Volume 1 No 1, 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Research Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

An In-Depth Review of ML and DL Approaches for Phishing Email Identification and Mitigation

Garima Mishra
Research Scholar
Department of Computer Science and Application
Mandsaur University
Mandsaur, Madhya Pradesh, India
mishragarima2708@gmail.com

Dr. Parth Gautam
Assistant Professor
Department of Computer Science and Application
Mandsaur University
Mandsaur, Madhya Pradesh, India
parth.gautam@meu.edu.in

***Abstract*—Businesses lose money, and consumers become annoyed by phishing emails, which are a major problem on the Internet these days. A form of artificial intelligence called machine learning has been shown to be a useful tool for spotting email threats; yet, there is currently no comprehensive answer to the issue of phishing email filtering. This study fills this gap by providing an extensive analysis of the most recent approaches to phishing detection, ranging from conventional ML methods to advance DL frameworks. The expanding use of artificial intelligence, namely ML and DL, in phishing email detection and mitigation is examined in this review. ML models with efficient feature selection and classification, such as Random Forest, Decision Tree, and SVM, exhibit high accuracy, while DL architectures like DNNs, CNNs, RNNs, and LSTMs excel in automated feature extraction and sequential data analysis, offering scalability and adaptability for real-time detection. When Natural Language Processing (NLP) and hybrid models are used, DL techniques demonstrate encouraging accuracy and resilience, lowering false positives and improving security. The study emphasizes the effectiveness of DL and ensemble models in thwarting advanced phishing efforts as well as the significance of AI-driven layered defenses in stopping the phishing lifecycle.**

Keywords—Phishing Detection, Email Spam, Machine learning (ML) and Deep Learning (DL) Approaches, Mitigation.

I. Introduction

Nowadays, most individuals can't imagine living without the Internet. Trying to picture life before the Internet is an exercise in futility. There are 4.66 billion people employing the Internet now, constituting 59.5% of the global populace, according to research on the global digital population published in January 2021. The majority of these people, namely 92.6%, access the Internet using their smartphones [1]. Communication, shopping, talking, and office work are just a few examples of how the Internet has revolutionized people's lives and careers. Traditional businesses, such as catering and retail, have moved their services online in response to a pandemic that began at the end of 2019. Users' personal details, financial details, account credentials, privacy questions, and passwords have been among the many vulnerable pieces of information that internet users have left behind. For the purpose of engaging in illicit online activity, cybercriminals get this information via a variety of illicit channels and masquerade as these individuals. The advent of the Internet brought with it new security concerns for networks. Numerous threats to network security have emerged in tandem with the evolution of Internet attack methods [2][3]. Network attack techniques and forms primarily classify cybersecurity challenges into the following: DoS [4], MitM [5], Threats, malware, DNS tunnelling, SQL injection, zero-day exploits, and phishing [6][7].

The deadly assault known as "phishing" may affect anybody, from individuals to whole countries. This social engineering technique is employed by criminals to deceive people into disclosing private information by posing as trustworthy sources [8][9]. According to the APWG, the second quarter of 2023 saw 1,286,208 phishing attempts, setting a new record. The financial sector is obviously at the forefront of this problem, since it is the target of a staggering 23.5% of all phishing assaults [10]. An overwhelming majority of malware infections begin with phishing emails (82% of all cases), and social engineering is often the first tactic used by cybercriminals. Phishers employ a variety of communication channels to launch their assaults; the most prevalent of them are email, social media, text, and phone calls [11][12]. In addition to the 320 billion unwanted emails sent daily, this is also the vector via which 94% of malware is distributed. Unsolicited commercial emails were sent to business email accounts, resulting in financial losses of an estimated $12 billion.

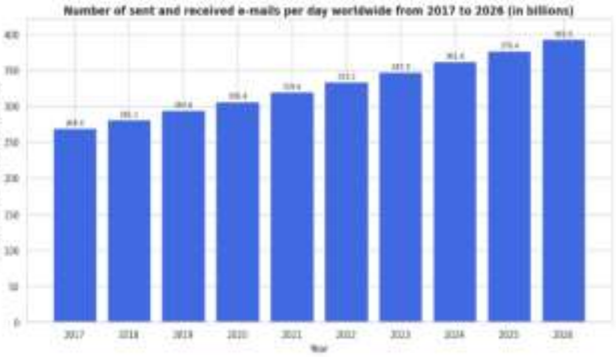

Fig. 1. Worldwide, Everyday Spam Emails[13]

Figure 1 displays the worldwide spam email volume for January 16, 2023, as reported by Statista. At 8.6 billion, the US topped the list, followed by the Czech Republic at 7.7 billion and the Netherlands at 7.6 billion.

The phrase "email phishing" is used to describe communications that are intentionally harmful. It was in 2018 that a famous email phishing attack happened [14]. Hackers were able to get sensitive information from naive users by getting them to open phishing emails and click on links inside [15]. To stop phishing efforts and stop crimes, it is crucial to recognize these emails [16]. Numerous businesses concentrate on enhancing their email security protocols using a variety of techniques [17]. One way to strengthen defence against email-based attacks is to implement subdomain limitations, It entails setting up a special domain just for email security [18].

