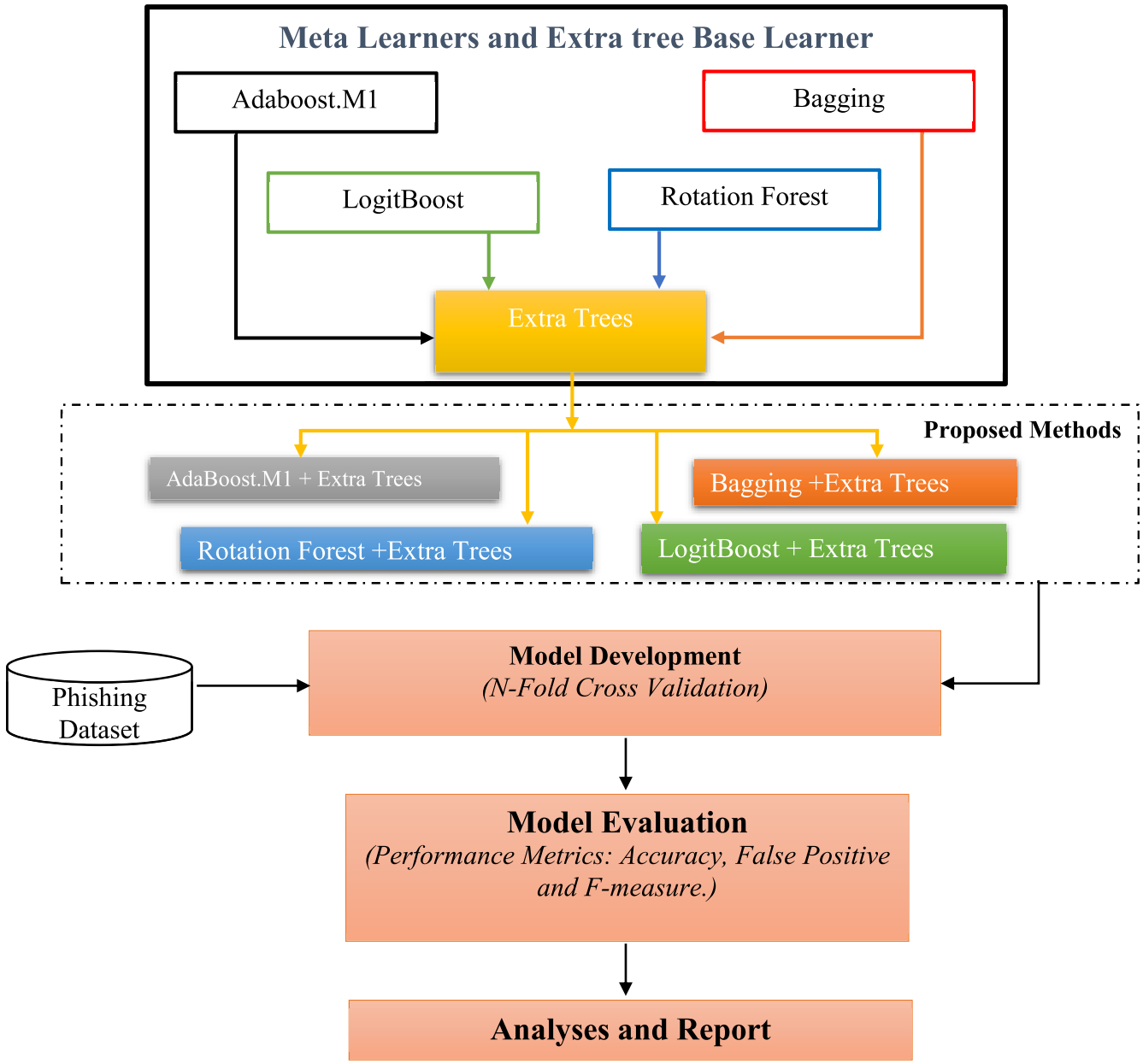

FIGURE 2. Experimental flowchart.

two (2) labels), the confusion matrix is used and values for each performance metric were calculated using the results obtained from each model confusion matrix.

The performances of the developed proposed models were evaluated using the following performances as widely used to evaluate existing methods for phishing website detection [3]. These metrics include Accuracy, False Positive (FP), False Negative (FN) and F-measure.

Accuracy measures the overall rate at which the actual labels of all instances are correctly predicted. It is calculated using (1):

$$Accuracy = \frac{\mathrm{TP} + TN}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{1}$$

False Positive measures the rate of good websites classified as a phishing website. It is calculated using (2)

$$FP = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{2}$$

False Negative is the rate of phishing websites classified as a good website. It is calculated using (3).

$$FN = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}} \tag{3}$$

F-measure is the weighted average of both the Recall (R) and Precision (P) metrics. It emphasizes how good a classifier is in maximizing both precisions and recall simultaneously. It is

Algorithm 1 The ABET Algorithm

Input: Training set S $= \{x_i, y_i\}, i = 1 \ldots m, y_i \in Y, Y = \{c_1, c_2\}, c_k$ is the class label;
The number of Iterations = 100;
Base Learner = ET.
1 Initializing weights distribution of $D_1\ (i) = 1/m$
2 For t = 1 to 100
3 Train classifier ET(S, D_t), get a weak hypothesis

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

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 $\varepsilon_t > 0.5$ 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

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

Algorithm 2 RoFET Algorithm

X = Training Set, Y = Class Label, and F = Attribute Set
ET = All Extra trees, $ET_1, ET_2, \ldots, ET_L$
Input: Training Data D = $\{x_i, y_i\}, x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$
1. X = $D \times n$ matrix.
1. K = 5, Then F is randomly divided into K distinct subsets while each subset must contain $N = 6$ number of features.
2. Select the corresponding columns of attributes in the subset $ET_{i,j}$ from the training dataset X, then form a new matrix $X_{i,j}$. Extract a bootstrap subset of objects 3/4 of X to make a new training dataset $X'_{i,j}$.
3. Use Matrix $X'_{i,j}$ as feature transform to produce the co-efficient in the matrix $P_{i,j}$, which jth column coefficient is the characteristic component jth.
4. Construct a sparse rotation matrix S_i using the obtained coefficient obtained in the matrix $P_{i,j}$.

Output: classifier ET_i of $d_{i,j}\left(XS_i^{\mathrm{f}}\right)$ to determine x belonging to the class y_i
Then, Calculate class confidence:

$$\alpha_j\ (x) = \frac{1}{ET} \sum_{i=1}^{L} d_{i,j}\left(XS_i^{\mathrm{f}}\right)$$

Assign the category with the largest $\alpha_j\ (x)$ value to x.

Algorithm 3 The BET Algorithm

Training Set = S
Base Learner (Inducer) = Extra-tree (ET)
Iteration (T) = 150
Input: S, ET, integer T.
1. for $i = 1$ to T {
2. $S' =$ bootstrap sample from S (i.i.d. sample with replacement)
3. $C_i = ET\left(S'\right)$
4. }
5. $C^*\ (x) = \arg\max \sum_{i:C_i(x)=y} 1$ (the most frequently predicted label y)

Output: classifier C^*

calculated using (4).

$$F - measure = \frac{2 \times \mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{4}$$

IV. EXPERIMENT AND RESULT ANALYSIS

In this section, the settings and the tools for conducting the experiments are being discussed. More so, the method for evaluating the developed proposed phishing website detection model was also covered in the discussion. Lastly, the results obtained from after evaluating the developed models' performance is being analysed.

A. EXPERIMENTAL SETTINGS

All proposed AI meta-learner models' for detecting phishing websites (i.e. ABET, RoFBET, BET, LBET) were developed having conducted the experiments on an Intel (R) Core (TM) i5-3230M CPU @2.60GHZ with 6GB RAM running the Windows 7 professional operating system.

As earlier stated, the dataset used is the widely used phishing website dataset created by [15]. The method of model development involved the application of 10-fold cross-validation. The Waikato Environment for Knowledge Analysis (WEKA) software was used for conducting all experiments, particularly version 3.8.1 which was run with a console. WEKA is a software created by [31] and released as open-source software. The Extra Tree algorithm was imported into the classifiers package.

B. RESULT ANALYSIS AND DISCUSSION

Following the development of all the proposed models and tentatively their evaluation, the performance of each model are analysed as follows:

1) PERFORMANCE EVALUATION OF ABET MODEL

The first model to be evaluated is the implemented ABET algorithm which was used to fit a model on the phishing dataset. The result of the evaluation is revealed in Table 1.

From Table 1, it is seen that the ABET model produced a very high accuracy of 97.485% while achieving a very

Algorithm 4 The LBET Algorithm

K = 100
N = 11,055
Base Learner = Extra-tree (ET)
1. Input data set N = $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots (x_n, y_n)\}$, where $(x_i \epsilon X)$ and, $y_1 \epsilon Y = \{-1, +1\}$
2. Number of iterations = K.
3. Initialized the weights $w_1 = 1/\mathrm{N}$, i = 1, 2, . . . , N;
4. Start ET function $f\ (x) = 0$ and probabilities estimates $\mathrm{P}(x_i) = 1/2$.
5. *Repeat for* $k = 1, 2, \ldots, K$:
   a. Calculate the weights and working response

$$w_i = p(x_i)(1 - p(x_i))$$
$$z_i = \frac{y_i - p(x_i))}{p(x_i)(1 - p(x_i))}$$

   b. Fit the function $fk(x)$ by a weighted least squared regression of z_i to x_i using weights w_i.
   c. Update

$$F\ (x) \leftarrow F\ (x) + \frac{1}{2} fk\ (x)$$

   and

$$p\ (x) \leftarrow \frac{e^{F(x)}}{e^{F(x)} + e^{-F(x)}}$$

6. Output the classifier:

$$LBET\ [F\ (x)] = LBET\left[\sum_{k=1}^{K} fk(x)\right]$$

TABLE 1. Performance evaluation of ABET's model.

Performance Metrics	ABET Model
Accuracy	97.485%
False Positive	0.016
False Negative	0.036
F-Measure	0.975

low false-positive value of 0.016 and false negative of 0.036. More so, the f-measure score of 0.0974 reveals how good the classifier is in detecting both labels of the class attributes.

2) PERFORMANCE EVALUATION OF RoFET MODEL

Following the implementation of RoFET algorithm, and fitting the same on the given dataset. The generated model was also evaluated and its result is being presented in Table 2.

As seen in Table 2, the RoFET model also yielded a highly predictive model for detecting phishing websites as it produced an accuracy of 97.449%. With an F-measure score of 0.974, a low positive value of 0.019 and a false negative value of 0.034, RoFET model is also a viable model for detecting phishing with lesser false alarm notification if finally implemented in real-time.

TABLE 2. Performance evaluation of RoFET's model.

Performance Metrics	RoFBET Model
Accuracy	97.449%
False Positive	0.019
False Negative	0.034
F-Measure	0.974

3) PERFORMANCE EVALUATION OF BET MODEL

BET's model is also being evaluated and the results obtained after the evaluation is presented in Table 3.

TABLE 3. Performance evaluation of BET's model.

Performance Metrics	BET Model
Accuracy	97.404%
False Positive	0.017
False Negative	0.038
F-Measure	0.974

BET's phishing website detection model produced an accuracy of 97.404% with a false positive score of 0.017 and a false negative value of 0.038. Finally, its efficiency in detecting both phishing and the non-phishing website is being evaluated and it resulted in an f-measure score of 0.974.

4) PERFORMANCE EVALUATION OF LBET MODEL

The fourth and final model to be evaluated in this research work is the LBET model. The LBET algorithm was fitted also on the given dataset and yielded the model whose performance was evaluated and thereby presented in Table 4.

TABLE 4. Performance evaluation of LBET's model.

Performance Metrics	LBET Model
Accuracy	97.576%
False Positive	0.018
False Negative	0.033
F-Measure	0.976

Table 4 revealed that the LBET model was able to achieve a very high predictive capability with an accuracy evaluated to score 97.576%. Also, the model's ability to raise the false alarm (i.e. its false-positive score) was evaluated and resulted in a low value of 0.018 while its false-negative score was 0.033. Finally, it established in dominance by yielding an f-measure value of 0.976.

Summarily, all the developed models of this research work achieved accuracy higher than 96% and produce a false positive rate lower than 0.02 and a false negative rate as low as 0.033 signifying the high predictive capabilities of all four models.

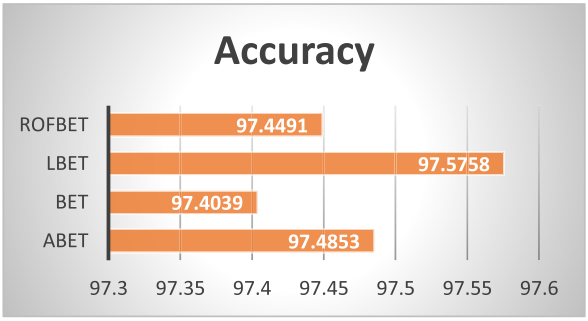

FIGURE 3. Accuracies of all developed models.

C. COMPARATIVE ANALYSIS OF PROPOSED METHODS WITH SOME EXISTING METHODS

As depicted in Figure 3 and 4, the LBET model outperformed all other models developed in this research work by producing the highest accuracy and f-measure scores as well as the lowest false-negative rate. Nevertheless, the ABET model had the lowest false positive rate of 0.016 which means in real-time application lowered notifications of false alarm if implemented.

In light of comparative analysis with existing methods, all implemented proposed methods of this research work produced accuracies that outperformed the content-based associative classification method presented by Dedakia and Mistry [11] which used the improved MCAC algorithm and achieved 92.48% accuracy of the test set. In the same vein, the accuracies of this research model outperformed the swarm intelligence DLNN method of Vrban i *et al.* [17] which produced 96.9% accuracy and outperformed other models compared against it such as NB, RT, LR, J48, etc.

The novel PWCAC method presented by Alqahtani [3] was also outperformed by this research in both accuracy and f-measure score. Although the MFPD approach of Yang *et al.* [4] research work produced a higher accuracy of 98.99%, it had a very high false-positive rate of 0.59 which undermines the efficiency of the method. The models of this research work had as its highest false-positive rate, a score of 0.028 which is drastically low when compared to the MFPD method and thus, make sense of the application of this research work proposed methods in real-time.

The hybridized evolutionary algorithm and DNN implemented by Ali and Ahmed [12] produced an accuracy of 88.77% which is very much significantly lower than the least achieved an accuracy of this research work. Concisely, all proposed methods for detecting phishing websites (i.e. ABET, RoFBET, BET, and LBET) of this research work are very much viable for application in real-time. The very low false-positive rate, as well as the high accuracy and f-measure scores, indicate the credibility and viability of these AI-based meta-learners using the Extra-tree base learner.

The recently published study of Zabihimayvan and Doran [5] that implemented the fuzzy rough set feature selection algorithm to enhance a homogeneous ensemble method reportedly achieved a maximum f-measure value of 95% (i.e. 0.95 which is lower than the various f-measures scores obtained respectively from all the proposed model of this study (i.e. proposed models achieved higher than 0.97 f-measure value).

Conclusively, the study of Zamir *et al.* [18] implemented some high-level AI meta-learner method for detecting

FIGURE 4. Visuals for FP, FN and F-measure (axis on the right) values.

phishing as reviewed in Section II. However, the most efficient and effective model of their study (i.e. Stacking1 (NN + RF + bagging) achieved and accuracy of 97.4 and f-measure of 0.97 which equal the performance of the BET method of this study as they both share similar bagging computation. However, Zamir *et al.* [18] most performing method was outclassed by other methods of this study (i.e. ABET, LBET, and RoFBET respectively).

V. CONCLUSION AND FUTURE WORK

The aim of this study is to provide an excellent solution to the menace of phishing in our modern society. By so doing, this study further aim to solve the existing shortcomings already in place. The shortcomings as previously identified in Section II as the inability of single classifier methods to detect phishing methods adequately, high false-positive rate, high false-negative rates, inadequacies of ensemble methods to perform excellently in detecting phishing websites as well as poor performances of some hybridized methods for detecting phishing websites when compared against single classifiers. Resolving these problems led to the pursuit of this research work. As a result, this research work proposed, implemented and presented four (4) different AI-based meta-learner models using Extra-tree algorithm base learner for detecting phishing websites. It lies in the heart of this study to produce credible and viable phishing website detection solutions that are of high predictive capability as well as with low false-positive and false-negative rates. The proposed methods (ABET, RoFET, BET and LBET) in this research work showcased the strength of AI meta-learners as an intelligent algorithm for developing models usable in detecting phishing websites. The methods produced extremely high predictive accuracy of approximately 98% by three of the proposed methods and also, a low false-positive rate of 0.018 by the ABET method and low false-negative rate of 0.033 from the LBET method. Evidently, the results indicate the effectiveness and efficiency of the proposed methods whose false alarm rate is drastically low while achieving high accuracy and f-measure scores. Comparative analyses established the excellent performances of the implemented methods proposed by this study. The methods presented by this study resolved all problems highlighted in the introduction section and sets a new performance standard for phishing website detection methods. In addition, this study presented AI phishing detection methods that are interpretable, unlike other black-box AI methods. The development of interpretable AI methods remains as a predominant concern in the AI community and the continuous contribution towards implementing interpretable AI models is essential.

In the future, we aim to consider other decision tree algorithms aside from the Extra-tree in other to produce an interpretable model. Further study that considers other families of AI algorithms whose models can be interpreted will be a considered research study of the future.

More so, in the context of developing a hybridized model, it was seen through the review of related works that some hybridized models for detecting phishing website performed poorly when compared against single classifier models. Thus, exploring feature selection or extraction algorithms with the implemented meta-learner methods of this study as a form of hybridization will be conducted in the future. Conclusively, the application of the implemented methods of this study in a real-time environment remains pivotal future work.

REFERENCES

[1] A. V. Elijah, A. Abdullah, N. Jhanjhi, M. Supramaniam, and B. Abdullateef, "Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: An empirical study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 520–528, 2019.

[2] A. Subasi, E. Molah, F. Almkallawi, and T. J. Chaudhery, "Intelligent phishing Website detection using random forest classifier," in *Proc. Int. Conf. Electr. Comput. Technol. Appl. (ICECTA)*, Nov. 2017, pp. 1–5.

[3] M. Alqahtani, "Phishing Websites classification using association classification (PWCAC)," in *Proc. Int. Conf. Comput. Inf. Sci.*, Apr. 2019, pp. 1–6.

[4] P. Yang, G. Zhao, and P. Zeng, "Phishing Website detection based on multidimensional features driven by deep learning," *IEEE Access*, vol. 7, pp. 15196–15209, 2019.

[5] M. Zabihimayvan and D. Doran, "Fuzzy rough set feature selection to enhance phishing attack detection," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2019, pp. 1–6.

[6] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2091–2121, 4th Quart., 2013.

[7] K. Gajera, M. Jangid, P. Mehta, and J. Mittal, "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)*, Jun. 2019, pp. 196–200.

[8] Y. Li, Z. Yang, X. Chen, H. Yuan, and W. Liu, "A stacking model using URL and HTML features for phishing Webpage detection," *Future Gener. Comput. Syst.*, vol. 94, pp. 27–39, May 2019.

[9] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing Websites based on self-structuring neural network," *Neural Comput. Appl.*, vol. 25, no. 2, pp. 443–458, Aug. 2014.

[10] M. K. Verma, S. Yadav, B. K. Goyal, B. R. Prasad, and S. Agarawal, "Phishing Website detection using neural network and deep belief network," in *Recent Findings in Intelligent Computing Techniques*. Singapore: Springer, 2019, pp. 293–300.

[11] M. Dedakia and K. Mistry, "Phishing detection using content based associative classification data mining," *J. Eng. Comput. Appl. Sci.*, vol. 4, no. 7, pp. 209–214, 2015.

[12] W. Ali and A. A. Ahmed, "Hybrid intelligent phishing Website prediction using deep neural networks with genetic algorithm-based feature selection and weighting," *IET Inf. Secur.*, vol. 13, no. 6, pp. 659–699, 2019.

[13] B. Wei, R. A. Hamad, L. Yang, X. He, H. Wang, B. Gao, and W. L. Woo, "A deep-learning-driven light-weight phishing detection sensor," *Sensors*, vol. 19, no. 19, p. 4258, Sep. 2019.

[14] R. P. Ferreira, A. Martiniano, D. Napolitano, M. Romero, D. D. De Oliveira Gatto, E. B. P. Farias, and R. J. Sassi, "Artificial neural network for Websites classification with phishing characteristics," *Social Netw.*, vol. 7, no. 2, pp. 97–109, 2018.

[15] R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing Websites using an automated technique," in *Proc. IEEE Int. Conf. Internet Technol. Secured Trans.*, Dec. 2012, pp. 492–497.

[16] E. Zhu, D. Liu, C. Ye, F. Liu, X. Li, and H. Sun, "Effective phishing Website detection based on improved BP neural network and dual feature evaluation," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl. Ubiquitous Comput. Commun. Big Data Cloud Comput. Soc. Comput. Netw., Sustain. Comput. Commun.*, Dec. 2018, pp. 759–765.

[17] G. Vrban i , I. Fister, and V. Podgorelec, "Swarm intelligence approaches for parameter setting of deep learning neural network: Case study on phishing Websites classification," in *Proc. 8th Int. Conf. Web Intell., Mining Semantics*, 2018, pp. 1–8.

[18] A. Zamir, H. U. Khan, T. Iqbal, N. Yousaf, F. Aslam, A. Anjum, and M. Hamdani, "Phishing Web site detection using diverse machine learning algorithms," *Electron. Library*, vol. 38, no. 1, pp. 65–80, Jan. 2020.

[19] A. O. Balogun, R. O. Oladele, H. A. Mojeed, B. Amin-Balogun, V. E. Adeyemo, and T. O. Aro, "Performance analysis of selected clustering techniques for software defects prediction," *Afr. J. Comput. ICT*, vol. 12, no. 2, pp. 30–42, 2019.
[20] B. Sun, S. Chen, J. Wang, and H. Chen, "A robust multi-class AdaBoost algorithm for mislabeled noisy data," *Knowl.-Based Syst.*, vol. 102, pp. 87–102, Jun. 2016.
[21] G. Haixiang, L. Yijing, L. Yanan, L. Xiao, and L. Jinling, "BPSO-AdaBoost-KNN ensemble learning algorithm for multi-class imbalanced data classification," *Eng. Appl. Artif. Intell.*, vol. 49, pp. 176–193, Oct. 2016.
[22] M. H. Kamarudin, C. Maple, T. Watson, and N. S. Safa, "A LogitBoost-based algorithm for detecting known and unknown Web attacks," *IEEE Access*, vol. 5, pp. 26190–26200, 2017.
[23] H. A. H. Al-Najjar, B. Kalantar, B. Pradhan, and V. Saeidi, "Conditioning factor determination for mapping and prediction of landslide susceptibility using machine learning algorithms," *Earth Resour. Environ. Remote Sens./GIS Appl. X*, vol. 11156, Oct. 2019, Art. no. 111560k.
[24] G. Collell, D. Prelec, and K. R. Patil, "A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data," *Neurocomputing*, vol. 275, pp. 330–340, Jan. 2018.
[25] S.-J. Lee, Z. Xu, T. Li, and Y. Yang, "A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making," *J. Biomed. Informat.*, vol. 78, pp. 144–155, Feb. 2018.
[26] H. Lu, L. Yang, K. Yan, Y. Xue, and Z. Gao, "A cost-sensitive rotation forest algorithm for gene expression data classification," *Neurocomputing*, vol. 228, pp. 270–276, Sep. 2017.
[27] L. Wang, Z.-H. You, S.-X. Xia, X. Chen, X. Yan, Y. Zhou, and F. Liu, "An improved efficient rotation forest algorithm to predict the interactions among proteins," *Soft Comput.*, vol. 22, no. 10, pp. 3373–3381, May 2018.
[28] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.
[29] S. K. Sahay and A. Sharma, "Extra-tree classifier with metaheuristics approach for email classification," in *Proc. Adv. Comput. Commun. Comput. Sci.* Singapore: Springer, 2019, pp. 437–446.
[30] S. A. Manaf, N. Mustapha, M. N. Sulaiman, N. A. Husin, H. Z. M. Shafri, and M. N. Razali, "Hybridization of SLIC and extra tree for object based image analysis in extracting shoreline from medium resolution satellite images," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 1, pp. 62–72, Feb. 2018.
[31] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA, USA: Morgan Kaufmann, 2011.

YAZAN AHMAD ALSARIERA received the bachelor's degree in computer science from Mutah University, Jordan, in 2010, the M.Sc. degree in computer science (minor in software engineering) from the University of Putra Malaysia (UPM), Malaysia, in 2013, and the Ph.D. degree in software engineering from the University of Malaysia Pahang (UMP), Malaysia, in 2018. He is currently a Faculty Member at the Department of Computer Sciences, Northern Border University. His main research interests include software engineering, software testing, artificial intelligence, computational intelligence, combinatorial optimization, cybersecurity, and secure software development.

VICTOR ELIJAH ADEYEMO received the B.Sc. degree in computer science from the University of Ilorin, Ilorin, Nigeria, in 2015. He is currently pursuing the Ph.D. degree with Leeds Beckett University, Headingly Campus, Leeds, U.K. His research interests include artificial intelligence, cyber security, sports analytics, software engineering, business data analysis, and big data mining.

ABDULLATEEF OLUWAGBEMIGA BALOGUN (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from the University of Ilorin, Ilorin, Nigeria, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS. He is currently an Academic Staff with the University of Ilorin. His research interests include machine learning, data mining, information security, and empirical software engineering.

AMMAR KAREEM ALAZZAWI received the bachelor's degree in computers' techniques engineering from the Islamic University College, Iraq, in 2013, and the master's degree in software engineering from University Malaysia Pahang, Malaysia, in 2016. He is currently pursuing the Ph.D. degree with Universiti Teknologi PETRONAS. His research interests include software testing, combinatorial testing, t-way testing, and artificial intelligence.