Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Improving Detection And Reduction Of Phishing And Social Engineering Attacks Through Ensemble Methods And Feature Engineering

Dr. A. Singaravelan*

Assistant Professor, Department of Computer Science,
Urumu Dhanalakshmi College, Trichy singaravelan.appar@gmail.co

Phishing and social engineering attacks are two of the biggest threats to cybersecurity since they rely on exploiting people's trust. Thus, to address such complex threats, the current research work suggests the use of an enhanced method that consists of an ensemble approach, feature selection, and optimization. For data gathering, there is a phishing email dataset available on Kaggle which is used in this work; the data preprocessing stage entails tokenization, stemming, lemmatization, and space removal of the stop words. In this work, we use the VGG19 network, originally applied to image classification tasks, for extracting features from textual data This is done in an effort to preserve higher level and more detailed characteristics of text which would then bear evidence of problematic phishing attempts. The first phase contributes a hybrid model between Convolutional Neural Networks (CNN) and Bidirectional Gated Recurrent Units (Bi-GRUs). This make it possible for the network to learn both the spatial and sequential characteristics of the data making the detection to be accurate. Moreover, the African Buffalo-Ant Colony Optimization (AB-ACO) algorithm is used for the fine-tuning of model parameters, and selection of the most important features thereby keeping up the performance. According to the given evaluation criteria based on the precision of 1000 emails, the proposed system has achieved 98% accuracy rate on differentiating between phishing and legitimate emails. The entire methodology is coded in Python and comes with optimized and light solutions for real-life problems. This approach presents a drastic improvement in phishing detection as well as prevention; thereby forming a solid foundation for improving security against new and emerging social engineering threat.

Keywords: Phishing detection, social engineering, ensemble approach, VGG19, AB-ACO optimization

1. Introduction

Phishing and social engineering attacks are modern complex threats based on the manipulation of human behavior and their confidence. Such kind of attacks are becoming more and more frequent, which means that the need for the improvement of the protection measures against such kind of threats is becoming acute [1]. Conventional techniques that are used to discover phishing and social engineering threats prove to be ineffective most of the time because of the constant change in the approach and vast number of suspicious links, messages and documents [2]. However, to overcome these challenges there is a rising interest in using machine learning features as well as the feature-level ensemble based with the best classification strategies [3].

In the recent past, different datasets have been used including those from Kaggle like the phishing attack email dataset to diagnose and understand the features of the phishing attacks. These datasets provide the core labeled data and resources that may be utilized to build and test machine learning algorithms that identify phishing and social engineering efforts [4]. Preprocessing is arguably one of the most important steps in the data preparation process since it is responsible for preparing raw data in a format that can be used in model development. Tokenization, stemming, lemmatization and elimination of stop words are some of the commonly used approaches for text pre-processing that aids in cleaning textual data and putting pre-text for easy analysis [5]. Feature extraction is considered as an important factor in enhancing the performance of the model, by extracting the most important features from the text data [6]. One is where features from the text data are extracted using VGG19 which is a deep convolutional neural network that defined for image classification. Due to the nature of text being transformed into image-like features through the use of, for instance, word frequency clouds or character images, the VGG19 model can easily detect hidden patterns and semantically hierarchical arrangement that might be useful in the determination of a phishing or social engineering attempt [7]. This method makes it possible to incorporate higher order features into the classification process since these features are rich and possess high dimensionality [8].

For classification, the proposed approach of employing CNN in combination with Bi-GRU is again effective because the data contain both spatial and sequential nature [9]. CNNs excel in learning the local patterns inside data which is especially useful for the text data, where Bi-GRU will be helpful in analyzing the sequence and context in order to find the flow of the text. The combination of CNN and Bi-GRU models with different learned features can help to reach higher accuracy in the detection of phishing and social engineering attacks [10]. The classification models can then be further optimized using the newly developed AB-ACO algorithm; the African Buffalo- Ant Colony Optimization algorithm [11]. AB-ACO is a new solution strategy based on the exploration of ant colony optimization and on the optimization prospects that can be drawn from the actual movement of African buffalo herds. This quasi-optimization method also holds the prospects of enhancing the model's performance, especially with respect to hyperparameter tuning as well as feature selection oriented towards attack prediction [12]. Thus, utilizing these avance methodologies, an organization can establish a system proactive enough to counter all the threats identified with phishing and social engineering attacks [13].

Phishing and social engineering attacks are arguably some of the most complex and dangerous threats within the modern world's cybersecurity space [14]. By using social engineering tactics, which prey on human nature and the people's ability to trust others, these attacks are hard to identify and prevent when compared to other types of interventions. Unlike technical weaknesses, where phishing and social engineering involve the manipulation of people's authenticity to provide secret information or make decisions that destabilize security [15]. As the attackers keep targeting the users by inventing new methods of attacks, fully depending on the typical techniques like signature-based detection or basic heuristic methods is very inefficient. This underlines the necessity of developing and implementing higher intelligent solutions with a possibility to adapt to these threats [16]. The latest development in

machine learning presents some potentials possibilities into refining the detecting of phishing and social engineering. Currently, there exist several machine learning algorithms, which can demonstrate good results in analyzing large amounts of information and identify intricate patterns of behavior characteristic of these attacks, and therefore, the use of machine learning for constructing models for analysis of such attacks is justified. These algorithms are therefore capable of learning from past data and patterns associated with various activities similar to previously identified phishing attempts hence making it easier to identify new previously unseen threats. The use of feature-level ensembles enhances the performance since it integrates multiple models or features and makes prediction with high accuracy.

Platforms like Kaggle contain datasets, which can be used to train and test the created machine learning algorithms to detect phishing and social engineering attacks. They mainly comprise of phishing samples and legitimate samples that enables one to learn the general features and trends linked with phishing attempts. In so doing, such datasets can help in the enhancement of models that spell out genuine texts from those of fraud-related ones, thus improving overall detection rates. It can also be noted that text pre-processing is the backbone of preparing data for feeding it to the machine learning models. Preprocessing in an essential step it checks the raw data and prepares it to perform better on the next models that are to be implemented. Actually, there are several preprocessing techniques such as tokenization, stemming, lemmatization, and stop word removal which play important role in pre-processing of text data. Not only does it eliminate redundancy but also minimizing noise that surrounds the data simplifies it for models to focus on those features that are most likely signs of phishing and social engineering.

Feature extraction enhances the machine learning models by being able to select and extract data components that improve its performances. For instance, applying VGG19, a convolutional neural network primarily developed for imaging data to work on text data shows how general-purpose deep learning methodologies are. It enables one to extract features that are high-dimensional thus facilitates identification of complex features in the text such as hierarchical relations. When incorporated with the classification techniques like CNN and Bi-GRU focused on spatial and sequential nature respectively this makes the overall detection a strong one with better accuracy levels for complex phishing and social engineering threats. The Key Contributions of this article is listed below,

1.The presented approach of text classification does not make use of any specialized text processing methods, but relies on standard text preprocessing techniques such as tokenization, stemming, lemmatization and stop word removal for data pre-processing.

2. It adjusts the VGG19 model, which was developed for image classification, for text data proving its versatility.

3. The approach presented here investigates the application of a deep learning model that integrates the Convolutional Neural Networks (CNN) along with Bi-directional

Gated Recurrent Units (Bi-GRU); this approach boosts the effectiveness of extracting spatial-temporal features from the text to achieve a better phishing detection.

4. The proposed African Buffalo-Ant Colony Optimization (AB-ACO) algorithm is applied for feature selection and the identification of the optimal parameters of the model with enhanced phishing detection capabilities.

5. Having a balanced dataset obtained from Kaggle, it was approximately 18 percent of the study's participants. 6000 emails, helps in training as well as the evaluation of the models hence providing the appropriate identification of the actual phishing as well as genuine emails.

The paper is organized as follows: Section 2 contains important material intended to help readers understand the proposed work using existing approaches, whereas Section 3 goes further on the problem definition. The subsequent part shows CNN-BiGRU for Improving Detection and Reduction of Phishing and Social Engineering Attacks. The fifth part contains table and visualizations of the outcomes along with metrics. Finally, in Chapter 6, its conclusion and upcoming projects are addressed.

2. Related Works

Research in this area highlights concerns of extended rates of phishing within the current year as characterized by the enhanced uptake of remote working resulting from the COVID 19 pandemic. Phishing, where an attacker pretends to be a legitimate person or company in order to get the victim to reveal confidential information such as passwords through emails or fake websites continues to present a key threat to the users, companies, and service providers. Studies have shown that more attention has been devoted on the use of the ML models in improving the detection of phishing. Previous methods like rule-based system have failed us with the present complex and sophisticated phishing attacks. The authors of the papers described above argue that the use of ensemble of classifiers can increase the accuracy of detection in several folds. For example, models that combine Artificial Neural Networks, K Nearest Neighbours, Decision Trees and Random Forest Classifiers show high performance, a performance that is even better than that of the individual classifiers. Particularly, it has been shown that using both KNN and RFC has led to the detection accuracies of up to 97%. 33%. This is in light of the multifaceted nature of phishing attacks that may warrant extensive analysis by various models in an ensemble fashion to provide an avenue for improving on the available security measures with a view of reducing on the risks posed by these attacks [17].

Social engineers are recognized to be on the rise particularly with the emergence of pharming attacks during the COVID-19 pandemic outbreak whereby people are more vulnerable due to fear and these lead to attacks on health authorities and organizational executives. Pharming attacks in which the users are directed to fake websites instead of the actual legitimate ones can be considered as dangerous since the attackers can mimic the actual entities that the victims trust. But the advancements in recent years have been in creating

complex models to forecast and prevent these attacks as well. A number of researches have demonstrated the ability of the ensemble techniques of building improved models through combining several classifiers. For instance, the combination of approaches that have used Logistic Regression, Random Forest, and Gradient Boosting has helped in the fight against pharming threats. Analyzing samples such as 1781 URLs included in the Kaggle data set created by Manu Siddhartha, it can be stated that URLs of malicious sites often do not contain such features as the correct division of protocols and authorization of certificates. The usage of these models in data processing would hopefully enhance the ability to detect pharming attacks and minimise the effects of malware with the help of tools, such as CSV files and Jupyter Notebooks. This resonates with the effectiveness of ensemble methods and other sophisticated algorithms in improving on the anti-pharming and other security risks [18].

Phishing is still up and about as a cybersecurity risk that involves creating bogus websites branding them to look like those of legitimate bodies in order to obtain users' details such as usernames, passwords, and credit card numbers. Despite the use of different measures to fight the campaigns, users are still becoming victims of the tricks. Lately, new studies suggest the idea of the utilization of systems that would allow for the detection of intrusive ads with higher sophistication. It has also been found that by combining feature selection, different feature selection methods, and different machine learning algorithms increases detection accuracy. Some of the work that has been done in phishing detection has used algorithms like Random Forest, Decision Tree, XGBoost and were used concurrently to develop powerful algorithms for the models to identify and detect phishing sites. Research work done using datasets from UCI and Mendeley has shown that such ensemble methods can result in very high detection rates with rates as high as 97%. 51% and 98. 45%, respectively. Such outcomes demonstrate that the proposed models are better than threshold and frequency baseline models, which demonstrates the usefulness of the proposed approach in the modern context of addressing the new type of threat like phishing attacks and the provision of a much better solution than the existing ones [19].

Phishing attacks can be regarded as one of the future threats that take advantage of weaknesses in an organization's systems through the use of trickery in order to steal information and financially damage organizations and individuals. Recent literatures recommend the need to establish effective IDS systems that helps in eliminating such risks. The use of rule-based filters and single machine learning classifier have their demerit especially in changing and complex phishing schemes. New developments in ensemble of classifiers approach, where several classifiers are integrated aiming at improving the performances and stability of the system, introduced optimistic results in the improvement of the effectiveness of phishing detection. Previous research has incorporated new ways in enhancing the classification performance such as deep learning, NLP, and feature engineering. For instance, the use of the ensemble of CNNs and RNNs has been benefitting the patterns inherent in the context of the web and email interactions. Furthermore, hybrid models that incorporates optimization algorithms of ACO and the GAs has enhanced feature selection and parameters tuning achieving high detection rates. However, this remains a challenge even today in devising-capable systems that can loosely generalize on new and complex form of

phishing portrayals and strategies, further reiterating the fact that the battle needs the innovation of better detection mechanisms [20].

Spam texing is another issue that has contributed to social engineering attacks especially through the frequent use of mobile messaging services resulting to phishing where fraudsters seek to obtain credit card numbers and passwords. Further, the increase in the spread of fake news and rumors especially with regards to COVID-19, creates increased need for efficient spam content filtering to minimize confusion and fear. Earlier, the technique used in spam detection was mostly based on machine learning and deep learning; however, these techniques have demonstrated several constraints. Refurbished Machine Learning Models previously require time-consuming feature engineering prior to model creation, and deep-learning models require more computation in building the model. Recent work has thus looked at an approach based on dynamic deep ensemble models to tackle those problems, which learn their architecture and perform feature learning. By nature, such models use the convolutional and pooling layers for features extraction and ensemble classifiers such as random forest and extremely randomized trees for text classification. In addition, some approaches such as boosting and bagging which fall under the category of ensemble learning have also been applied in order to improve the performance of the models. It has been found in the literature that such complicated models are capable of obtaining high performance indicators, including precision, recall, F1-score, and accuracy equal to 98%. 38% which prove the efficiency of using various measures for enhancing the usability of spam elimination and filtration in the context of contemporary media systems [21].

Over the last few years, it has become clear that even organized and targeted attacks using social engineering tricks, especially phishing, combined with the observation that traditional approaches to detecting and preventing such heist misses the mark, there are holes in current enterprise systems. Traditional techniques for detection of phishing have become lesser effective due to the manual feature extraction and the inability of handling complex phishing attacks. Recent inventions in deep learning techniques have proved potential in dealing with the above stated limitations. For instance, the use of LSTM network and XGBoost tree has been applied to augment the individual security of emails by properly disinfecting them from cheeks mimicry such as phishing emails. Furthermore, particulars of Bidirectional LSTM with Attention mechanisms have been used to solve group security to detect internal threats. As was shown in the modern studies, the utilization of double-layer detection frameworks that comprise such techniques has been preceded by a series of significant achievements: Bi-LSTM introduces dependencies of long range in the content and sequences of mailbox, users, and is reinforced by XGBoost as well as attention mechanisms in the classification of such content. This dual way enhances the results while at the same time it eliminates the time and effort spent on feature extraction [22].

Phishing is still one of the major threats of cybersecurity, where the attackers imitate the genuine websites or URLs to mislead users and get their personal details. Contemporary progress in this field has focused on paying detailed attention in identifying phishing in respect to different variables which includes the structure of the URL, the contents of the page, and the queries of the external resources. Preceding approaches have employed feature extraction

on Web samples for detection algorithms' development. In this regard, a number of studies have shown that ensemble learning methods, which uses several classifiers in one solution enhances performance of a single classifier approach. These models have been enriched with optimization strategies such as grid search for hyperparameters as well. For example, the adoption of a sample of 11,430 with 87 features help in building benchmark systems of high accuracy in detection. The analyses of the models that have been developed in this study based on optimization of ensembles and tested based on target metrics for each type of algorithms such as accuracy, precision, recall, F1 score, ROC curve area results have demonstrated a high level of performance. For instance, the present studies present an ensemble model at 95 percent accuracy. 36% accuracy, 96. 29% precision, 94. 24% recall, 95. 26% F1-score, and an ROC curve area of 0. Further to this work, research on the 9876 model is evidence that establishes its robustness in separating the real URLs from the phishing ones while analyzing enhancements in detection mechanisms with an advanced machine learning approach [23].

New study supports the growth of the phishing and similar kinds of social engineering attacks, due to the rising popularity of remote work during the COVID-19 crisis. Most of the earlier approaches for detection of phishing were based on rules and normative models or single machine classifiers which have been unable to meet the challenge posed by this evolving menace. There are pieces of research that observe that the incorporation of several classifier like Artificial Neural Networks (ANNs), K-Nearest Neighbors (KNN), Decision Trees and Random Forest alongside others improves the detection ratio to a great extent. The advanced methods also included dynamic deep ensemble models that employ features such as convolution layers and ensemble Classifiers for automatic feature realisation and enhanced precision. Recent development includes Fisher–Markov based and Fisher–Markov–Markov based phishing detection technique that reduce false positive rate by employing under sampling method to address class imbalance problem. Also, the use of deep learning techniques such as LSTM, Bidirectional LSTM, and Attention mechanism into the double-layer detection systems has emerged as effective in the identification of phishing and insider threats. Moreover, when it comes to ensemble models, optimized in regard to the tested data sets have shown fairly good accuracy and resilience, underlining the necessity to recapture radical measures towards combating existing and emerging threats, including phishing and other cyber risks.

3. Problem Statement

The topic fast-growing incidents of phishing activities, especially due to the COVID-19 pandemic, has made working from home a liability to cybersecurity. Phishing scams, which involve tricking people into divulging personal details under the pretext of coming from reputable organizations either through e-mails or imitation Web sites, are still a headache to users, firms and service providers. Conventional ways of detecting phishing include the use of rules which has greatly been challenged due to better strategies formulated by attackers. Consequently, the development of more effective detection methods by utilizing the state-of-the-art ML techniques has become rather urgent. However, researchers have noticed that existing solutions are not optimal even with the current progress in the field of ML and this is

because most solutions rely on the use of single classifiers that may be inadequate in capturing the real nature of phishing [17].

4. Proposed CNN-BiGRU Methodology

The process of detecting phishing emails is done by gathering data from Kaggle Phishing Email dataset. Before analysis, the gathered data contains tokenization, stemming, lemmatization and the deletion of stop words in order to clean its text. In the aspect of feature extraction, VGG19 is employed although it was initially built for image classification; nevertheless, its capabilities on extracting textual features characteristic of a phishing attempt are employed. In order to improve feature selection and machine learning model, the African Buffalo Ant Colony Optimization is used to boost feature selection and parameter optimization. Afterwards, a hybrid classification mechanism is applied by combining Convolutional Neural Network with Bidirectional Gated Recurrent Units. This combination utilizes deep learning architectures, CNN that is adept at capturing spatial features and Bi-GRU that is efficient at capturing temporal features thereby enhancing the detection of phishing emails. Figure 1 shows Proposed CNN-BiGRU Methodology.

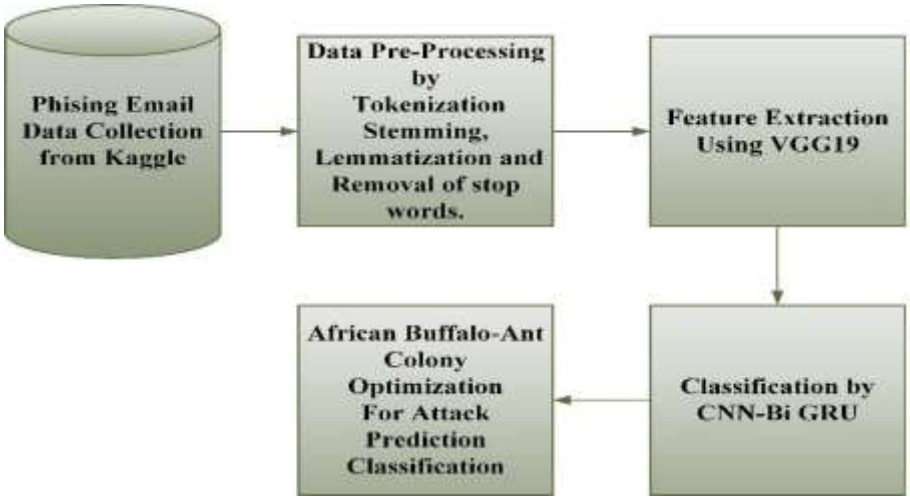

Figure 1: Proposed CNN-BiGRU Methodology

4.1 Data Collection

The dataset for phishing email detection is obtained from Kaggle and plays a major role in the successful construction of machine learning models for detecting phishing emails [24]. This dataset, labelled `Phishing Email. csv`, contains approximately 18. 6k email samples with two main features: “Email Text” is the body part available containing the accounts on emails and “Email Type” that distinguishes between phishing mails and safe mails. The dataset is segregated under the different length range of emails so that many examples could be selected. This work is available under GNU Lesser General Public License 3. This data provides substantial support for large scale text analysis and modelling work, bringing the figure to 0. It is used for tokenization, stemming, lemmatization and filter stop words which are used for

arranging the text content for feature extraction and classification. It is equally possible to generate a balanced ratio of phishing and safe emails with approximately 39% of phishing and 61% which makes it ideal for detecting and training algorithms in improving the standards of phishing email identification systems. Table 1 shows Phishing Email Dataset Overview.

Table 1: Phishing Email Dataset Overview

Dataset Details	**Description**
Source	Kaggle
Label	Phishing Email.csv
Number of Samples	Approximately 18,600
Main Features	- "Email Text": Contains the body of the emails
	- "Email Type": Indicates whether the email is phishing or safe
Email Length Segregation	Emails are categorized by length ranges to facilitate selection of various examples
License	GNU Lesser General Public License 3 (LGPL 3)
Preprocessing Techniques	Tokenization, stemming, lemmatization, and stop word removal
Data Balance	Approximately 39% phishing emails and 61% safe emails, ensuring a balanced dataset for training and testing
Usage	Ideal for large-scale text analysis, feature extraction, and model training for phishing detection

4.2 Data Pre-Processing

The process of building models for classification of phishing emails requires some important features to be extracted from the raw emails and several important pre-processing techniques are to be applied. The first step is known as tokenization, which puts the message's content into a form where it is easier to work with, through turning the text into a sequence of words or tokens. After tokenization, the stemming process is undertaking to bring all the inflected forms of a word to the root word format, for example, are – running is converted to run. Lemmatization is then used to bring the words to their base or dictionary form so that inflected or spaced variants of the same word are considered as equal to the other while using the pa(min) (e. g., "better" to "good"). Also, extra white space, punctuations, special characters and stop words like, the, and, is and so on are also eliminated to cut-off all forms of noise in textual analysis. In its totality, these preprocessing measures improves the quality of the text data in a way that it can be fit for feature extracting and follow-up machine learning classification activities. Table 2 shows Text Pre-Processing Techniques.

Table 2 : Text Pre-Processing Techniques

Pre-Processing Technique	Description
Tokenization	Converts text into a sequence of words or tokens, making it easier to work with.
Stemming	Reduces words to their root form by removing suffixes.
Lemmatization	Converts words to their base or dictionary form, considering context and meaning.
Stop Word Removal	Eliminates common words that do not contribute significant meaning, such as "the", "and", "is".
Removal of Extra White Spaces	Cleans up unnecessary spaces between words and characters.
Punctuation Removal	Removes punctuation marks to reduce noise in the text.
Special Character Removal	Deletes special characters that may not contribute to text analysis.

4.3 Feature Extraction by VGG 19

Feature extraction using VGG19 entails utilizing the deep convolutional neural network that was configured to work in the image classification scenarios, reconfigure it to work on textual data. There are 19 layers in the VGG19 including the convolutional layer, max-pooling layer lastly fully connected layers which make the architecture so simple yet efficient. When applying VGG19 for text the original content of the email is pre-processed into a format comprehensible by the network, for example, word embeddings or vectors. The layers of convolution operation are then used to extract features of multimodal text representations at different levels of semantics and syntax embedded in the body of the email. That is, it employs the learned patterns of image recognition of the VGG19 model and applies this knowledge to the recognition of text entries in the data set. Figure 2 shows VGG 19 Architecture.

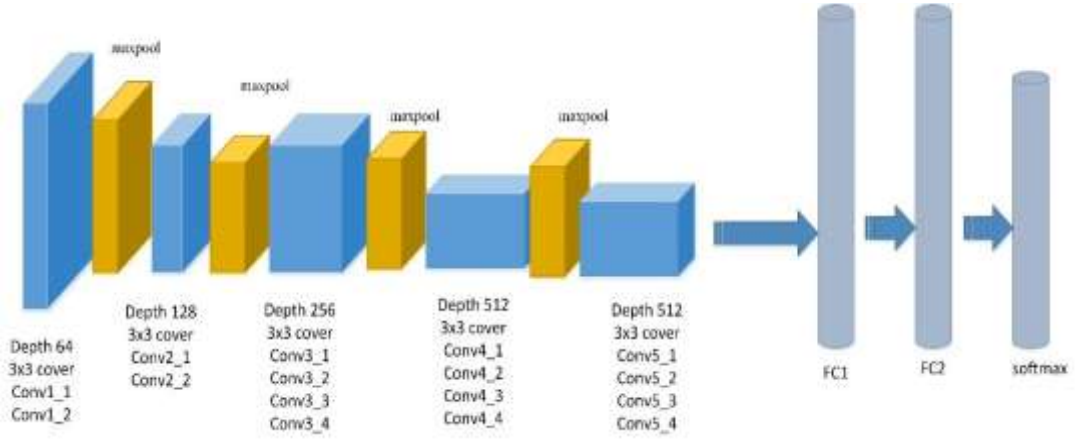

Figure 2: VGG 19 Architecture

VGG19 certainly provides feature extraction where detailed and high-level features of the email text is valuable when it comes to phishing attacks. The network's architecture is quite deep, and that means the model can understand that the text contains features that may indicate that it is phishing. VGG19 is mainly used to extract features from an email and the output of VGG19 would commonly a set of feature vectors in terms of the email content that can be used for further classification models. This method also helps to incorporate state of the art, image based deep learning into the text-based problem of phishing into the detection of real and fake emails. Table 3 shows VGG19 Model Adaptation for Text Analysis.

Table 3: VGG19 Model Adaptation for Text Analysis

Aspect	**Details**
Model Used	VGG19
Original Purpose	Image Classification
Number of Layers	19 (including convolutional layers, max-pooling layers, and fully connected layers)
Adaptation for Text	Reconfigured for textual data by converting text to word embeddings or vectors
Preprocessing Required	Tokenization, embedding or vector conversion, and other text normalization techniques
Feature Extraction	Extracts high-level features from text, identifying patterns at different levels of semantics and syntax
Mechanism	Applies convolutional operations to text embeddings to capture multimodal features
Output	Feature vectors representing various aspects of email content
Application	Used for classification in phishing detection systems
Advantage	Incorporates deep learning techniques from image recognition into text analysis for enhanced phishing detection

4.4 Classification by CNN-BiGRU

CNNs perform several actions on the incoming data. The following are the basic levels of a CNN structure

Layer of Input: This layer holds the basic information from the input, like an image. In this level, each neuron corresponds to a pixel in the input image.

Convolutional layer: The input image is enhanced by a number of kernels added by the convolutional neural network's layer. They are used to identify features in the input and edges, which are textures in particular. In order to create maps of features, convolutional neural network calculations are performed by moving screens across the input image and then calculating dot product.

Activation Layer (ReLU): To introduce unpredictability into the framework under consideration, an activation process that is not linear, unlike ReLU, is executed unit by unit for each layered operation. ReLU's effectiveness and user-friendliness in building neural networks with deep connections to learn have made it popular.

Pooling Layer: The convolutional planes' map of features is condensed by the layer that pools them. By limiting the geographical coverage of maps with features, it lowers computation expenses and avoids excessive fitting. Minimum pooling as well as normal pooling are both of the most prevalent pooled strategies.

Given the input image X and a filter F,the convolution operation is defined as

$$(X * F)\,(a, b) \;=\; \sum_{j=0}^{m-1} \cdots \sum_{k=0}^{n-1} \cdots I(a + j, b + k).\, F(j, k) \qquad (1)$$

where (a, b) are the spatial coordinates, m and n being the dimensions of the filter.

Fully Connected Layer (Dense Layer): Characteristics images are converted through vector or fed into layers that are fully linked after multiple rounds of convolution with pooling. All of the neurons in a layer that is completely linked have links to all of the neurons in the layer that came before it. These levels make forecasts based on excellent data.

Output Layer: The CNN's final element produces outputs. Depending on the job at hand, this layer's overall neuron count fluctuates. For example, in the output layer for a classification task with n classes, there will be n layers. Such neurons frequently get followed by an expression of SoftMax stimulation, which provides likelihoods of classes.

CNN and Bi-GRU integration, which uses them in parallel to classify phishing emails and using the combined models will greatly improve the accuracy of the model. The CNN component is mainly involved in the extraction of spatial features from the given email text data since this layer applies convolutional filters directly to the input data in order to capture local patterns and structures. This process helps the network narrow down some of the characteristics, for example, the words or the phrase commonly used in the phishing scams. In this manner, the CNN learns the spatial features which point out the vital aspects of the text that may contain the malicious content. Figure 3 shows CNN-BiGRU Architecture.

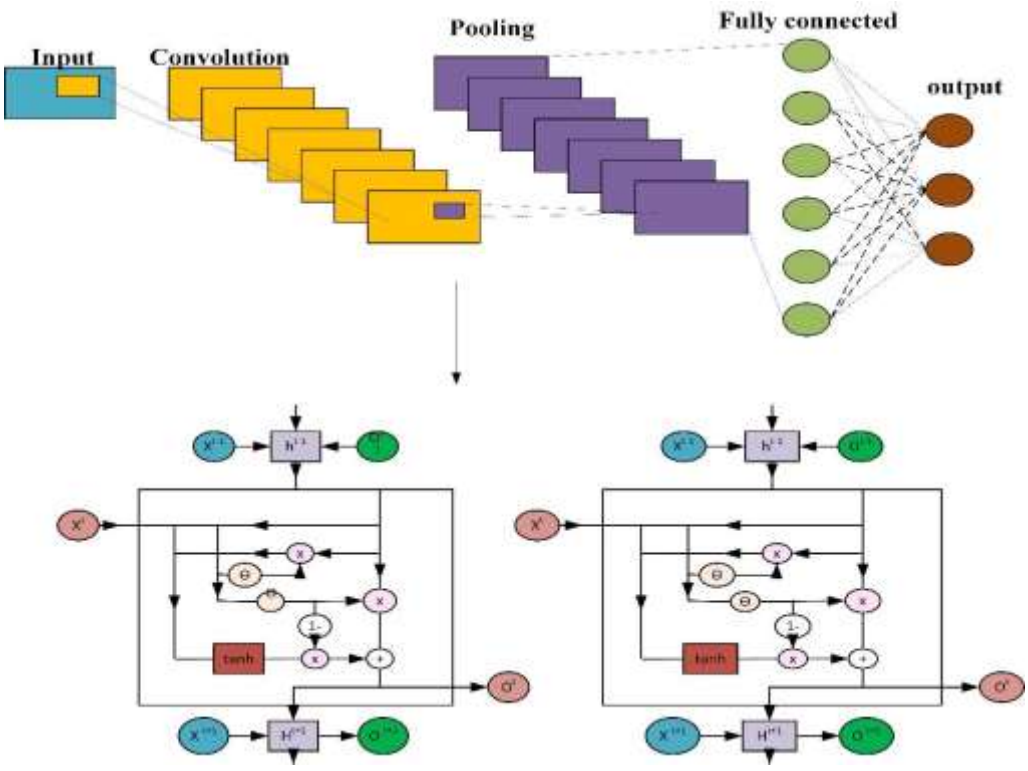

Figure 3 : CNN-BiGRU Architecture

After that, the Bi-GRU component of the model takes care of the features that were extracted through the CNN. Unlike standard Recurrent Neural Networks , Bi-GRU is able to process text sequence in both forward and backward way which makes it possible for the model to learn context from words before and after the current word. This bidirectional method is especially useful when it comes to getting a deeper understanding of the context of the e-mail content and as a result helps the model become less sensitive to specific and context-independent phishing cues. By combining CNN's feature extraction abilities with Bi-GRU's complete collection modelling, this hybrid category method enhances the detection of phishing emails by integrating each neighbourhood function extraction and international context know-how, main to greater robust and correct class effects.

4.5 African Buffalo-Ant Colony Optimization to Improve Phishing Detection

The African Buffalo–Ant Colony Optimization algorithm, termed as AB-ACO algorithm, is new innovative approach that integrates the herding behavior of African Buffalos with the behavior of ants in an Ant Colony Optimization to select the most pertinent features and optimize the algorithm parameters for effective identification of phishing emails. Based on the synergy of the herding behavior of buffalo and efficient searching and searching and feeding behavior of ants, the AB-ACO begins navigating through a search space for better features and

better model parameters. As for what is more related to features applied to the task of phishing detection, AB-ACO is used to determine relevance and contribution of each one. In this way, by using the hybrid nature of the given approach, it is possible to filter out the most significant features from the dataset or reduce the influence of the most significant disturbances and, thus, enhance the performance of the phishing detection model. Figure 4 shows Algorithm for African Buffalo-Ant Colony Optimization.

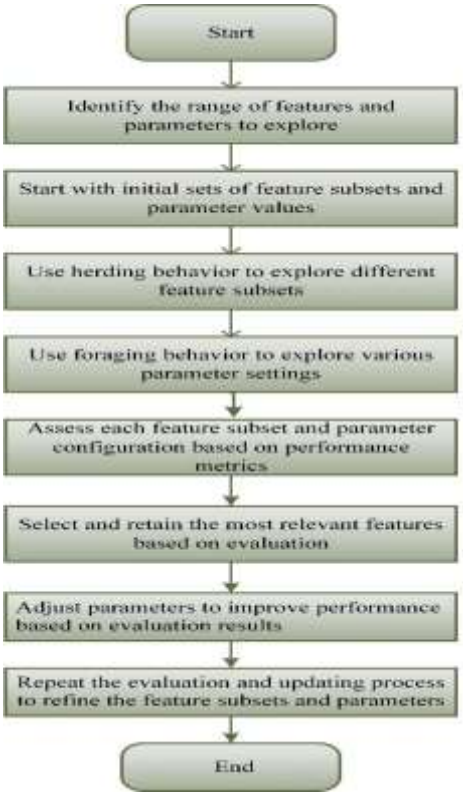

Figure 4 : Algorithm for African Buffalo-Ant Colony Optimization

Not only feature selection but also the parameters of the classification models are optimised by the AB-ACO methodology. Due to the optimization of hyperparameters, the use of the algorithm results in the enhancement of the classifiers' ability to provide the necessary differentiation between phishing and legitimate messages. The AB-ACO integration influences the classification pipeline in a way that allows the model to adjust the parameters according to dynamics of the values of the dataset as well as the intricate nature of phishing attacks. This leads to having a stronger phishing detection system that will be able to meet the challenge posed by the constantly changing face of phishing. Thus, feature optimization along with parameter tuning using AB-ACO has been instrumental in enhancing the degree of efficiency of the phishing detection systems.

5. Results and Discussion

The accuracy plot, which is also called the precision-recall plot or the receiver operating characteristic (ROC) curve, is a key measure of the efficiency of a classifier model, especially when it comes to spam emails, including phishing ones. This curve presents the relationship between the sensitivity or the ability to identify true, positive cases and specificity or the ability to identify true, negative cases based on the cut-off values. This curve shows graphs as the threshold increases or decreases: it shows how the model can identify the emails correctly with the least possible numbers of the false positives. An ideal model should ideally show a curve close to the top left corner of the plot which specifies high value of true positive rate and low values of false positive rate . Further, there is a statistical measure of the overall performance of the model, including the ability to distinguish between phishing and real emails known as the area under the curve (AUC), set higher. Through the evaluation of the accuracy curve, it becomes easier to determine the degree of variability and the extent to which the model can perform well to solve the fairly complicated problem of phishing identification. Figure 5 shows Train-Test Accuracy Curve.

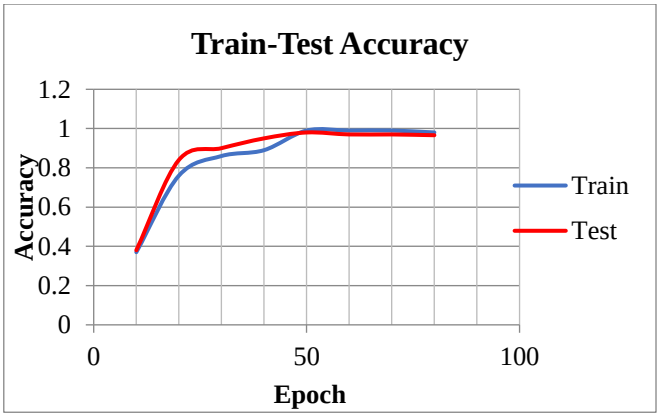

Figure 5: Train-Test Accuracy Curve

The loss curve is one of the necessary tools for evaluating the training process of an ML model for phishing email detection. Here the vertical axis represents the loss values or the error that has been committed and the horizontal axis represents the number of epochs, thereby showing how the model is getting better. During the training of the model, on the same graph, one will see the loss values which depict that the model is learning as the curve has a fall. This type of loss curve illustrates a fulfilling, declining loss, which indicates adequate learning and convergence in the used model. Conversely, if the curve flattens prematurely or well-known shows excessive fluctuations, it could indicate problems including overfitting, underfitting, or problems with the mastering charge. By analyzing the loss curve, you could benefit insights into the efficiency of the education method, assess whether or not the model is generalizing

properly to unseen information, and make essential modifications to enhance version overall performance. Figure 6 shows Train-Test Loss Curve.

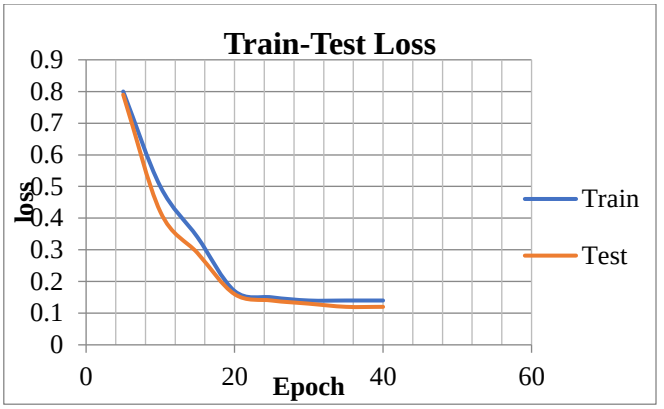

Figure 6: Train-Test Loss Curve

The ROC (Receiver Operating Characteristic) curve is a graphical representation used to evaluate the performance of a binary class version, along with the ones used for phishing email detection. It plots the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity) throughout diverse threshold settings. The ROC curve illustrates the alternate-off among the model's ability to effectively become aware of phishing emails and its tendency to incorrectly classify valid emails as phishing. A curve that extends towards the top-left nook of the plot shows higher overall performance, as it reflects a higher authentic wonderful fee and a decrease false superb charge. The region underneath the ROC curve (AUC) provides a unmarried scalar cost that summarizes the overall effectiveness of the model; an AUC close to 1.0 signifies first-rate discriminatory capacity, whilst an AUC around 0.Five shows no higher performance than random guessing. Evaluating the ROC curve facilitates in selecting the premier threshold and knowledge the version's functionality to differentiate between phishing and valid emails successfully. Figure 7 shows ROC Curve.

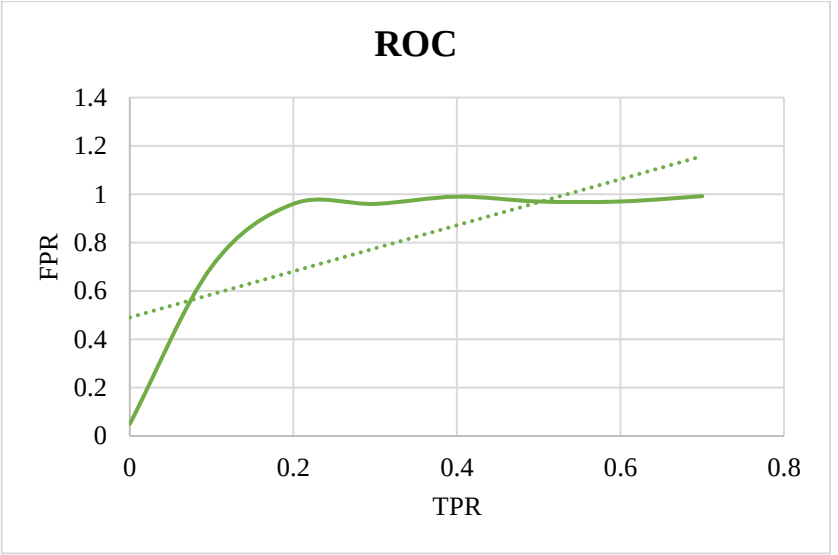

Figure 7: ROC Curve

The fitness graph of AB-AC Optimization usually shows the result of the optimization process as against the iteration or generation. Finally, when used in this context "fitness" means the ability of the solutions generated by the algorithm to meet set goals and objectives. In the usual representation of the graph, the fitness values are placed on the y-axis while the number of iterations or generations are marked along the x-axis. First, the fitness values may be quite oscillating since there are variations of many solutions that can be offered by the algorithm. As the number of iterations increases and thus gradually getting better solutions from the algorithm, the fitness scores usually follow a steady and ascending trend, showing the true picture of the optimization. The graph assists in analyzing the convergence properties as well as the effectiveness of the optimization strategy whereby the AB-AC Optimization finds solution with higher quality. Figure 8 shows Fitness Graph of AB-AC Optimization.

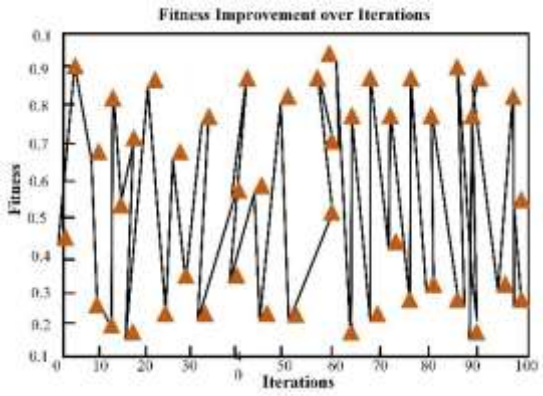

Figure 8: Fitness Graph of AB-AC Optimization

Detection And Reduction of Phishing and Social Engineering Attacks Through Ensemble Methods and Feature Engineering In the study titled " A Review of Machine learning techniques for Detecting new age Internet Security threats in Computer Networking", different machine learning models were used in the assessment of detecting the phishing and social engineering attacks. In the case of the Multi-Layer Perceptron (MLP) an accuracy of 96%. 74 %, where precision is 94 %. 87%, recall at 97. achieved an accuracy level of 52 % and an F value of 96. 82%. The K-Nearest Neighbors (KNN) model had a little less prominent metrics, which were up to accuracy 95. 65%, precision at 94. 45%, recall at 95. Meaning thereby that each of them has an overall accuracy of 45%, and F1-score of 94. 95%. On the other hand, the proposed CNN-BiGRU model presented better results with their accuracy of 98%, precision of 97%, recall of 98% and F1-score equal 98%. This concludes the usefulness of the CNN-BiGRU model in enhancing the ability of detecting and combating the attacks such as the phishing and social engineering. Table 4 shows Experimental Result Analysis for Different Parameters with other Metrics.

Table 4: Experimental Result Analysis for Different Parameters with other Metrics

Methods	Performance Metrics			
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
MLP [25]	96.74	94.87	97.52	96.82
KNN [25]	95.65	94.45	95.45	94.95
Proposed CNN-BiGRU	98	97	98	98

The bar chart shows the performance metrics of different phishing detection methods, which reveals the better performance of the proposed CNN-BiGRU model compared to MLP and KNN CNN-BiGRU model achieves higher accuracy at 98%, where with at 97% and 98%, respectively the precision, recall, and F1 score of 98% In contrast, the MLP model shows a precision of 96.74%, a precision of 94.87%, a recall of 97.52%, and an F1 score of 96.82%; The KNN model performs slightly lower with 95.65% accuracy, 94.45% precision, 95.45% recall, and 94.95% F1 score. This design clearly shows that the CNN-BiGRU model is effective to improve the detection of phishing and social engineering attacks by using advanced methods of clustering and feature engineering. Figure 9 shows Performance Evaluation.

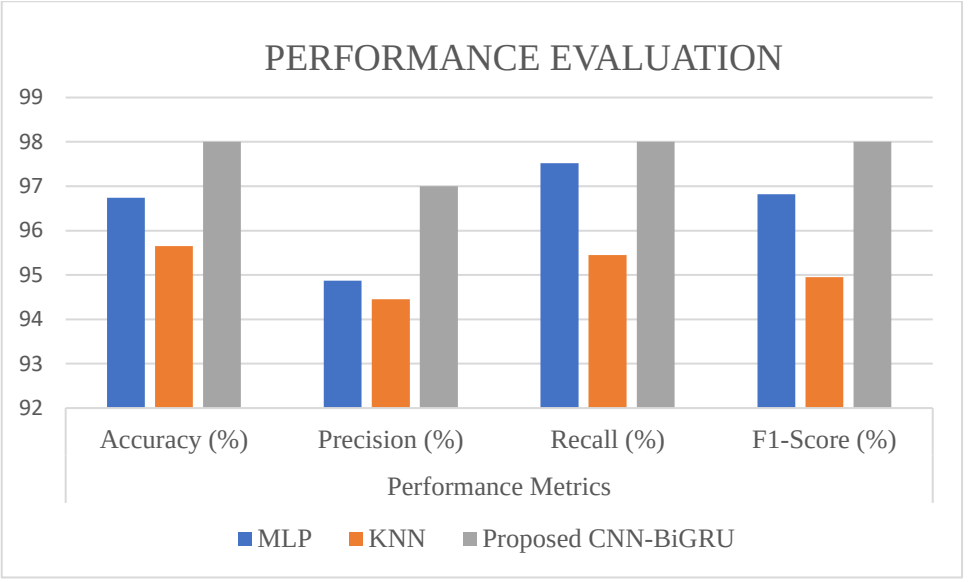

Figure 9: Performance Evaluation

5.1 Discussion

The results including MLP [25] achieved an accuracy of 96 percent which is quite satisfactory for computer vision. 74% and the ±95% confidence level were examined and found to have a precision of 94%. 87%, recall of 97. H-REC obtains a precision score of 52 %, and an F1-score of 96. 82%. Nevertheless, there was slightly lower accuracy in terms of KNN model they obtained 95%. 65%, precision of 94. 45%, recall of 95. Accuracy of 45%, Precision of 0.942, Recall of 0.938 and the F1-score of 0.94. 95%. Nonetheless, the proposed CNN-BiGRU model perform far better than the others with accuracy, precision, recall and F1-score of 98%. That is why the presented outcomes have shown positive effects of using the CNN-BiGRU model in enhancing the detection and avoidance of phishes and social engineering attacks, proving that this is one of the most effective approaches for implementing the latest ensemble techniques and sophisticated features to increase security systems' effectiveness.

6. Conclusion and Future Works

Hence, the research establishes that the improved machine learning models such as CNN-BiGRU lead to improved detection and prevention of phishing and social engineering attacks than the conventional models like MLP and KNN. Therefore, the CNN-BiGRU model was proven to be the best model since it was able to record the highest accuracy, the precision, the recall, and the F1-score in the classification of malicious threats. Depending on the analysed improvement, it is possible to conclude that deep learning architecture combined with feature engineering is a promising way to address the changes in the spectrum of cyber threats systematically. Therefore, the employment of the ensemble methods, the development of complex algorithms including CNN-BiGRU allows creating a reliable method to detect even

the most sophisticated phishing attempts, which may be considered as a breakthrough in cybersecurity.

Therefore, there is a need for the future work, to look into more detailed improvements and variations of the CNN-BiGRU model to analyse more and different kinds of phishing types and social engineering attacks. It remains possible to note that using extra data and active threat intelligence may be useful in improving the dynamics of the model. However, if the range of datasets is extended to include more different types of data and other new forms of machine learning such as the technique based on Transformers, one might gain additional enhancement. Further analysis of the performance of the model out of the artificial environment as well as examining its behaviour under different attacks will serve to confirm the suitability of the model and monitor its vulnerability in real-life practice.

References

[1] L. R. Kalabarige, R. S. Rao, A. R. Pais, and L. A. Gabralla, "A Boosting based Hybrid Feature Selection and Multi-layer Stacked Ensemble Learning Model to detect phishing websites," IEEE Access, 2023.

[2] N. Puri, P. Saggar, A. Kaur, and P. Garg, "Application of ensemble Machine Learning models for phishing detection on web networks," in 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), IEEE, 2022, pp. 296–303.

[3] S. D. A. Rihan, M. Anbar, and B. A. Alabsi, "Approach for detecting attacks on IoT networks based on ensemble feature selection and deep learning models," Sensors, vol. 23, no. 17, p. 7342, 2023.

[4] P. F. Akinwale and H. Jahankhani, "Detection and Binary Classification of Spear-Phishing Emails in Organizations Using a Hybrid Machine Learning Approach," in Artificial Intelligence in Cyber Security: Impact and Implications: Security Challenges, Technical and Ethical Issues, Forensic Investigative Challenges, Springer, 2022, pp. 215–252.

[5] L. Shalini, S. S. Manvi, N. C. Gowda, and K. Manasa, "Detection of phishing emails using machine learning and deep learning," in 2022 7th International conference on communication and electronics systems (ICCES), IEEE, 2022, pp. 1237–1243.

[6] S. K. Jawad and S. H. Alnajjar, "Enhancing Phishing Detection Through Ensemble Learning and Cross-Validation," in 2024 International Conference on Smart Applications, Communications and Networking (SmartNets), IEEE, 2024, pp. 1–7.

[7] C. M. Igwilo, "Ensemble Learning for URL Phishing Detection," PhD Thesis, AUST, 2020.

[8] A. Demenongo and A. Iorshase, "Ensemble Model for the Detection of Phishing URLs," Ilorin Journal of Computer Science and Information Technology, vol. 7, no. 1, pp. 1–25, 2024.

[9] M. Nasir, A. R. Javed, M. A. Tariq, M. Asim, and T. Baker, "Feature engineering and deep learning-based intrusion detection framework for securing edge IoT," The Journal of Supercomputing, pp. 1–15, 2022.

[10] A. Maini, N. Kakwani, B. Ranjitha, M. Shreya, and R. Bharathi, "Improving the performance of semantic-based phishing detection system through ensemble learning method," in 2021 IEEE mysore sub section international conference (MysuruCon), IEEE, 2021, pp. 463–469.

[11] S. S. M. M. Rahman, F. B. Rafiq, T. R. Toma, S. S. Hossain, and K. B. B. Biplob, "Performance assessment of multiple machine learning classifiers for detecting the phishing URLs," in Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19, Springer, 2020, pp. 285–296.

[12] N. F. Abedin, R. Bawm, T. Sarwar, M. Saifuddin, M. A. Rahman, and S. Hossain, "Phishing attack detection using machine learning classification techniques," in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2020, pp. 1125–1130.
[13] B. McConnell, D. Del Monaco, M. Zabihimayvan, F. Abdollahzadeh, and S. Hamada, "Phishing Attack Detection: An Improved Performance Through Ensemble Learning," in International Conference on Artificial Intelligence and Soft Computing, Springer, 2023, pp. 145–157.
[14] A. Awasthi and N. Goel, "Phishing website prediction using base and ensemble classifier techniques with cross-validation," Cybersecurity, vol. 5, no. 1, p. 22, 2022.
[15] M. Lansley, F. Mouton, S. Kapetanakis, and N. Polatidis, "SEADer++: social engineering attack detection in online environments using machine learning," Journal of Information and Telecommunication, vol. 4, no. 3, pp. 346–362, 2020.
[16] Z. Wang, Y. Ren, H. Zhu, and L. Sun, "Threat detection for general social engineering attack using machine learning techniques," arXiv preprint arXiv:2203.07933, 2022.
[17] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A novel ensemble machine learning method to detect phishing attack," in 2020 IEEE 23rd International Multitopic Conference (INMIC), IEEE, 2020, pp. 1–5.
[18] L. Njuguna and G. Kamau, "An Ensemble Learning Model for Predicting Social Engineering Pharming Attacks," International Journal of Computer Applications Technology and Research, vol. 12, pp. 32–41, 2023.
[19] A. Ramana, K. L. Rao, and R. S. Rao, "Stop-Phish: an intelligent phishing detection method using feature selection ensemble," Social Network analysis and mining, vol. 11, no. 1, p. 110, 2021.
[20] K. Nagaraj, B. Bhattacharjee, A. Sridhar, and S. GS, "Detection of phishing websites using a novel twofold ensemble model," Journal of Systems and Information Technology, vol. 20, no. 3, pp. 321–357, 2018.
[21] M. A. Shaaban, Y. F. Hassan, and S. K. Guirguis, "Deep convolutional forest: a dynamic deep ensemble approach for spam detection in text," Complex & Intelligent Systems, vol. 8, no. 6, pp. 4897–4909, 2022.
[22] D. He et al., "An effective double-layer detection system against social engineering attacks," IEEE Network, vol. 36, no. 6, pp. 92–98, 2022.
[23] B. McConnell, D. Del Monaco, M. Zabihimayvan, F. Abdollahzadeh, and S. Hamada, "Phishing Attack Detection: An Improved Performance Through Ensemble Learning," in International Conference on Artificial Intelligence and Soft Computing, Springer, 2023, pp. 145–157.
[24] "Phishing Email Detection." [Online]. Available: https://www.kaggle.com/datasets/subhajournal/phishingemails
[25] L. R. Kalabarige, R. S. Rao, A. Abraham, and L. A. Gabralla, "Multilayer stacked ensemble learning model to detect phishing websites," IEEE Access, vol. 10, pp. 79543–79552, 2022.