Furthermore, it is essential to educate users and analyze phishing attack histories in order to guarantee the safety of people and organizations [19]. Email systems have substantial and intricate hurdles when it comes to spam detection and filtering. The shortcomings of content-based solutions, real-time blackhole listings, and blocklists highlight the need for new approaches to identification. To get around these limitations and make spam detection more accurate, more advanced ML and DL techniques have been developed. There has been a lot of talk on the potential of ML and DL approaches to enhance email spam detection. Therefore, a thorough literature study is essential to creating a current, fact-based knowledge of current research on using these techniques to combat this enduring issue [20][21]. Finding and evaluating several ML and DL strategies for spam detection is the main goal of this study. It will also suggest topics for future research that address any gaps in their current understanding. The strengths and shortcomings of current approaches will be determined by integrating and evaluating results from various investigations. The goal of this project is to use the most modern ML and DL algorithms to find trustworthy and effective ways to identify spam emails. Here are the main elements of the review article:

- Phishing attacks are growing in complexity, extending beyond emails to include QR codes, fake apps, SMS (smishing), and voice phishing (vishing), making conventional detection techniques less and less efficient.
- The power of AI and ML in particular to analyse massive datasets has made them essential for phishing detection, find previously unseen trends, and adapt to new phishing tactics.
- ML techniques such as RF, DT, and SVM use structured data analysis and efficient feature extraction to classify phishing emails with high accuracy.
- DL models like DNNs, CNNs, RNNs, and LSTMs automate feature learning and excel in handling sequential and unstructured data (e.g., email content, URLs), providing real-time detection capabilities.
- DL approaches integrated with Natural Language Processing (NLP) and hybrid architectures (e.g., CNNRNN) enhance the understanding of context and improve detection accuracy while reducing FP.
- The research highlights the significance of security frameworks driven by multi-layered AI that can identify and counteract phishing attempts throughout the whole phishing lifecycle.

A. *Structure of the Paper*

The structure of this paper is as follows: In Section II, it learns about phishing attempts. Section III discusses the types of phishing attacks; Section IV explores ML and DL approaches, including advantages and techniques. Section V reviews relevant literature studies, and Section VI concludes with limitations and future research directions.

II. Understanding of Phishing Attacks

Common cyberattack tactics like "phishing" attempt to deceive people into disclosing important information by fabricating an email or conversation that seems to be from a trustworthy source [22]. An effort at phishing goes through many stages, as seen in Figure 2. An evildoer will first construct Phishing websites that seem a lot like the genuine thing. One side of the coin consists of attackers who construct valid URLs (particularly domain names and network resource directories) using tactics such as misspellings or similar alphabetic letters. The URL "https://aimazon.amz-z7acyuup9z0y16.xyz/v" (found on May 9, 2021) is a spoof of https://www.amazon.com. After clicking on a link, the user's browser displays the URL address, however most users have trouble distinguishing between these and actual URLs just by looking at or memorising them. There is a lot of significance to the issue of online content imitation. Web layouts, logos, information, and more may be stolen from legitimate websites by attackers using scripts. Most commonly, cybercriminals pose as legitimate-looking form submission sites in order to trick consumers into divulging critical information [23].

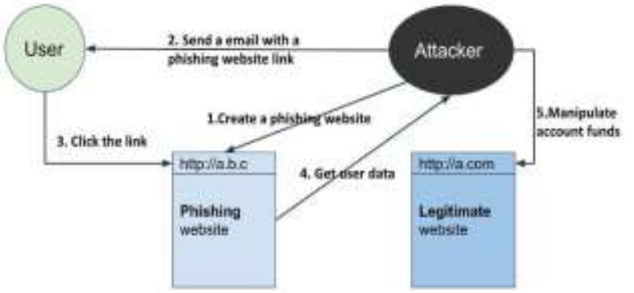

Fig. 2. Phishing Life Cycle [24]

Phishing attacks secondly include persuading victims to click on harmful links that are sent in several methods, including as text messages, phone calls, emails, QR codes, and phoney mobile apps, among others. Attackers take advantage of the proliferation of cell phones and social media by luring victims with misleading text and graphics. For instance, scammers may impersonate telecom customer service to pressure users into making payments. These messages usually use social engineering strategies, such as creating a sense of urgency, trust, or terror, to deceive recipients into clicking on links to bogus websites. These websites impersonate trustworthy companies and steal personal information, including login passwords and payment details, by using similar user interfaces, logos, and content. Once the information is submitted, attackers can access victims' accounts, especially if credentials are reused across platforms, and may use the stolen data for further criminal activities. As phishing evolves with internet technologies shifting targets like online payments, its impact remains severe, accounting for nearly 30% of cybercrime complaints and over $54 million in losses in 2020 alone. Thus, recognizing phishing websites is critical, and users need visual tools to distinguish fake sites from legitimate ones.

A. *Anti-Phishing*

Figure 2 illustrates the five steps an attacker must take before using user data improperly or stealing money from their account. Thereby, a phishing attempt might be prevented by preventing any step. In this article, it will go over the anti-phishing strategy at each stage [25].

1) Web Scraping

Although criminals' ability to create websites is hard to prevent, there are a number of approaches that might increase their cost. Phishing assaults are carried out by hackers using scripts to create crawlers. These crawlers then automatically scrape important data from legitimate websites and paste it into hoax websites [26]. Consequently, genuine websites might use various methods of data obfuscation, such as CSS sprites, to shield sensitive information from web scrapers by portraying it as visuals rather than text.

2) Spam Filter

Preventing users from opening or clicking on links in unwanted emails is the primary function of spam filtering solutions. It can filter spam with any of the popular email services like Gmail, Yahoo!, Outlook, or AOL. First filters were based on blacklists, whitelists, and empirical criteria [27]. To further eliminate non-listable spam, some filters use intelligent prediction models grounded on ML, a feature made possible by advancements in AI technology. For instance, consider Gmail's spam filter; it employs ML to thwart an additional 100 million spam emails daily.

3) Detecting Fake Websites

People, especially those working for new or start-up businesses, often forget the domain name, making it impossible for consumers to differentiate the difference between a genuine website and a phishing page that looks exactly like it. Chrome, one of many online browsers with built-in security features to identify malicious or phishing websites, alerts users with warning messages whenever it accesses such a site. In 2007, Google introduced Google Safe Browsing, which is now a part of other Google products including Gmail and Google Search. One security feature of Google Chrome is Safe Browsing, which uses a blacklist to block malicious or fraudulent websites. Furthermore, several browser addons exist to identify fraudulent websites. Problematically, unknown phishing websites cannot be addressed by methods that rely on blacklists or whitelists [28]. Unfortunately, new ideas and methods for detecting phishing efforts have surfaced due to the rapid advancement of AI technology. The ML-based prediction approach transcends the boundaries of present regulation by having the ability to identify phishing URLs that aren't on the whitelist.

4) Second Authorization Verification

The criminal may visit the real website, take over the account, and steal funds after obtaining sensitive user details. Instant action must be done to confirm the user's identity if a website notices a discrepancy between the logged-in user's IP address and device details and their previously used information. Additionally, the verifications are frequently biological and dynamic, such voiceprint analysis or the identification of facial movements or mood.

III. TYPES AND TECHNIQUES OF PHISHING ATTACKS

Psychological manipulation or technological approaches are used by phishers to assault people and get them to provide personal information. False positives based on human psychology are more often used by phishers than technological measures. Figure 3 below shows an types of attacks of email phishing as follows:

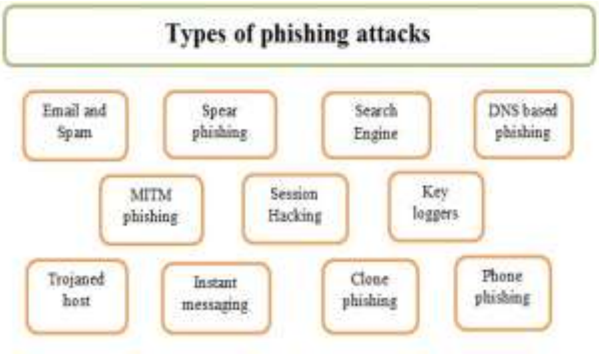

Fig. 3. Types of Phishing Attacks[29]

There are many varieties of phishing assaults, including:

A. Email and Spam:

The vast majority of phishing attempts take this form. To get people to click on harmful links or download infected files, attackers send out mass emails pretending to be legitimate organizations. Typical subject lines for these emails include account breach or urgent payment demands in an effort to scare the recipient into taking immediate action.

B. Spear Phishing:

Spear phishing attempts to target specific individuals, in contrast to the generalized phishing attacks. Everyone or any group may have it customized. In order to send convincing messages, the attacker researches the target's personal details on social media and other platforms. Cybercriminals often use these seemingly legitimate emails for business espionage or to breach protected systems.

C. Search Engine Phishing:

In this method, cybercriminals create fake websites that appear in search engine results. These websites seek to collect personal information, including social security numbers, passwords, and credit card details, by mimicking legitimate corporate portals. Victims are tricked into visiting these sites by searching for services or deals online.

D. DNS-Based Phishing:

The goal of this assault, which goes by the name "pharming," is to reroute consumers from a safe website to a malicious one secretly by manipulating the Domain Name System (DNS) [30]. The victim thinks they are visiting a genuine site, but their information is being captured by attackers. It is particularly dangerous because it can affect multiple users simultaneously without requiring individual targeting [31].

E. MITM (Man-In-The-Middle) Phishing:

This attack occurs when a hacker overhears a two-person discussion, usually between a user and a website. The message is surreptitiously relayed or modified by the attacker. It is commonly used on unsecured Wi-Fi networks where attackers can eavesdrop on online transactions or steal login credentials.

F. Session Hacking:

A user's private information might be compromised when hackers take advantage of current sessions by taking the session ID, which allows for illegal access to the user's account. Once the session is compromised, attackers can perform unauthorized actions like fund transfers or data theft as if it were the legitimate user.

G. *Trojaned Host:*

Trojan horses are harmful programs that masquerade as useful ones [32]. When a user installs the Trojan, it gives attackers remote access to their system. This access can be used to monitor activity, steal data, or deploy additional malware. It's a silent and highly effective form of phishing.

H. *Instant Messaging Phishing:*

Also called smishing or IM phishing, this method uses instant messaging platforms to send malicious links or attachments. The attacker may pose as a known contact or organization, persuading the victim to click on harmful content. With the widespread use of apps like WhatsApp, this technique has become increasingly prevalent [33].

I. *Clone Phishing:*

The tactic known as "clone phishing" is sending a user an email that seems like it came from a trusted source but has malicious attachments added. The modified email is then sent from an address resembling the original sender. Victims are often tricked because the email appears familiar and trustworthy [34].

J. *Phone Phishing (Vishing):*

In this type, attackers use phone calls instead of emails or messages. It impersonates officials from banks, tax departments, or tech support, trying to extract sensitive information like bank details or OTPs. The technique is based on linguistic manipulation of the victim via social engineering.

IV. Machine And Deep Learning for Phishing Email Detection

Cybersecurity and phishing attempts are only two areas that have been impacted by the widespread use of AI [35]. AI has improved email security by making investigations faster, more accurate, and more comprehensive. ML may use datasets to identify many forms of assault, including spam, phishing, and spear phishing. Trust in social services, such as online ones, is likely to take a hit when these kinds of assaults occur [36][37]. The shortcomings of conventional phishing detection techniques have recently been shown to be solved by methods based on DL and ML [38]. ML methods may be utilized to train models that identify phishing emails [39]. These programs may study large databases on the topic and pick out phishing trends and characteristics. It is necessary to identify key characteristics of phishing attempts before training can begin [40]. This often requires knowledge of the subject matter and a meticulous selection of key characteristics that lead to effective detection algorithms. With DL, unlike ML algorithms, important features may be automatically extracted from raw data. Due to their state-of-the-art capabilities, DNN are already finding effective applications in several fields [41]. Neural networks are useful for text and picture categorization because it can handle and learn from massive volumes of data [42][43]. This research seeks to determine the optimal method for detecting email phishing using DL and ML approaches [44]. This enables the identification of the optimal designs for ML and DL as well as well as fresh perspectives on the relative advantages of different phishing detection models and recommendations for choosing suitable techniques for building practical phishing detection systems based on thorough empirical evaluation [45]. The flowchart of email phishing websites employing ML and DL is illustrated in Figure 4 below:

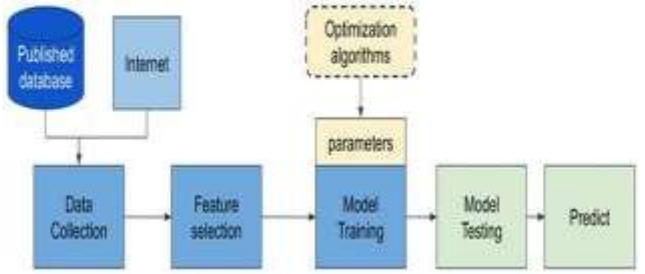

Fig. 4. Flow chart for identifying Email Phishing websites using machine learning and deep learning [24]

A. *Data Collection and Data Preprocessing*

The data serves as the foundation for every method and has a significant impact on the results. Extracting URLs from the Internet and loading publicly available datasets are the two main ways to get data [46]. Preprocessing data is getting the raw data ready to be used in a machine learning model. An essential step in making it easier to derive valuable insights is improving the data's quality via preprocessing.

B. *Feature Selection*

A "feature selection" is an automated training procedure for ML models that prioritizes which attributes are most valuable. Including relevant characteristics in the input may lead to enhanced model performance, quicker training (especially for DL models), and less overfitting. A feature selection algorithm may take one of three primary forms: the filter, wrapper, or embedding methods.

C. *Machine Learning Modeling*

The two most common types of models built using ML are supervised and unsupervised. Due to labelled data, models may be trained to make predictions or classifications in supervised learning, which depends on known input-output pairs [47][48]. Finding groupings, structures, or patterns in data without labels is the purpose of unsupervised learning. Binary categorisation issues come up while attempting to identify websites that are phishing. A number of widely used classification methods are shown below.

- **SVM:** Supervised learning algorithms, or SVMs, divide data points into two groups and then utilize those groups to forecast future data points. Having two labelled classes and a hyperplane classifier with N features (which is proportional to the number of features) makes it ideal for linear binary classification. When training the SVM model using the UCI dataset, for instance, it gets two classes: phishing and genuine, along with a hyperplane with 29 dimensions.
- **Decision Tree:** A common ML approach, decision trees are structured like the model logic. There are feature nodes, feature values, and possible outcomes shown at each stem of the decision tree, with the result being presented at the very end.
- **Random Forest:** For regression and classification, a random forest combines several decision trees. Through training-process tree classification or average output, random forests mitigate the overfitting issue. So, in comparison to decision tree algorithms, random forests tend to be more accurate.
- **k-NN:** K-NN is a non-parametric classification technique that computes the distance between the target and its closest neighbours to identify comparable

data points and produce predictions [49]. Real-time scenarios are thus often inappropriate for this method.

- **Bagging:** If you're doing a regression or classification analysis using many ML algorithms, you may boost their performance via bagging, which is another name for bootstrap aggregating [50]. The bootstrapping method uses resampling methods to make sure that each component of the initial training dataset has the same size, then conducts classification in N rounds to make sure that parallel execution is possible. At last, the aggregating mechanism averages or votes together the results of N classifiers.
- **Naive Bayes:** A probabilistic statistical strategy exhibiting strong independent properties, NBC is according to the Bayes theorem. The theory of conditional probability is known as Bayes' theorem. Common names for it include independence, Bayes and simple Bayes. ML algorithms that were utilized to detect phishing attempts are summarized in Table 1.

Table 1: Machine Learning Stems for Detecting Phishing Emails

Algorithms	Dataset	Accuracy
Random forest [51]	Website URL features	Acc: 99.50%
Bayes Net (naive Bayes) [52]	Phishing Corpus and Spam Assassin	Acc: 92%
Bootstrap aggregating + logistic model tree [53]	UCI	Acc: 97.42%
SVM[54]	Phishing Corpus and Spam Assassin	Acc: 97.25%
KNN [55]	Symantec's enterprise emails	F1: 90%, FPR: 0.1
Decision Tree Ensemble (C4.5 and CART)[56]	anti-phishing website	Acc: 99.27%

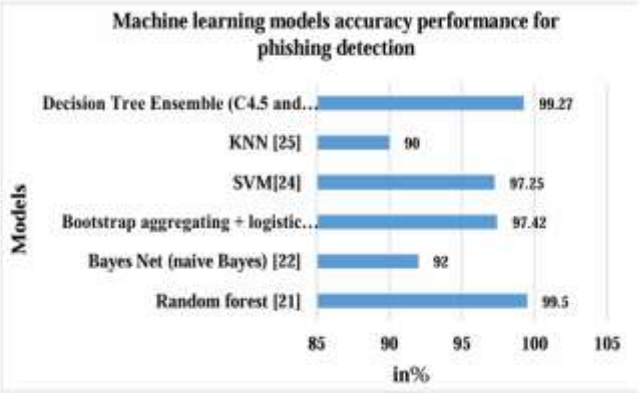

Fig. 5. Bra graph of ML accuracy performance for phishing detection

Figure 5 is a bar graph that shows how different algorithms for phishing detection performed on different datasets in terms of accuracy. RF reached the greatest accuracy of 99.50% utilising website URL characteristics, followed closely by the Decision Tree Ensemble (C4.5 and CART) with 99.27% on an anti-phishing website dataset. Bootstrap Aggregating combined with a Logistic Model Tree reached 97.42% accuracy on the UCI dataset, while SVM attained 97.25% using the Phishing Corpus and Spam Assassin data. Naive Bayes (Bayes Net) showed a comparatively lower accuracy of 92% on the same dataset. K-Nearest Neighbors (KNN) applied to Symantec's enterprise emails demonstrated strong performance with an F1score of 90% and a low FPR of 0.1, emphasizing its effectiveness in practical scenarios.

D. Deep Learning-Modelling

The use of deep structured architectures in the construction of ML is known as deep learning. A few popular DL algorithms are CNNs, RNNs, and LSTM networks. Several DL-based methods for phishing detection have been presented due to the fast development of DL algorithms and NLP [57][58]. The fundamental structure of methods based on DL is displayed in Figure 6.

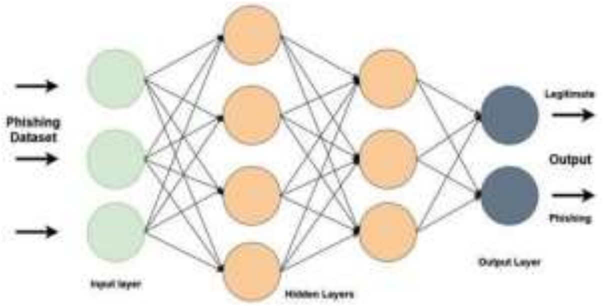

Fig. 6. Deep learning for phishing detection[59].

Researchers have proposed a plethora of DL models, capitalizing on the popularity of NLP and the rapid development of DL. Even without features retrieved from web page source codes, these models are able to infer information and sequential patterns by URL strings. It doesn't need cybersecurity experts to have knowledge of phishing as it relies on third-party services to capture features. A number of well-known DL algorithms are listed below.

- **DNN:** As an alternative, DNNs are feedforward neural networks that have several hidden layers; these networks are able to simulate complicated data connections and are often used for regression and classification assignments. While DNNs excel with structured input, LSTMs are better suited for data with temporal or sequential patterns.
- **CNN:** One popular feedforward deep learning approach for picture categorization is the CNN[60]. The traditional architecture of a CNN includes input, hidden, and output layers. Convolutional, pooling, and completely linked layers are commonly used to construct hidden layers.
- **RNN:** RNNs are DNN that can process inputs of various lengths, including text, thanks to their internal memory function. Consequently, text mining has made good use of it.
- **LSTM:** Time series prediction and NLP are two applications that greatly benefit from LSTM networks, a kind of RNN that is specifically intended to grasp long-term relationships in sequential data. A summary of DL techniques that may identify phishing attempts is shown in Table 2.

Table 2: Deep Learning Systems for Detecting Phishing Emails.

Algorithms	Dataset	Accuracy
DNN [61]	ISCX-URL-2016	97.45
CNN [62]	Websites	95.02%
RNN [63]	URL phishing	94.26
LSTM [64]	URL phishing	93.28
Recurrent Neural Network (RNN) + Convolutional Neural Network (CNN) [65]	Websites	95.79%
Genetic algorithm (GA) + DNN [66]	UCI	89.50%

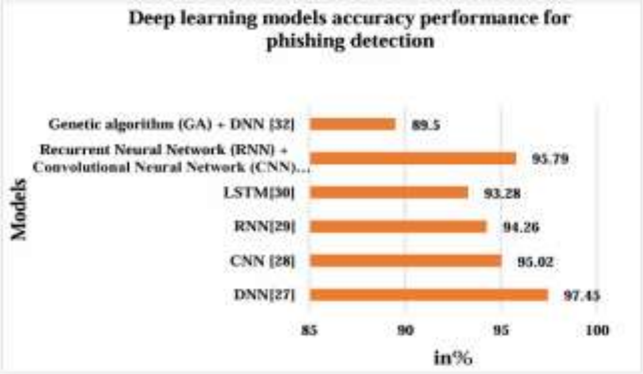

Fig. 7. Bra graph of DL accuracy performance for phishing detection

The performance of several DL algorithms to detect phishing attempts is displayed in Figure 7. On the ISCX-URL-2016 dataset, the DNN showed the best prediction ability with an accuracy of 97.45%. A combination of RNN and CNN also performed well, with 95.79% accuracy on website data, slightly outperforming standalone CNN (95.02%) and RNN (94.26%) models. LSTM, designed for sequence-based data like URLs, reached an accuracy of 93.28%. In contrast, the Genetic Algorithm combined with DNN yielded the lowest accuracy at 89.50% on the UCI dataset, suggesting that while optimization techniques can support performance, this may not always surpass standalone DL models. The findings show that DL models, especially DNNs and hybrid architectures, are good at detecting phishing attempts.

E. Performance Evaluation

The performance throughout the testing process. It is common practice to save 20% of the initial dataset for testing and utilize 80% for training. It is used for statistical metrics to measure TP, TN, FP, and FN labelled by the model, indicating the classifier's performance on the testing dataset. To that end, there are a variety of widely used metrics at disposal. To measure how accurate the predictions are in comparison to the total number of predictions, it uses the classification accuracy:

- **Accuracy:** The proportion of accurate predictions to all predictions is one way to measure a classification system's accuracy.
- **Recall:** The number of times the model has properly identified the data is called the recall.
- **Precision:** The accuracy of the model is the number of correctly identified positive data points among all possible ones.
- **F-measure:** The F-measure, which is often called the F score, is, in the end, the recall sum precision.

1) Advantages of ML and DL for phishing email detection and mitigation

The following are some of the benefits of using ML and DL to detect and prevent phishing emails:

- High Accuracy and Adaptability By discovering patterns in massive volumes of email data, ML/DL models may get very high detection accuracy. It can generalize well to detect both known and previously unseen (zero-day) phishing attacks.
- Automated Feature Extraction (in DL) There is no longer any need for human feature engineers thanks to DL models that use architectures such as CNNs and LSTMs to automatically extract complicated features from unprocessed email content.
- Detection in Real Time Training ML/DL models allows for real-time detection of phishing emails, allowing for quick reaction and mitigation of phishing assaults.
- Scalability Highly scalable, these models can handle and analyze massive amounts of email for businesses of any size.
- Understanding the email's context and semantics is something that deep learning models, particularly those based on natural language processing (e.g., BERT, RNNs), which aids in distinguishing subtle phishing efforts from genuine correspondence.
- Reduced False Positives Advanced models can learn nuanced differences among phishing and legitimate emails, reducing false alarms and improving user trust in the detection system.
- Integration with Security Infrastructure ML/DL-based system integration with firewalls, spam filters, and other cybersecurity resources may be used to build a multi-layered defence system.

V. Literature Review

The aim of this section is to provide a comprehensive literature review on the subject of detecting and avoiding phishing emails using ML and DL techniques. To offer a brief overview, Table 3 provides a summary of the papers that were examined.

Alhuzali et al. (2025) aims to fill that need by creating a new paradigm for developing and testing algorithms that can identify phishing emails. A combined dataset developed for this work and nine publically accessible datasets make up the 10 datasets used to assess fourteen ML and DL models. Various performance measures are used in the examination to make sure it's a thorough comparison. The experimental findings show that when comparing accuracy and resilience, DL models are always better than ML ones. More specifically, as compared to traditional ML approaches, transformer-based models RoBERTa (99.08% accuracy) and BERT (98.99% accuracy) get better results on the combined balanced dataset, beating them by 4.7% on average [67].

Zhao and Jin (2024) trace the development of phishing email content and investigate how it affects the efficacy of detection algorithms. It suggests Fewshing, a method for detecting phishing emails using a few-shot learning technique. The experimental results show that Fewshing achieves an F1score of 92.4% and an accuracy of 98.6% on the limited and imbalanced training datasets, which demonstrates Fewshing's efficacy in identifying phishing emails. Attacks using phishing emails continue to jeopardise people's personal information and possessions [68].

Qi et al. (2023) propose two new under-sampling methods that rely on Identification of Fisher Markov phishing ensembles: the FMPED technique and the FMMPED method. Algorithms under-sample remaining innocuous emails after filtering out overlapping ones. The emails are then trained and classified as either phishing or benign using ensemble learning techniques. Results from experiments show that the suggested algorithms perform better than competing ML and DL algorithms, with an F1score of 0.9945, an accuracy of 0.9945, an AUC of 0.9828, and a Gmean of 0.9827 [69].

Ripa, Islam and Arifuzzaman (2021) looks at how a Twitter spear phishing bot that uses ML can detect phishing URLs, emails, and websites. Several classifiers were utilized to detect phishing URLs; for optimal performance, the dataset was trained with a view towards time. The XGBoost classifier was shown to be more accurate (94.44% of the time) and to have a quicker processing time. In terms of identifying phishing emails, NBC reached a success rate of 95.15 percent. Their website identification approaches used several classifiers, with the RFC yielding the highest accuracy at 96.80% [70].

Zhang and Wu (2020) This study lays forth a systematic approach to detecting convincing phishing emails. A fundamental idea in persuasion is to search the inbox for the word that matches the characteristic. In the end, 25 features are produced for detection after an information gain technique is applied to each feature. "Phishing emails" are malicious emails with links that lead to impersonated websites or pages with malicious HTML code inserted into them. Financial data, account numbers, and login credentials are among the sensitive pieces of information that aim to gain unauthorised access. When these emails were ultimately examined, their accuracy percentage was shown to be 99.6% [71].

Loh, Lee and Balachandran (2024) created and implemented a hybrid security architecture that uses generative and AI-assisted platforms to identify phishing attempts and provide ongoing end-user education. This is their contribution to the research. This platform enables dynamic and scenario-specific user education for handling ever-more complex phishing email assaults. Both systems' functional aspects and technical architecture are shown and explored. According to performance testing, the CNN DL model architecture had the best overall performance, phishing attack detection sub-system, which had over 94% accuracy, over 95% precision, and over 94% recall [72].

Table 3: Summary of the Related Work for Email Spam Phishing Attacks Detection Using ML-DL Models

Study	Study Focus	Methodology	Datasets	Approaches /Models Used	Challenges	Limitations	Future Work
Ahluwalia et. al (2025)	Develops and evaluates a novel framework for phishing email detection	Experimental evaluation across 14 ML/DL models	9 public datasets + 1 merged balanced dataset	BERT, RoBERTa, traditional ML models	Handling varied data sources, ensuring robustness	May require high computational resources for DL models	Refinement of a framework for real-time deployment
Zhao and Jin (2024)	Analyzes phishing email evolution & proposes a Few-shot learning model (Fewshing)	Few-shot learning approach with minimal training data	Limited, imbalanced datasets	Fewshing model	Detection with limited labeled data	Performance on large-scale or diverse data is unexplored	Extend Fewshing for multilingual/phishing variants
Qi et al. (2023)	Proposes novel under-sampling ensemble methods	Fisher–Markov based ensemble techniques	Custom datasets for under-sampling	FMPED and FMMPED algorithms	Overlapping benign/phishing regions	Potential overfitting from reduced data	Apply to real-time phishing scenarios
Ripa, et. al. (2021)	Detects phishing in URLs, emails, and websites	Supervised learning with multiple classifiers	Twitter-based phishing data	XGBoost, Naïve Bayes, Random Forest	Handling different phishing vectors	Dataset size and scope limitations	Incorporate deep learning and real-time detection
Zhang et. al. (2020)	Detection based on persuasion principles in phishing	Feature selection using information gain	Text-based phishing email data	Persuasion principle + IG-based selection	Identifying relevant linguistic cues	May not generalize across all phishing styles	Test on larger, multilingual datasets
Loh, Lee et. al. (2024)	Proposes integrated AI framework for detection and user education	Hybrid of CNN-based detection + Gen AI for education	Not explicitly specified	CNN model; Generative AI-based education	Adapting to advanced, evolving phishing tactics	Dataset and platform details limited	Expand customizable scenarios and real-time feedback

VI. Conclusion and Future Work

Phishing tactics, which involve psychological manipulation and technology methods, are a significant and ongoing concern in cybersecurity. The goal is to deceive people into handing over vital information. This research shows that ML and DL techniques are effective in detecting and avoiding phishing emails. ML models like RF and DT have shown excellent performance in structured classification tasks, while DL models like DNNs, CNNs, and LSTMs provide superior accuracy and automation by learning complex patterns from raw data. Integrating these models into security infrastructures significantly improves real-time detection and minimizes false positives, making them highly effective against both known and emerging phishing threats. While ML and DL models have significantly advanced phishing detection, it face limitations such as dependence on labeled data, lack of interpretability, high computational demands, and vulnerability to adversarial attacks. Future research should aim to develop lightweight, explainable algorithms that can respond to new dangers as they emerge in real-time. Incorporating techniques like Explainable AI (XAI), user behavior analytics, and hybrid ML-DL-rule-based systems can enhance model transparency, scalability, and resilience against sophisticated phishing tactics.

References

[1] J. Clement, "Global Digital Population as of April 2020," 2020.

[2] S. S. S. Neeli, "Critical Cybersecurity Strategies for Database Protection against Cyber Attacks," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 1, no. 1, p. 5, 2023.

[3] S. Duary, P. Choudhury, S. Mishra, V. Sharma, D. D. Rao, and A. Paul Aderemi, "Cybersecurity Threats Detection in Intelligent Networks using Predictive Analytics Approaches," in *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, Feb. 2024, pp. 1–5. doi: 10.1109/ICIPTM59628.2024.10563348.

[4] S. B. Shah, "Machine Learning for Cyber Threat Detection and Prevention in Critical Infrastructure," *J. Glob. Res. Electron. Commun.*, vol. 2, no. 2, pp. 1–7, 2025, doi: 10.5281/zenodo.14955016.

[5] N. Patel, "AI-Enhanced Zero Trust Security Architecture for Hybrid and Multi-Cloud Data Centers: Automating Trust Validation, Threat Detection, and Mitigation," *Int. J. Nov. Trends Innov.*, vol. 3, no. 1, pp. a13–a18, 2025.

[6] P. Saravanan and S. Subramanian, "A Framework for Detecting Phishing Websites using GA based Feature Selection and ARTMAP based Website Classification," in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.04.116.

[7] S. Arora, S. R. Thota, and S. Gupta, "Data Mining and Processing in the Age of Big Data and Artificial Intelligence - Issues, Privacy, and Ethical Considerations," in *2024 4th Asian Conference on Innovation in Technology (ASIANCON)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ASIANCON62057.2024.10838087.

[8] Y. Wei and Y. Sekiya, "Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3224781.

[9] S. Pandya, "A Machine and Deep Learning Framework for Robust Health Insurance Fraud Detection and Prevention," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, pp. 1332–1342, Jul. 2023, doi: 10.48175/IJARSCT-14000U.

[10] M. Shah, P. Shah, and S. Patil, "Secure and Efficient Fraud Detection Using Federated Learning and Distributed Search Databases," in *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ICAIC63015.2025.10849280.

[11] N. Abdelhamid, F. Thabtah, and H. Abdel-Jaber, "Phishing detection: A recent intelligent machine learning comparison based on models content and features," in *2017 IEEE International Conference on Intelligence and Security Informatics: Security and Big Data, ISI 2017*, 2017. doi: 10.1109/ISI.2017.8004877.

[12] S. Murri, "Data Security Challenges and Solutions in Big Data Cloud Environments," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 06, Jun. 2022, doi: 10.14741/ijcet/v.12.6.11.

[13] E. H. Tusher, M. A. Ismail, M. A. Rahman, A. H. Alenezi, and M. Uddin, "Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open Research Problems," *IEEE Access*, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3467996.

[14] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, and R. Alaiz-Rodríguez, "Phishing websites detection using a novel multipurpose dataset and web technologies features," *Expert Syst. Appl.*, 2022, doi: 10.1016/j.eswa.2022.118010.

[15] O. Christou, N. Pitropakis, P. Papadopoulos, S. McKeown, and W. J. Buchanan, "Phishing URL Detection Through Top-level Domain Analysis: A Descriptive Approach," in *International Conference on Information Systems Security and Privacy*, 2020. doi: 10.5220/0008902202890298.

[16] A. Yasin and A. Abuhasan, "An Intelligent Classification Model for Phishing Email Detection," *Int. J. Netw. Secur. Its Appl.*, vol. 8, no. 4, pp. 55–72, 2016, doi: 10.5121/ijnsa.2016.8405.

[17] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, "Phishing Detection System Through Hybrid Machine Learning Based on URL," *IEEE Access*, 2023, doi: 10.1109/ACCESS.2023.3252366.

[18] A. Safi and S. Singh, "A systematic literature review on phishing website detection techniques," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 2, pp. 590–611, 2023, doi: https://doi.org/10.1016/j.jksuci.2023.01.004.

[19] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, "Defending against phishing attacks: taxonomy of methods, current issues and future directions," *Telecommun. Syst.*, 2018, doi: 10.1007/s11235-017-0334-z.

[20] S. Pahune and M. Chandrasekharan, "Several Categories of Large Language Models (LLMs): A Short Survey," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 7, pp. 615–633, 2023, doi: 10.22214/ijraset.2023.54677.

[21] R. Tarafdar, "AI-Powered Cybersecurity Threat Detection in Cloud," *Int. J. Comput. Eng. Technol.*, 2025.

[22] R. Alabdan, "Phishing Attacks Survey: Types, Vectors, and Technical Approaches," *Futur. Internet*, vol. 12, no. 10, p. 168, Sep. 2020, doi: 10.3390/fi12100168.

[23] J. Lee, Y. Lee, D. Lee, H. Kwon, and D. Shin, "Classification of Attack Types and Analysis of Attack Methods for Profiling Phishing Mail Attack Groups," *IEEE Access*, vol. 9, pp. 80866–80872, 2021, doi: 10.1109/ACCESS.2021.3084897.

[24] L. Tang and Q. H. Mahmoud, "A Survey of Machine Learning-Based Solutions for Phishing Website Detection," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 3, pp. 672–694, Aug. 2021, doi: 10.3390/make3030034.

[25] G. Varshney, R. Kumawat, V. Varadharajan, U. Tupakula, and C. Gupta, "Anti-phishing: A comprehensive perspective," *Expert Syst. Appl.*, vol. 238, p. 122199, Mar. 2024, doi: 10.1016/j.eswa.2023.122199.

[26] M. Khder, "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application," *Int. J. Adv. Soft Comput. its Appl.*, vol. 13, no. 3, pp. 145–168, Dec. 2021, doi: 10.15849/IJASCA.211128.11.

[27] M. Caliendo, M. Clement, D. Papies, and S. Scheel-Kopeinig, "Research Note —The Cost Impact of Spam Filters: Measuring the Effect of Information System Technologies in Organizations," *Inf. Syst. Res.*, vol. 23, no. 3-part-2, pp. 1068–1080, Sep. 2012, doi: 10.1287/isre.1110.0396.

[28] M. Shah and A. V. Hazarika, "An In-Depth Analysis of Modern Caching Strategies in Distributed Systems: Implementation Patterns and Performance Implications," *Int. J. Sci. Eng. Appl.*, vol. 14, no. 1, pp. 9–13, 2025.

[29] K. D. Tandale and S. N. Pawar, "Different Types of Phishing Attacks and Detection Techniques: A Review," in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, IEEE, Oct. 2020, pp. 295–299. doi: 10.1109/ICSIDEMPC49020.2020.9299624.

[30] A. Gogineni, "Edge-Aware DNS Routing in Kubernetes for Multi-Region Deployments," *Int. J. Innov. Res. Creat. Technol.*, vol. 9, no. 3, 2023.

[31] A. Gogineni, "Artificial Intelligence-Driven Fault Tolerance Mechanisms for Distributed Systems Using Deep Learning Model," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 1, no. 4, 2023.

[32] D. D. Rao, "Multimedia Based Intelligent Content Networking for Future Internet," in *2009 Third UKSim European Symposium on Computer Modeling and Simulation*, IEEE, 2009, pp. 55–59. doi: 10.1109/EMS.2009.108.

[33] M. I. Khan, A. Arif, and A. R. A. Khan, "AI's Revolutionary Role in Cyber Defense and Social Engineering," *Int. J. Multidiscip. Sci. Arts*, vol. 3, no. 3, pp. 57–66, 2024.

[34] M. S. S. Lingolu and M. K. Dobbala, "A Review on Data Security and Privacy in Serverless Computing: Key Strategies, Emerging Challenges," *J. Artif. Intell. Cloud Comput.*, 2024.

[35] S. A. Jawaid, "Artificial Intelligence with Respect to Cyber Security," *J. Adv. Artif. Intell.*, vol. 1, no. 2, pp. 96–102, 2023, doi: 10.18178/JAAI.2023.1.2.96-102.

[36] S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, Dec. 2022, pp. 260–264. doi: 10.1109/IC3I56241.2022.10073077.

[37] T. K. Kota and S. Rongala, "Implementing AI-Driven Secure Cloud Data Pipelines in Azure with Databricks," *Nanotechnol. Perceptions*, vol. 20, no. S15, 2024, doi: 10.62441/nano-ntp.vi.4439.

[38] R. P. Sola, N. Malali, and P. Madugula, *Cloud Database Security: Integrating Deep Learning and Machine Learning for Threat Detection and Prevention*. Notion Press, 2025.

[39] N. Prajapati, "The Role of Machine Learning in Big Data Analytics : Tools , Techniques , and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, pp. 16–22, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[40] V. Kolluri, "Revolutionary Research on the AI Sentry: An Approach to Overcome Social Engineering Attacks Using Machine Intelligence," *IJCRT - Int. J. Creat. Res. Thoughts (IJCRT), ISSN 2320-2882*, vol. 7, no. 3, pp. 2320–2882, 2024.

[41] A. Zamir *et al.*, "Phishing web site detection using diverse machine learning algorithms," *Electron. Libr.*, 2020, doi: 10.1108/EL-05-2019-0118.

[42] S. Nokhwal, P. Chilakalapudi, P. Donekal, S. Nokhwal, S. Pahune,

and A. Chaudhary, "Accelerating Neural Network Training: A Brief Review," in *2024 8th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI)*, New York, NY, USA: ACM, Apr. 2024, pp. 31–35. doi: 10.1145/3665065.3665071.

[43] K. S. Adewole, A. G. Akintola, S. A. Salihu, N. Faruk, and R. G. Jimoh, "Hybrid Rule-Based Model for Phishing URLs Detection," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 2019. doi: 10.1007/978-3-030-23943-5_9.

[44] S. Atawneh and H. Aljehani, "Phishing Email Detection Model Using Deep Learning," *Electron.*, 2023, doi: 10.3390/electronics12204261.

[45] G. Vrbančič, I. Fister, and V. Podgorelec, "Swarm intelligence approaches for parameter setting of deep learning neural network: Case study on phishing websites classification," in *ACM International Conference Proceeding Series*, 2018. doi: 10.1145/3227609.3227655.

[46] S. Murri, M. