Journal of Informatics and
Web Engineering

Vol. 3 No. 2 (June 2024) eISSN: 2821-370X

Unveiling the efficacy of AI-based algorithms in phishing attack detection

Tajamul Shahzad[1,*], Kashif Aman[2]

[1]Riphah International University Islamabad, Pakistan
[2]Bahria University Islamabad, Pakistan
*corresponding author: (tajamul.shahzad@riphah.edu.pk, ORCiD: 0009-0008-1670-1000)

Abstract - Phishing poses a significant challenge in an ever-evolving world. The increased usage of the Internet has resulted in the emergence of a different kind of theft referred to as cybercrime. The term cybercrime describes the act of invading privacy and illegitimately obtaining personal information using digital platform. Primarily an approach named phishing is employed, which involves the use of spoof emails or bogus websites by the attackers to get the victim's personal information like their account credentials, debit, or credit card’s number, etc. To give the brief knowledge of phishing attacks and their types of the objective of this work is to investigate various AI algorithms. Through a detail literature 14 AI algorithms which are repeatedly used for detection, and these are Random Forests, Convolutional Neural Network, Naïve Bayes, K-Nearest Neighbours algorithm, Decision Trees, long short-term memory, gated recurrent unit, Artificial Neural Network, AdaBoost, Logistic Regression, Gradient Boost, Multi-layer perceptron, Recurrent Neural Network, Extreme gradient boosting, and Support Vector Machine to detect phishing attacks. To verify the effectiveness of these algorithms an experiment is performed on two datasets. Among all the algorithms Convolutional Neural Network, Multi-layer perceptron and AdaBoost achieved more than 90% accuracy, precision and sensitivity and it was showed through results that these algorithms are very efficient and can achieve high accuracy if used to the requirements of specific scenario with proper planning. Moreover, the paper shows how different AI techniques have been employed in multiple studies to detect and address phishing attacks. Also, this paper gives a complete list of current problems with phishing attacks and ideas for future studies in this area.

Keywords - Phishing attacks. Phishing website detection. Cyber Security. Deep learning. Machine Learning

Received: 13 October 2023; Accepted: 5 January 2024; Published: 16 June 2024

I. INTRODUCTION

Today, it can be seen a revolutionary transformation in different fields like communication, marketing, and banking due to internet. However, attackers are inventing diverse methods to interrupt the communication. To obtain sensitive data, these adversaries trick the user by installing malicious software or phishing websites. The act of tricking individuals online, commonly known as phishing, is one method employed by the attackers. To deceive people into becoming victims, the phisher delivers a bait that is a copy of the legitimate website. The success of the phisher occurs when someone falls for the trick by placing their confidence in the fake website. [1]. Data interchange over the Internet

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2024.3.2.9

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

has become commonplace due to the exponential advancement of technology, as well as the expansion of businesses worldwide and the establishment of offices in various locations. These developments have necessitated the need to make data usable and accessible from anywhere, which means that data sent via a could potentially result in serious security issues involving breaches of integrity, confidentiality, and authentication of data [2].

Nowadays, to remain competitive and current, organizations are embracing technology more and more. On the other hand, a higher reliance on technology and networking increases one's vulnerability to cyber risk. Establishing an internet connection enables organizations to become visible in a globalized environment where disruptions can occur without warning, leading to serious harm and monetary losses. To make matters worse, the Covid-19 outbreak hastened the transition to remote and digital work, which has increased vulnerabilities and the threat of cyberattacks [3]. Cybercriminals can use a social engineering attack named as phishing to spread ransomware, steal money, theft and commit financial fraud. Furthermore, it permits government actors to obtain essential access to places of significance. Through the development of fake websites that appear trustworthy, phishing is employed to deceive individuals into revealing personal and crucial data, including passwords or credit card numbers [4]. Most phishing attacks occur when an individual clicks on a fraudulent email link, redirecting them to a fake website. The impacts of phishing attacks can be far-reaching, causing identity theft and inflicting both psychological and financial hardships on victims [5]. With the passing of time, the quantity of such websites is constantly growing. The APWG (Anti-Phishing Working Group) received 211,032 reports of phishing incidents in the last three months of 2016. Furthermore, there was a 12% increase in the number of reports received in the final quarter of 2018, reaching a total of 239,910. The existence of fraudulent websites that deceive both individuals and businesses pose a significant issue. The attacker can make unauthorized purchases, steal money, or steal the identity of individuals using their personal details. Even though corporations permit their employees to navigate around security measures, they also put themselves at risk as employees can introduce malicious software and obtain unauthorized access to protected data within the company's network. These attacks will lead to a decrease in the organization's market share [6].

Phishing attacks consist of malware-based phishing and social engineering. Social engineering attacks frequently leverage people's vulnerabilities and thoughts in order to manipulate them into revealing private information. Phishing with malware involves the manipulation of your computer through malicious software or unwanted applications. This malevolent program employs a key logger and screen logger to save the keystrokes and webpages you interact with. These various attacks include phone phishing, session hijacking, key loggers, content-injection phishing, link manipulation, DNS phishing and system reconfiguration [7]. Artificial Intelligence (AI) has significantly influenced numerous industries, including cyber-security. Fast, precise, and thorough investigation capabilities have been achieved through the implementation of AI in email security. By utilizing datasets containing previous experiences, AI can recognize spam, phishing, suspicious links, and other forms of attacks. AI can rely on stored information to identify spam, phishing, suspicious links, and different types of attacks [8].

Cyberattacks are becoming more frequent due to their ease, cost-effectiveness, and lower risk compared to physical attacks. With internet accessibility and a computer, anyone has the potential to engage in cybercriminal activities. Moreover, the enormous nature of the Internet makes it hard to identify and apprehend criminals involved in cybercrimes. Having access to the Internet and a computer is all that is necessary for participating in cybercriminal activities [9]. Phishing attacks are increasing all around the world. In December 2021, the APWG Numerous researchers have discovered various approaches to classify methods for detecting phishing. The counter measurements can be categorized into four groups: Machine Learning, Deep Learning, Scenario-based Techniques, and Hybrid Techniques. Deep learning encompasses the field of machine learning, aiding in the identification of patterns and enabling the development of fully autonomous systems capable of making end-to-end predictions without requiring human input [10]. (APWG) documented 316,747 attempts, a figure that stood as the highest recorded number of attacks in a month for the organization [11] which is shown in Figure 1.

The quick advancement of smart methods like deep learning and machine learning, these are part of artificial intelligence, are helpful in keeping computer operations and cybersecurity management safe. Artificial intelligence (AI) has a wide range of abilities, like finding and predicting patterns, keeping things secure, and adjusting to new surroundings. These skills are crucial in important technology systems, including computer vision [12].

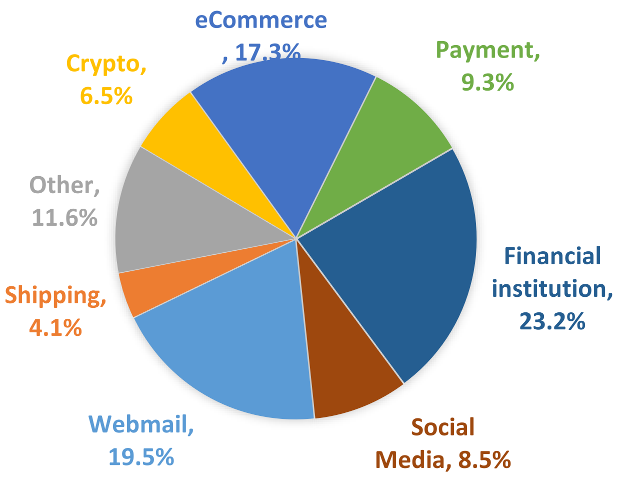

Figure 1: Phishing Attacks on Industries in 2021 [11]

According to available information, there are limited surveys in previous research that provide a broad understanding of techniques for detecting attacks. These surveys lack information on a wide range of deep learning, hybrid learning, machine learning, and scenario-based learning techniques. Nevertheless, these surveys fail to adequately address the ongoing and future obstacles associated with identifying phishing attacks. Considering the mentioned restrictions, this article presents the following contributions:

- Offer a brief overview of diverse phishing strategies for detection including machine learning, scenario-based learning, hybrid learning, and deep learning.
- There is ample discussion revolving around the diverse ways people employ to deceive others on the internet, including the act of pretending to be someone they are not.
- Furthermore, compare the varieties of results obtained from different research studies regarding these kinds of attacks.
- Describe how phishing attacks are currently detected, outline the difficulties encountered, and mention the research endeavors being conducted in this domain.
- Performed testing of these algorithms on two random datasets to verify the effectiveness of these.

The work is split into different parts. In Section I, the concept of phishing attacks is elaborated. Section II discusses literature about various methods for identifying phishing attacks, including deep learning, scenario-based detection, hybrid learning and machine learning. Section III gives brief introduction about different approaches of AI used in Literature. Section IV addresses Discussion. Section V Test Experiment on datasets and Section VI is about conclusion and recommendations for further research.

II. LITERATURE REVIEW

The term "phishing" describes the act of misleading individuals into providing their online passwords and login information, like luring fish with bait. During the 1970s, "ph" was derived from the practice of "phone phreaking," a widely favored technique for exploiting phone systems. In 1996, some attackers used the word "phishing" for the first time online when they tricked AOL (unaware America Online) employees into revealing their passwords to steal accounts on the service. Phishing is a kind of automated identity theft that preys on behavior of people and the Internet to deceive millions of people out of their money. [13]. The most recent AI contribution to detect and prevent phishing attacks are as follows.

In [14] the authors utilized C4.5, Rotation Forest, KNN, Artificial Neural Network (ANN), SVM, and RF for phishing attacks detection. They briefly detailed about how these classifiers can identify phishing attacks with better accuracy. They indicate that the RF can only be accurate to a maximum of 97.26%. Similar accuracy was obtained by all other classifiers as stated in the study.

The authors in [15], proposed methods and conducted tests on more than two datasets. First place goes to Phishtank, which lists 1528 phishing websites. Second place goes to Openphish, which lists 613 phishing sites. Third place goes to Alexa, which lists 1600 legitimate sites. Fourth place goes to a payment gateway, which lists 66 sites. Fifth place goes to banking websites, which lists 252 web sites. They increased phishing detection accuracy by using machine-learning algorithms. They used NN, NB, LR, RF, SVM, and neural networks. On the client-side, a feature extraction strategy was used and achieved 99.09% accuracy.

Today, email continues to be the most popular method of disseminating phishing scams. Because of this, phishing email detection has been regarded as a crucial problem in the field of cybersecurity. The authors in [16], To identify phishing emails, deep learning techniques including CNN, LSTM, RNN and MLP were combined with neural bag-of-ngrams and word embedding. To elicit semantic and syntactic similarity across emails, word embedding and Neural Bag-of-Ngrams are both helpful. At IWSPA-AP 2018, all experiments were conducted using a shared task corpus for anti-phishing. Each model did well during the training phase and the obtained accuracy was 99.1%.

In [17] the scholars proposed a methodology that primarily uses two procedures: identifying and extracting login credentials. According to the suggested methodology, after image processing, the logo was extracted from the image using a two-dimensional code. The identity detection process then evaluated the consistency between the website's described identity and its actual identity. The website is legitimate if the identity is real; otherwise, it is a phishing site. They extracted data from 726 web pages and created two separate, non-overlapping datasets. Both trustworthy and phishing websites are included in the dataset. While the phishing pages are taken from Phishtank, the legitimate pages are taken from Alexa and by applying the proposed methodology they achieved 98.3% accuracy.

In [18], heuristics, visual similarity, Blacklist and whitelist, are the three methods used in the hybrid solution was proposed by the authors. The suggested methodology keeps track of all activity on the user's computer and evaluates each website link against a list of trustworthy websites. The website evaluates different information for attribute. There are three types of websites: phishing websites, legitimate websites, and suspicious websites. The AI-based machine learning classifier is used to collect data and generate a score. They immediately blacklisted the URL as a phishing attempt if the score was higher than the criteria. To predict the accuracy of their test websites, they used LR, DT, and RF and they achieved 96.58% accuracy.

In [19], the authors employed a dataset from the UCI machine learning repository that consists of 2456 unique URL instances, 11,055 total URLs, 6157 were phishing websites, and the legitimate webpages were 4898. They took 30 characteristics out of URLs and used those characteristics to identify the phishing attack. There were two outcomes: either the user needed to be informed that the website was a phishing scam or they already knew the website was secure. They applied machine learning (ML) algorithms like Gradient Boosting (GBM), RF, PCA, Generalized Linear Model (GLM), and DT and the achieved accuracy was 98.40%.

While feature engineering plays a significant role in detection solutions of phishing websites, the effectiveness of the detection heavily depends on the features' prior knowledge but extracting such features are time taking. To overcome these drawbacks the authors in [20] suggest a MFPD methodology named as (multidimensional feature phishing detection), which detects using deep learning technique. First, sequence features of character of the provided URL are retrieved and then used for rapid classification by deep learning, this phase doesn't need outside help or any prior knowledge of phishing. At stage 2, we combined rapid classification of deep learning output with statistical data of URL, webpage text features, webpage code features, and multidimensional features. The method can decrease the amount of time needed to detect something. The accuracy is tested on a dataset with millions of legitimate and phishing URLs, and it reaches 98.99%.

In [21], the scholars used the Random Forest algorithm to identify phishing websites. When combined with the 26 attributes of phishing websites, the Random Forest technique provides high accuracy of 98.8%, which is employed for better performance.

The authors suggested a model named THEMIS, a new DL model for identifying phishing emails in [22]. The model utilizes a more advanced RCNN (convolutional neural network) that incorporates multilevel vectors as well as an attention mechanism, allowing for concurrent modelling of an email at the character, body, header, and word levels. According to the findings, THEMIS had a 99.848% accuracy rate.

Another hybrid framework was proposed in [23] by the authors named HEFS (Hybrid Ensemble Feature selection) for the detection of phishing. To produce primary feature subsets in the first stage of HEFS, an algorithm named CDF-g (Cumulative Distribution Function gradient) was used. To produce secondary subset of features the primary subset was feed into data perturbation. Results achieved by performing experiments, by combining with Random Classifier HEFS performs best, where the baseline features can successfully identify 94.6% of phishing websites.

In [24], The authors suggested a way to solve the problem of identifying phishing websites by using images and website addresses as a classification task. Convolution neural networks (CNNs) are used to extract the most crucial information from website photos and URLs, and then they are used to categories them into phishing pages and benign pages. The model uses website links and pictures to identify phishing attacks. The experiment's results have a 99.67% accuracy rate, demonstrating how well the suggested model works to identify web phishing attacks.

Moreover, in [25], Four meta-student models were employed by the authors, all of which made use of the extra-tree basis classifier: Bagging-Extra Tree (BET), LogitBoost-Extra Tree (LBET), AdaBoost-Extra Tree (ABET) and Rotation Forest-Extra Tree (RoFBET). The performance of the suggested meta-algorithms was evaluated after fitting them to datasets from phishing websites. Furthermore, the suggested models outperformed other ML-based models at detecting phishing attacks with 97% accuracy rate.

In [26], the authors aimed at Explaining how to extract content of email and behavior features, which features are suited for unsolicited bulk emails (UBEs) detection, and how to choose the most discriminating feature set. In order to effectively address the threat posed by UBEs, the authors also supported a thorough comparative analysis employing a variety of machine learning methods. Regarding categorizing UBEs, our suggested models had a 99% overall accuracy rate.

The scholars in [27], developed a detection system for content-based phishing that examines the text and other aspects of the website to determine whether it contains false content or not. To comprehend and evaluate the offered models in this strategy, the scholars used 8 various machine learning models. The proposed methodology showed best accuracy with RF with accuracy of 97.91% according to experimental data.

In [1], LSTM (Long short-term memory) models, convolution neural network (CNN), deep neural network (DNN), and other models were used by the authors to detect phishing. The used models achieved a good detection rate, with LSTM's accuracy being 99.57%. These models were made to be reliable and effective by utilizing just one third-party service feature.

The authors in [4], proposed an approach which gave classification with better accuracy between phishing and legitimate websites by the help of CNN-based algorithms. By performing numerous experiments these algorithms were quite successful to identify unknown phishing sites and the detection rate was 98.2% which is better than conventional machine learning classifier.

The authors in [28], For the purpose of identifying phishing websites, an improved stacking ensemble model was presented. Random forests, GradientBoost, AdaBoost, Bagging, XGBoost and LightGBM were among the first machine learning techniques whose parameters were optimised using a genetic algorithm (GA). The top three models were then selected as the base classifiers of a stacking ensemble approach after the optimised classifiers were graded. The experimental findings demonstrated that the suggested optimised stacking ensemble approach outperformed existing machine-learning-based detection methods, with an accuracy of 97.16%.

The goal of authors in [29] research was to lessen spam production by utilizing a classifier to identify it. Machine learning techniques allow for the classification of spam to be as accurate as possible. To identify spam in email messages, an NLP (natural language processing) method of analysis was adopted. The following machine learning algorithms were chosen for comparison: NB, DT, KNN, SVM, RF, and LR. The dataset used for training was already created. The most accurate methods were NB and logistic regression, which achieved 99% accuracy.

In [30], the CNN and RF-based integrated phishing website detection approach was proposed by the authors. Without gaining access to online content or utilizing outside services, the approach is able to determine the legitimate of URL. The suggested method converts URLs into fixed size matrices using character embedding techniques. At various phases, features were retrieved using CNN, and RF classifiers then categorized those features. The proposed model was shown to have a 99.35% accuracy rate.

Moreover, for phishing attacks detection the authors in [31], had applied different ML-based algorithms like LR, DT, RF, AB and GB. An accuracy of 97% was attained by the authors using a dataset from UCI and a novel fusion classifier.

In [32], a series of recommendations for creating extendible and repeatable datasets for the detection of phishing website was proposed by the authors. Machine learning-based systems may use constructed datasets that match the suggested principles as benchmarks. To examine previous findings, a sample dataset was gathered in accordance with the suggested guidelines. A 96.61% accuracy rate in experiments demonstrated the efficiency of using RF classifier-based algorithms in browsers to detect phishing web sites.

Although there are numerous ways to spot phishing websites, including third-party methods, URL methods, and source code methods, people are still tricked into disclosing their private information. In [33], research the authors presented a novel method for identifying word-embedded phishing websites domain-specific text and using plain text that has been retrieved from the source code. the authors used ensemble and multimodal methodologies to evaluate our model using a variety of word embeddings. A significant accuracy of 99.34% was attained using multimodal with domain-specific text, according to the experimental assessment.

In order to obtain reliable performance, the authors introduced in [34] the PDGAN phishing detection model, which solely relies on a website's uniform resource location (URL). A CNN was used to identify of phishing URL, and a LSTM network was used to create artificial phishing URLs. According to the testing findings, the PDGAN was able to detect objects with an accuracy of 97.58% without the use of other services.

Another phishing detection method was proposed by the authors in [35], named ODAE-WPDC (deep autoencoder network based website phishing detection and classification). To eliminate missing values from the dataset, the suggested ODAE-WPDC model first pre-processes the input data. After that, features extraction and selection were done by the AAA (Artificial Algae Algorithm). The classification procedure was carried out by the DAE model utilizing the received features, and improved performance was achieved by parameter modifying the DAE technique using the IWO (invasive weed optimization) algorithm. The experimental results support the better accuracy of 99.28% of the ODAE-WPDC model's performance.

In [36], the authors proposed a multi-feature extraction and DL-based phishing detection technique. To achieve self-feature this technique used MLP (multiple perception), a CNN to achieve image feature, a RNN to achieve text feature, and using a classification network to combine the features and reach the conclusion and the achieved accuracy after experiments was 97.75%.
To improve the scheme of anti-phishing techniques the authors proposed a predictive model based on ML in [37]. This model included a module called Feature Selection for feature vector’s creation. SVM and NB, which were trained on a 15-dimensional feature set, are used in the suggested model. According to the experimental findings, an excellent outcome with 99.96% accuracy was achieved.

In [38], the authors proposed a phishing detection method in which AV-BMEO (AV-shape transfer function), KNN classifier, and BMEO (Binary Modified Equilibrium Optimizer) were used. The feature selection and classifier's hyperparameter optimization were done using the high exploration and exploitation capabilities of AV-BMEO. To improve the suggested algorithm's exploration capabilities, the AV-shape transfer function was created via opposition-based learning. By applying the suggested model, they got 97.64% accuracy.

In the field of cybersecurity, it might be challenged to quickly identify newly created phishing websites. To solve these challenge, the authors in [39] suggested an anti-phishing technique based on hybrid feature to extracts only client-side's features only. Additionally, the authors created a fresh dataset for tests using machine learning classification methods. Their test results demonstrated that the suggested phishing detection method is superior to conventional methods, with a detection accuracy of 99.17% using the XG Boost technique.

Moreover, DL techniques acquire feature at character level from URLs while neglecting the links between words. The authors in [40], suggested a CNN based methodology named HDP-CNN (highway deep pyramid convolutional neural network) that mixed character and word level representation of data. HDP-CNN take input from string sequences of URL, after that embedding of character and word level is performed. From experiments it was showed that their methodology achieved 98.30% accuracy.

Another research direction was semantic counterfeiting of phishing websites. To overcome this issue The authors in [41], suggested three deep learning-based phishing detection models like MIF (Multi-scale In depth Fusion) model, MFF (Multi-scale Feature-layer Fusion) model, and MDF (Multi-scale Data-layer Fusion) with varying semantic levels. All the three models showed high approaches, according to experimental results on a generated complicated dataset, however the MIF model performs best on a complex dataset, with an AUC value of 99.93%.

Users may have their private information accessed, which puts them at risk of financial loss or identity theft. A system that effectively detects phishing websites must therefore be created. LSTM, CNN, and finally an LSTM-CNN-based strategy are three separate deep learning-based techniques that the authors suggested in [12] for the detection of phishing websites. Through Experimental results it was shown that the proposed strategies were accurate; their respective accuracy rates for CNN, LSTM-CNN, and LSTM are 99.2%, 97.6%, and 96.8%.

The hacker obtains access to your information when you read and respond to the spam email that you received from them. It has become a significant issue for everyone in recent years. To overcome this issue the authors in [42] applied various phishing and valid data sizes, identified newly emails, and applied various algorithms like NB, SVM and LSTM for classification and characteristics. According to comparison it was showed that NB, SVM and LSTM have better performance in terms of detection of phishing attacks. With accuracy rates of 97%, 99.62%, and 98%, respectively, NB, SVM, and LSTM classifiers were used to categorize email threats.

Moreover, to reduce the limitation of existing techniques for phishing attacks detection in [43] the authors proposed methodology named HELPHED and the aim of this methodology was phishing email detection by using Ensemble learning methods (ELM) having features hybrid. As proposed by authors HELPHED two methods Stack Ensemble learning and Soft Voting Ensemble Learning were used. To improve the performance of model both methods used two various ML algorithms for the controlling of hybrid features concurrently and independently and thus reducing the complexity of features. A 99.42% accuracy rate in experimental tests demonstrated that the overall detection performance is improved when hybrid features and ensemble learning are combined.

For the detection of phishing scams the authors proposed another method in [11] named CNN-Fusion a powerful and compact character-level CNN that can distinguish between benign and harmful URLs by extracting multi-level features from raw URLs without the use of specialized knowledge or any other services. The authors used SpatialDropout1D because it strengthens the model and prevents it from memorizing the training data. Form Experiments it was showed that proposed method is superior to existing ones in that it requires 5 times less training time and consumes significantly more memory, with an average accuracy of above 99%.

Table 1 represents which classifier is used by the researcher in their work year wise and how much that classifier achieved the accuracy.

Table 1: Phishing Identification Using AI-based Approaches.

References	Year	Focus	Classification Technique	Accuracy
[14]	2017	Phishing	RF, ANN, C4.5, RF, SVM, KNN,	97.36%
[15]	2018	Phishing	LR, NN, SVM, RF	99.09%
[16]	2018	Phishing	LSTM, MLP, RNN, CNN	99.1%
[17]	2018	Phishing	-	98.3%
[18]	2018	Phishing	LR, DT, RF	96.58%
[19]	2018	Phishing	DT, RF, GBM	98.40%
[20]	2019	Phishing	LSTM, RNN, CNN	98.99%
[21]	2019	Phishing	RF	98.8%
[22]	2019	Phishing	RCNN, THEMIS	99.848%
[23]	2019	Phishing	NB, RF, PART, C4.5, SVM,	94.6%
[24]	2020	Phishing	CNNs	99.67%
[25]	2020	Phishing	ABET, BET, RoFBET, LBET	97%
[26]	2020	Phishing	RF	99%
[27]	2020	Phishing	NB,RF, K-NN,DT, Multilayer perceptron, XGBoost, SVM, LR,	Highest Accuracy RF: 97.91%
[1]	2020	Phishing	CNN, DNN, LSTM	99.57%
[4]	2020	Phishing	CNN	98.2%
[28]	2021	Phishing	GA, RF, AdaBoost, LightGBM, GradientBoost Bagging, XGBoost	97.16%
[29]	2021	Spam	NB, DT, KNN, SVM, RF, LR,	99%
[30]	2021	Phishing	RF, CNN	99.35%
[31]	2021	Phishing	LR, SVM, KNN, LG, DT, RF, EL, NB, AB, GB,	97%
[32]	2021	Phishing	DT, RF, LG, SVM, NB	96.61%
[33]	2022	Phishing	SVM, RF, LR, DT, XGBoost	99.34%
[34]	2022	Phishing	PDGAN	97.58%
[35]	2022	Phishing	ODAE-WPDC	99.28%
[36]	2022	Phishing	CNN, RNN MLP	97.75%
[37]	2022	Phishing	SVM, NB	99.96%
[38]	2022	Phishing	KNN	97.46%
[39]	2022	Phishing	XG Boost	99.17%

[40]	2022	Phishing	CNN, HDP-CNN	98.30%
[41]	2022	Phishing	MIF, MDF, MFF	99.93%
[12]	2023	Phishing	LSTM, CNN, LSTM-CNN	CNN: 99.2%, LSTM–CNN: 97.6%, LSTM: 96.8%
[42]	2023	Phishing	SVM, NB, LSTM	SVM:99.62% , NB:97%, LSTM:98%
[43]	2023	Phishing	HELPHED	99.43%
[11]	2023	Phishing	CNN-Fusion	99%

III. AI ALGORITHMS USED IN LITERATURE FOR PHISHING DETECTION

The process of recognizing a cyberattack depends on the method employed to identify it. While there are many different algorithms that can be used to ensure accuracy, there are differences in their detection efficiency. In order to construct a comprehensive phishing detection model, several researchers have employed different strategies, which are listed in this section, as shown in Table 6. The most common techniques include the following:

- **ADABOOST (AB):** By developing an array of several classifiers, the self-adaptive boosting technique known as AdaBoost that can improves the working of weak classifiers. A wide range of concerns have been raised since it automatically adjusts to the core algorithm's error rate during training through dynamic management of each sample's weight [44].
- **NAIVE BAYES (NB):** Based on the Bayesian theorem, the Nave Bayes classification approach performs better with high data dimensionality. The Bayesian classifier can figure out the result based on the information given that is most expected. It's not difficult to upgrade the probabilistic classifier at runtime by adding additional raw data [45].
- **K-NEAREST NEIGHBOR (KNN):** KNN can be simply defined as unknown categorization of data point on the basis of it nearest neighbor. The k-value, which sets the number of nearest neighbors to be taken into account and, consequently, the class of a sample data point, is used in this approach to calculate the nearest neighbor. The term "KNN" refers to a technique where the classification of a particular data point is determined by using more than one nearest neighbor. As a result of the requirement that data points be present in memory during execution, this algorithm is known as a memory-based technique [45].
- **RECURRENT NEURAL NETWORK (RNN):** For the purpose of modelling time-series data, recurrent neural networks, a variation of conventional FFNs, were first introduced in the 1980s. Since RNN units are connected in a cyclic fashion, computing current states is made easier by the ability to carry over information from earlier time steps. This has achieved good results in a number of well-established artificial intelligence tasks in the areas of speech processing, computer vision, and natural language processing [16].
- **EXTREME GRADIENT BOOSTING (XGB):** An optimised distributed gradient boosting library is XGBoost. Tianqi Chen, a PhD student at the University of Washington working on a research project, developed XGBoost. XGBoost gained popularity in the machine learning community after taking first place in multiple contests [46].
- **ARTIFICIAL NEURAL NETWORK (ANN):** To find hidden patterns and relationships within a dataset, this method uses a group of algorithms that simulate how the human brain functions. When we refer to neural networks, we imply neuronal systems, whether they be biological or artificial [47].

- **GATED RECURRENT UNIT (GRU):** A simplified variant of LSTM is GRU. It is built with a more straightforward architecture that combines the gates and integrates the states [48].
- **GRADIENT BOOST (GB):** Machine learning techniques such as gradient boosting include classification and regression. A final prediction model could be combined with a sizable number of decision-making bodies using this technique [49].
- **MULTI-LAYER PERCEPTRON (MLP):** A classification technique for artificial neural networks (ANNs) has been developed using multilayer perceptron (MLP). Z-score normalization has been performed before to the algorithm training using a "Normalizer" node because MLP requires normalized data as input. The test data have then been subjected to the same methodology. The highly effective family of nonlinear statistical models known as MLPs is made up of layers upon layers of connected nodes in a directed graph. Input, hidden, and output layers are the three different types of layers. As a result, every node—aside from the input nodes—is a neuron (or other processing element) with a nonlinear activation function [50].
- **SUPPORT VECTOR MACHINES (SVM):** For text classification techniques, this classifier provides a quick and effective supervised approach. An ideal two-dimensional line for separating categories is called a hyperplane, which is produced by the input training set. Decision boundary is represented by this hyperplane [51].
- **LOGISTIC REGRESSION (LR):** A possibility of a binary response (features) is estimated using a binary logistic model using one or more predictor variables. It permits indicating that a risk factor increases the possibility of a particular outcome by a certain proportion [51].
- **DECISION TREE (DT):** This method uses a model of actions and effects that resembles a tree, encompassing utility, resource costs, and outcomes of chance events. Using this method, a conditional-only algorithm can be presented. In operation research, particularly decision analysis, DT is used to identify the most likely path to achieving a goal. Additionally, it is a widely utilized machine learning method [51].
- **RANDOM FOREST (RF**): With this method, many DT results are combined to provide a single result. It has been more popular as a classification and regression tool due to its ease of use and adaptability [51].
- **LONG-SHORT TERM MEMORY (LSTM):** A type of recurrent neural network (RNN) called LSTM (long short-term memory) achieves better outcomes when working with time-series data by removing vanishing gradients and long-term dependencies. Input, output, and forget gates are the three components of the LSTM's cell-based architecture [12].
- **CONVOLUTIONAL NEURAL NETWORK (CNN):** A particular type of neural network that needs a lot of labelled data for training is a convolutional neural network. CNNs are important in solving numerous issues, including picture classification, object recognition, phishing detection, and disease diagnosis. For the purpose of building a CNN, the primary layers required are input, convolution, pooling, and fully linked layers [12]. List of abbreviations which are used in literature are explained in the following Table 2.

Table 2: List of abbreviations used in the literature.

Algorithm	**Abbreviation**
Convolutional Neural Network	CNN
Random Forest	RF
Decision Tree	DT
Support Vector Machine	SVM
Naive Bayes	NB
K-Nearest Neighbours algorithm	KNN
Long short-term memory	LSTM
Gated recurrent unit	GRU
Artificial Neural Network ANN	ANN
Logistic Regression	LR
AdaBoost	AB
Gradient Boost	GB
Multi-layer perceptron	MLP
Recurrent Neural Network	RNN
Extreme gradient boosting	XGB
Artificial Neural Network	ANN

Table 3 presents an impressive visual representation, exemplifying the application of advanced AI technology in academia. With the synergy of human intelligence and AI capabilities, it investigates numerous research topics across various domains. This table exhibits the diverse applications of AI in scientific research. AI algorithms are very important. Comparatively, the algorithms can be viewed as the instruments in a correlation. The following segment of the table illustrates the strong compatibility between researchers' ideas and these algorithms.

Table 3: AI algorithms used by the researchers to build phishing detection.

References	CNN	SVM	LSTM	RF	DT	KNN	XGB	NB	AB	GB	LR	ANN	MLP	RNN
[14]	✓	✓	-	✓		✓	-	-	-	-	-	✓	-	-
[15]	-	✓	-	✓	-	-	-	✓	-	-	✓	-	-	-
[16]	✓	-	✓	-	-	-	-	-	-	-	-	-	✓	✓
[18]	-	-	-	✓	✓	-	-	-	-	-	✓	-	-	-
[19]	-	-	-	✓	✓	-	-	-	-	-	-	-	-	-
[20]	✓	-	✓	✓	-	-	-	-	-	-	-	-	-	✓
[21]	-	-	-	✓	-	-	-	-	-	-	-	-	-	-
[22]	-	-	-	-	-	-	-	-	-	-	-	-	-	✓
[23]	-	✓	-	✓	-	-	-	✓	-	-	-	-	-	-
[24]	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
[25]	-	-	-	✓	-	-	-	-	✓	-	-	-	-	-
[26]	-	-	-	✓		-	-	-	-	-	-	-	-	-
[27]	-	✓	-	✓	✓	✓	✓	✓	-	-	✓	-	✓	-
[1]	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-
[4]	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
[28]	-	-	-	✓	-	-	✓	-	✓	✓	-	-	-	-
[29]	-	✓	-	✓	✓	✓	-	✓	-	-	✓	-	-	-
[30]	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-
[31]	-	✓	-	✓	✓	✓	-	✓	✓	✓	✓	-	-	-
[32]	-	✓	-	✓	✓	-	-	✓	-	-	-	-	-	-
[33]	-	✓	-	✓	✓	-	✓	-	-	-	✓	-	-	-
[36]	✓	-	-	-	-	-	-	-	-	-	-	-	✓	✓
[37]	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-
[38]	-	-	-	-	-	✓	-	-	-	-	-	-	-	-
[39]	-	-	-	-	-	-	✓	-	-	-	-	-	-	-
[40]	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
[12]	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-
[42]	-	✓	✓	-	-	-	-	✓	-	-	-	-	-	-
[11]	✓	-	-	-	-	-	-	-	-	-	-	-	-	-

IV. DISCUSSION

Phishing involves using deceitful methods to tricks individuals to get their sensitive details, such as usernames and passwords, with the intention of unlawfully obtaining their confidential information. Phishers capitalize on people's emotions and willingness to follow instructions. Phishing is a very common and ongoing danger in the online world. To deceive individuals and successfully execute their scams, attackers in phishing consistently introduce new tricks. Above section presents a comparison between different techniques used in research. In the past, techniques such as deep learning, machine learning, scenario-based, and hybrid approaches rooted in AI have been employed to identify and mitigate phishing attacks. After studying and comparing different techniques, researchers found that ML-based learning methods are the most used and successful ways to identify a phishing attack. To classify items, the above approaches have been utilized. Using techniques that reduce the number of features results in improved performance and we can detect phishing attacks. In our work we have mentioned 14 AI-based classifiers namely RF, SVM, LSTM, ANN, RNN, MLP, GB, AB, XGB, NB, DT, ANN, k-NN which are widely used in literature, and we have showed in literature that by using these classifiers the researcher has got better accuracy with low false positive rate.

In our work we have showed that in how many references a classifier is used and with what accuracy rate. The RF method proves to be the most precise and effective compared to other classification methods across a range of datasets. It has been demonstrated through literature that the RF classification technique is effective among other classifiers in identifying attacks with a precision of more than 97% and used in 13 citations in our literature. 10 references specifically mention SVM, making it the second most frequently cited classifier. Additionally, it performs exceptionally well, boasting a highest recorded success rate of 99.62%. CNN is cited in 8 references with high accuracy, reaching an impressive, reported rate of 99.57%. There are six sources that mention LSTM. With an accuracy range of 96.8% to 99.1%, it is a popular option for sequence data even if it is not the most accurate. Five references mention LR, which performs well in detection of phishing scenarios with 97% accuracy rate. KNN, whose accuracy ranges from 94.6% to 97%, is referenced four times.

A popular probabilistic classifier for text classification and spam detection is Naïve Bayes (NB) receives four mentions in some references and has an accuracy of between 94.6% and 99.96%. Although ANN isn't as often quoted as other classifiers, it is noted in a few references. There are differences in its claimed accuracy. MLP is employed in several scenarios of phishing detecting and is referenced a few times with fluctuates precision. Although gradient boosting isn't as popular as other classifiers, it is referenced in a few sources. 2 publications mention AdaBoost (AB), and different studies use it in different ways. The DT is referenced 13 times and occasionally used in conjunction with other classifiers. XGBoost is used for phishing detection and is referenced 5 times. With a maximum recorded accuracy of 99.96%, it delivers great precision. An artificial neural network with numerous layers of nodes is called a multilayer perceptron (MLP). It is applied to supervised learning tasks, such as classification. It is listed with other classifiers in reference [27] and also it is used in 4 references, where the test accuracy is 93%. RNNs) and CNNs are combined to form RCNNs. It's applied to image analysis and sequential data tasks. It is used in reference [22], but the precise accuracy is not stated. According to reference [28], GA is a search heuristic with an accuracy of 97.16% that is used in conjunction with other classifiers. Reference [11] mentions CNN-Fusion, which reaches 99% accuracy.

In the past researchers had utilized the widely used dataset named UCI machine learning to identify Incidents of phishing attacks. Additionally, the researchers utilized a scenario-based approach in various studies to detect phishing attacks. However, these solutions can only be implemented in specific contexts. Each person in a company acts in their own way, and people in the company sometimes know about these situations. The technique of identifying phishing attacks employs an alternative method called hybrid learning. Sometimes, it can be more accurate than using a RF in certain situations. Researchers believe that using multiple models together can help improve their performance even more. Defending against phishing attacks has become a difficult task for system experts of security systems nowadays. It is essential to have a highly accurate system that can effectively detect the attack with minimum false rate. The defense techniques that have been discussed are based on algorithms that use AI technology to learn. Despite their costly nature and tendency to make errors, these methods are effective in identifying phishing attacks. An informed and vigilant employee is your best line of defense against phishing attempts. However, they are still people and curiosity are one of their inherent qualities. Their desire to learn more and explore is strong. Organizations should aim to prevent employees from accessing their essential processes to reduce the risk that they may fall victim to phishing attacks and foster a mentality in their staff that would prevent them from clicking on bogus links and websites.

V. EXPERIMENT ON DATASETS

To show the performance and accuracy of AI-classifiers mentioned in literature we tested these classifiers on two random datasets dataset1 is Web page phishing detection" and dataset2 is "Phishing Dataset for Machine Learning: Feature Evaluation". Dataset1 has 11430 URLs with approximately 87 extracted features. The purpose of the dataset is to serve as a standard for phishing detection systems that rely on machine learning. Features are divided into three classes: twenty-six are derived from the content of the corresponding pages, twenty-six are derived from the structure and syntax of URLs, and seven are derived from external service queries. With precisely 50% phishing and 50% genuine URLs, the dataset is equally balanced. Dataset2 includes 48 features taken from 5000 authentic websites and 5000 fake websites. By applying AI-based classifiers as briefed in discussion section on dataset1 following are the results we achieved.

From Figure 2 it can be noted that AdaBoost and MLP give better accuracy with sensitivity and precision as compared to LR and RNN.

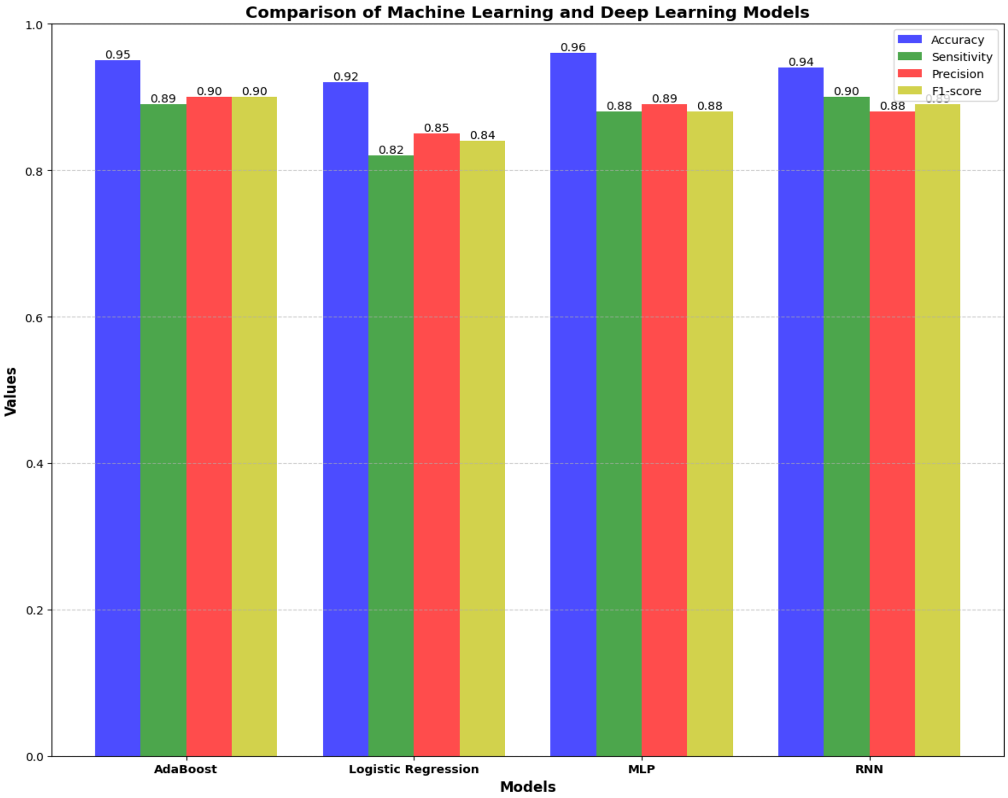

Figure 2: Comparison of AB, LR, MLP, RNN

XGB give better accuracy with high sensitivity and precision as compared to ANN, KNN, DT, NB as showed in Figure 3.

CNN outperformed among others with high accuracy, precision and sensitivity as shown in Figure 4.

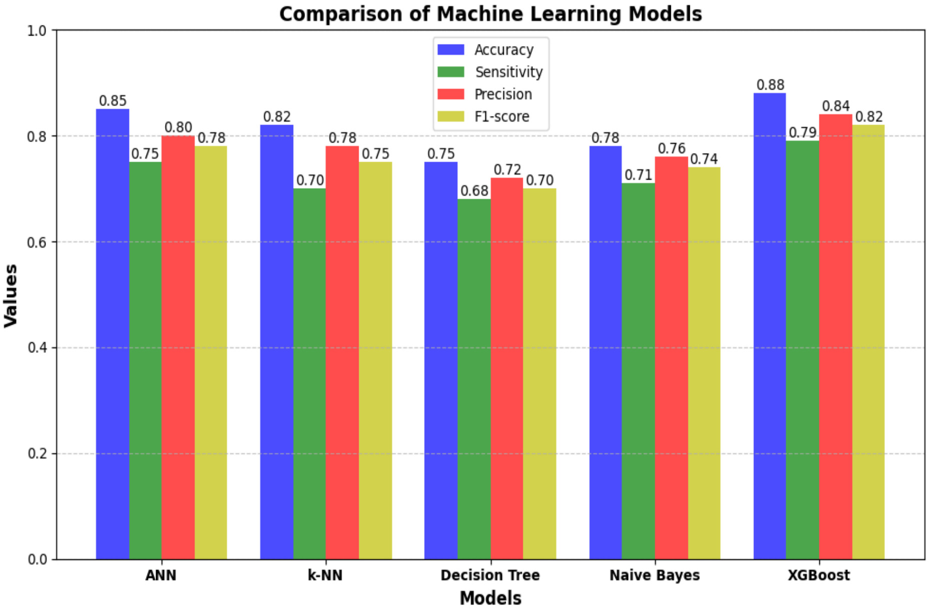

Figure 3: Comparison of ANN, KNN, DT, NB, XGB

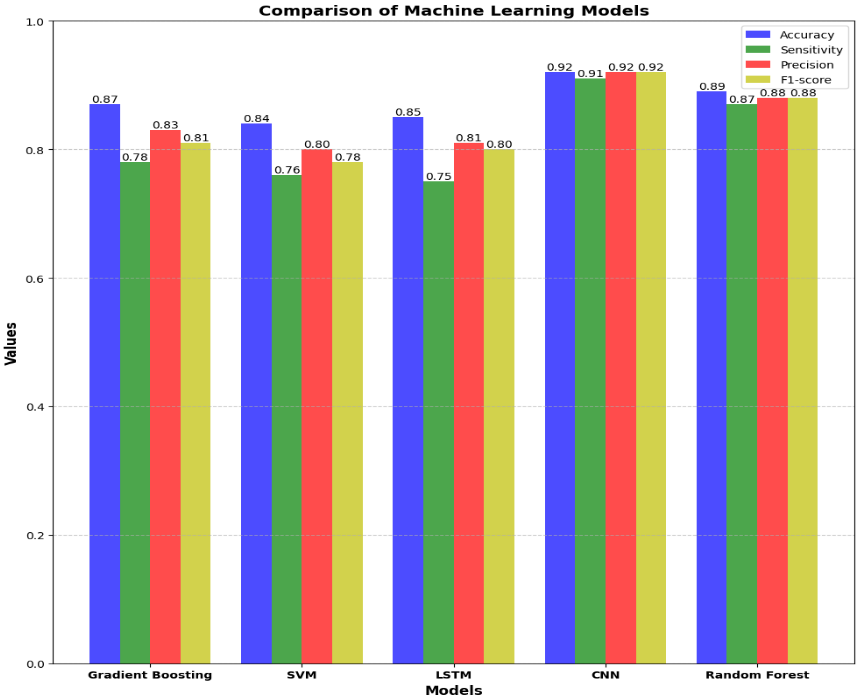

Figure 4: Comparison of GB, SVM, LSTM, CNN, RF

Following is the heat map for dataset 1 as shown in Figure 5.

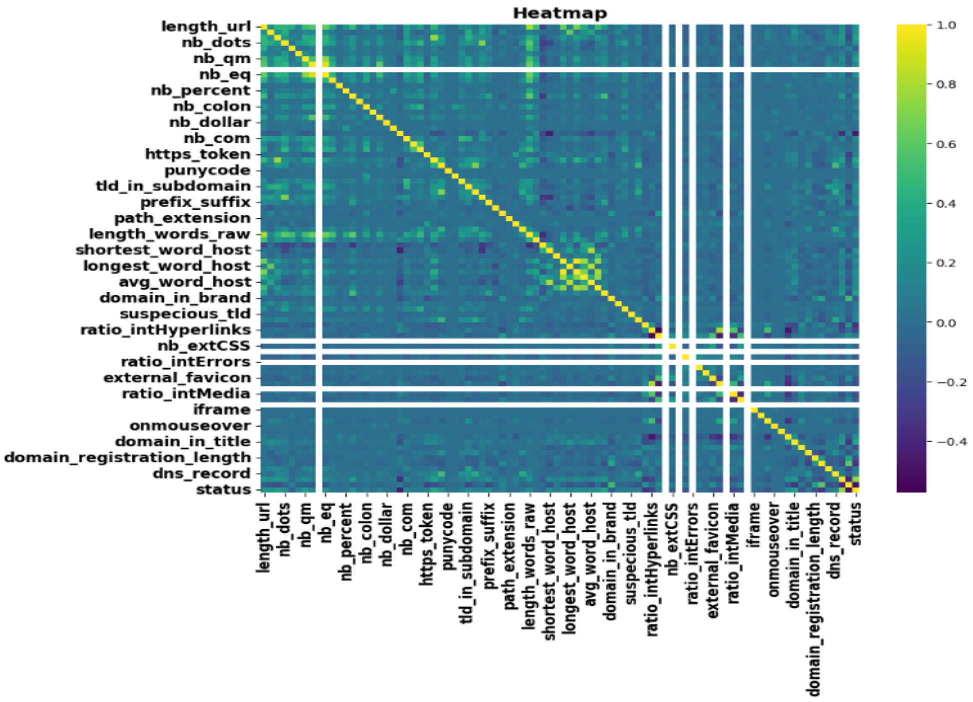

Figure 5:Correlation Matrix of dataset1

By applying classifiers LR, SVM, DT, KNN, RF, GB, AB, MLP on dataset2 except SVM and KNN all classifiers achieved more than 90% accuracy shown in Figure 6.

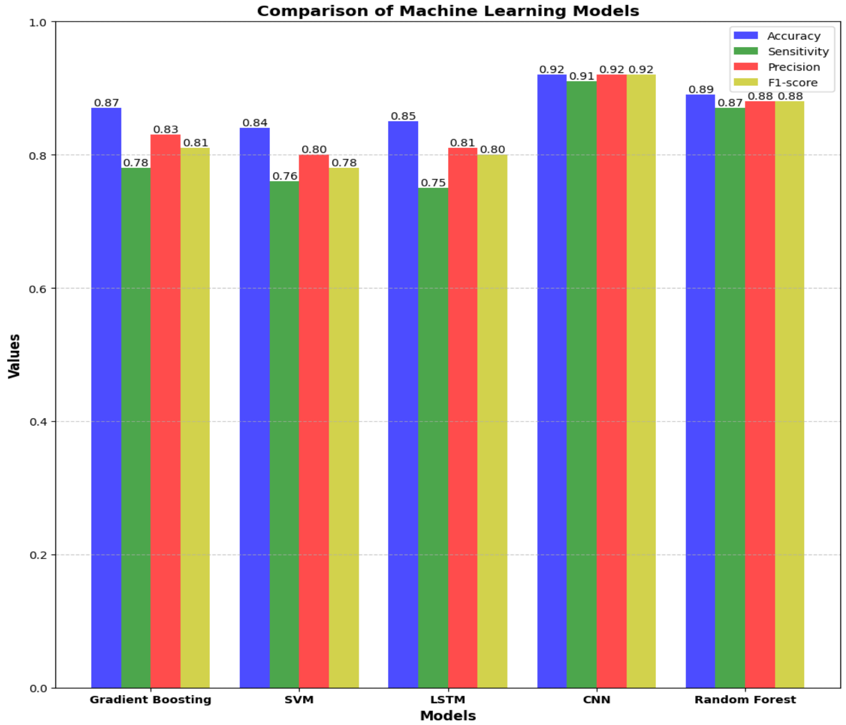

Figure 6: Comparison of Accuracy of different classifiers

While working on dataset2 following is the result we achieved. Except KNN and SVM all the classifiers outperformed with more than 90% accuracy as shown in Figure 7 (the ROC curve for dataset2).

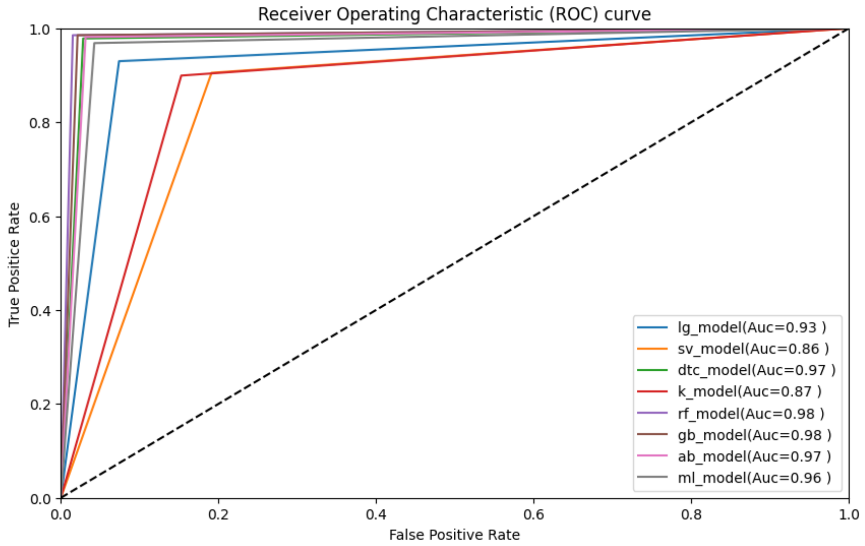

Figure 7: ROC Curve for Comparison of different classifiers on dataset2.

VI. CONCLUSION & FUTURE WORK

The purpose of this work is to assist researchers in understanding diverse methods for the detection of phishing attacks, in addition to the difficulties and patterns encompassed in such methods. It is hard to mitigate against phishing attacks in the field of system security in today's fast-paced world. The ability to easily identify phishing attacks is a key characteristic of a reliable detection system. attacks that rarely give incorrect results. Phishing attacks, a type of scam, are now being detected and prevented using artificial intelligence. Classifying data can be done through a range of methods such as RF, SVM, LSTM, ANN, RNN, MLP, GB, AB, XGB, NB, DT, ANN, k-NN. These techniques are extremely effective in detecting phishing attacks and prove to be highly useful in their identification. Increased studies can be carried out with the aim of discovering an enhanced and more reliable means of detecting whether a website is secure or involved in fraudulent activities intended to deceive individuals into disclosing their personal information. This can involve using smart tools to mark websites as either trustworthy or potentially dangerous.

ACKNOWLEDGEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Tajamul Shahzad: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Project Administration, Resources, Supervision.
Kashif Aman: Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

REFERENCES

[1] M. Somesha, A.R. Pais, R.S.b Rao and V.S.Rathour, "Efficient deep learning techniques for the detection of phishingwebsites.", Sādhanā 45 (2020): 1-18.

[2] J.F. Lai and S.H. Heng, "Secure file storage on cloud using hybrid cryptography.", Journal of Informatics and Web Engineering 1, no. 2 (2022): 1-18.

[3] T. Munusamy and T. Khodadi, "Building Cyber Resilience: Key Factors for Enhancing Organizational Cyber Security.", Journal of Informatics and Web Engineering 2, no. 2 (2023): 59-71.

[4] S.Y. Yerima and M.K. Alzaylaee, "High accuracy phishing detection based on convolutional neural networks.", In 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1-6. IEEE, 2020.

[5] C. Opara, B. Wei and Y. Chen, "HTMLPhish: Enabling phishing web page detection by applying deep learning techniques on HTML analysis.", In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.

[6] A. Abuzuraiq, M. Alkasassbeh and M. Almseidin, "Intelligent methods for accurately detecting phishing websites.", In 2020 11th International Conference on Information and Communication Systems (ICICS), pp. 085-090. IEEE, 2020.

[7] C. Singh, "Phishing website detection based on machine learning: A survey.", In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 398-404. IEEE, 2020.

[8] A. Basit, M. Zafar, X. Liu, A.R. Javed, Z. Jalil and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques.", Telecommunication Systems 76 (2021): 139-154.

[9] H.F. Atlam and O. Oluwatimilehin, "Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review, Electronics 2023, 12, 42." (2022).

[10] C. Catal, G. Giray, B. Tekinerdogan, S. Kumar and S. Shukla, "Applications of deep learning for phishing detection: a systematic literature review.", Knowledge and Information Systems 64, no. 6 (2022): 1457-1500.

[11] M. Hussain, C. Cheng, R. Xu and M. Afzal, "CNN-Fusion: An effective and lightweight phishing detection method based on multi-variant ConvNet.", Information Sciences 631 (2023): 328-345.

[12] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q.E.U. Haq, K. Saleem and M.H. Faheem, "A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN.", Electronics 12, no. 1 (2023): 232.

[13] B.B. Gupta, A. Tewari, A.K. Jain and D.P. Agrawal, "Fighting against phishing attacks: state of the art and future challenges.", Neural Computing and Applications 28 (2017): 3629-3654.

[14] A. Subasi, E. Molah, F. Almkallawi and T.J. Chaudhery, "Intelligent phishing website detection using random forest classifier.", In 2017 International conference on electrical and computing technologies and applications (ICECTA), pp. 1-5. IEEE, 2017.

[15] A.K. Jain and B.B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach.", Telecommunication Systems 68 (2018): 687-700.

[16] V. Ra, B.G. HBa, A.K. Ma, S. KPa, P. Poornachandran,and A. Verma, "DeepAnti-PhishNet: Applying deep neural networks for phishing email detection.", In Proc. 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Secur. Privacy Anal.(IWSPA), pp. 1-11. Tempe, AZ, USA, 2018.

[17] W. Yao, Y. Ding and X. Li, "Logophish: A new two-dimensional code phishing attack detection method.", In 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), pp. 231-236. IEEE, 2018.

[18] V. Patil, P. Thakkar, C. Shah, T. Bhat and S.P. Godse, "Detection and prevention of phishing websites using machine learning approach.", In 2018 Fourth international conference on computing communication control and automation (ICCUBEA), pp. 1-5. Ieee, 2018.

[19] I. Tyagi, J. Shad, S. Sharma, S. Gaur and G. Kaur, "A novel machine learning approach to detect phishing websites.", In 2018 5th International conference on signal processing and integrated networks (SPIN), pp. 425-430. IEEE, 2018.

[20] P. Yang, G. Zhao and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning.", IEEE access 7 (2019): 15196-15209.

[21] E. Benavides, W. Fuertes, S. Sanchez and M. Sanchez, "Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review.", Developments and Advances in Defense and Security: Proceedings of MICRADS 2019 (2020): 51-64.

[22] Y. Fang, C. Zhang, C. Huang, L. Liu and Yue Yang, "Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism.", IEEE Access 7 (2019): 56329-56340.

[23] K.L. Chiew, C.L. Tan, K. Wong, K.S. Yong and W.K. Tiong, "A new hybrid ensemble feature selection framework for machine learning-based phishing detection system.", Information Sciences 484 (2019): 153-166.

[24] S. Al-Ahmadi, "A deep learning technique for web phishing detection combined URL features and visual similarity.", International Journal of Computer Networks & Communications (IJCNC) Vol 12 (2020).

[25] Y.A. Alsariera, V.E. Adeyemo, A.O. Balogun and A.K. Alazzawi, "Ai meta-learners and extra-trees algorithm for the detection of phishing websites.", IEEE access 8 (2020): 142532-142542.

[26] T. Gangavarapu, C.D. Jaidhar and B. Chanduka, "Applicability of machine learning in spam and phishing email filtering: review and approaches.", Artificial Intelligence Review 53, no. 7 (2020): 5019-5081.

[27] U. Ozker and O.K. Sahingoz, "Content based phishing detection with machine learning.", In 2020 International Conference on Electrical Engineering (ICEE), pp. 1-6. IEEE, 2020.

[28] M. Al-Sarem, F. Saeed, Z.G. Al-Mekhlafi, B.A. Mohammed, T. Al-Hadhrami, M.T. Alshammari, A. Alreshidi and T.S. Alshammari,"An optimized stacking ensemble model for phishing websites detection.", Electronics 10, no. 11 (2021): 1285.

[29] Y. Kontsewaya, E. Antonov and A. Artamonov, "Evaluating the effectiveness of machine learning methods for spam detection.", Procedia Computer Science 190 (2021): 479-486.

[30] R. Yang, K. Zheng, B. Wu, C. Wu and X. Wang, "Phishing website detection based on deep convolutional neural network and random forest ensemble learning.", Sensors 21, no. 24 (2021): 8281.

[31] A. Lakshmanarao, P.S.P. Rao and M.B. Krishna, "Phishing website detection using novel machine learning fusion approach.", In 2021 international conference on artificial intelligence and smart systems (ICAIS), pp. 1164-1169. IEEE, 2021.

[32] A. Hannousse and S. Yahiouche, "Towards benchmark datasets for machine learning based website phishing detection: An experimental study.", Engineering Applications of Artificial Intelligence 104 (2021): 104347.

[33] R.S. Rao, A. Umarekar and A.R. Pais, "Application of word embedding and machine learning in detecting phishing websites.", Telecommunication Systems 79, no. 1 (2022): 33-45.

[34] S. Al-Ahmadi, A. Alotaibi and O. Alsaleh,"PDGAN: Phishing detection with generative adversarial networks.", IEEE Access 10 (2022): 42459-42468.

[35] H. Alqahtani, S.S. Alotaibi, F.S. Alrayes, I. Al-Turaiki, K. A. Alissa, A.S.A. Aziz, M. Maray and M. Al Duhayyim, "Evolutionary Algorithm with Deep Auto Encoder Network Based Website Phishing Detection and Classification.", Applied Sciences 12, no. 15 (2022): 7441.

[36] Yu, Shuaicong, Changqing An, Tao Yu, Ziyi Zhao, Tianshu Li, and Jilong Wang, "Phishing Detection Based on Multi-Feature Neural Network.", In 2022 IEEE International Performance, Computing, and Communications Conference (IPCCC), pp. 73-79. IEEE, 2022.

[37] A.A. Orunsolu, A.S. Sodiya and A.T. Akinwale, "A predictive model for phishing detection.", Journal of King Saud University-Computer and Information Sciences 34, no. 2 (2022): 232-247.

[38] S. Minocha and B. Singh, "A novel phishing detection system using binary modified equilibrium optimizer for feature selection.", Computers & Electrical Engineering 98 (2022): 107689.

[39] D. Guptta, Sumitra, K.T. Shahriar, H. Alqahtani, D. Alsalman and I.H. Sarker, "Modeling hybrid feature-based phishing websites detection using machine learning techniques.", Annals of Data Science 11, no. 1 (2024): 217-242.

[40] F. Zheng, Q. Yan, Victor C.M. Leung, F. R. Yu and Z. Ming, "HDP-CNN: Highway deep pyramid convolution neural network combining word-level and character-level representations for phishing website detection.", Computers & Security 114 (2022): 102584.

[41] D.J. Liu, G.G. Geng and X.C. Zhang. "Multi-scale semantic deep fusion models for phishing website detection." Expert Systems with Applications 209 (2022): 118305.

[42] U.A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi and A. Ahmadian, "Cloud-based email phishing attack using machine and deep learning algorithm.", Complex & Intelligent Systems 9, no. 3 (2023): 3043-3070.

[43] P. Bountakas and C. Xenakis, "Helphed: Hybrid ensemble learning phishing email detection.", Journal of network and computer applications 210 (2023): 103545.

[44] P. Wu and H. Zhao, "Some analysis and research of the AdaBoost algorithm.", In International Conference on Intelligent Computing and Information Science, pp. 1-5. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[45] I.A.A. Amra and A.Y. Maghari,"Students performance prediction using KNN and Naïve Bayesian.", In 2017 8th international conference on information technology (ICIT), pp. 909-913. IEEE, 2017.

[46] P. Carmona, F. Climent and A. Momparler, "Predicting failure in the US banking sector: An extreme gradient boosting approach.", International Review of Economics & Finance 61 (2019): 304-323.

[47] V.H. Nhu, A. Shirzadi, H. Shahabi, S.K. Singh, N. Al-Ansari, J.J. Clague, A. Jaafari et al, "Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms.", International journal of environmental research and public health 17, no. 8 (2020): 2749.

[48] A. Aldweesh, A. Derhab and A.Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues.", Knowledge-Based Systems 189 (2020): 105124.

[49] O. Wisesa, A. Adriansyah and O.I. Khalaf, "Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm.", In 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), pp. 101-106. IEEE, 2020.

[50] P.F. Orrù, A. Zoccheddu, L. Sassu, C. Mattia, R. Cozza and S. Arena, "Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry.", Sustainability 12, no. 11 (2020): 4776.

[51] H.F. Atlam and O Oluwatimilehin, "Business email compromise phishing detection based on machine learning: A systematic literature review.", Electronics 12, no. 1 (2022): 42.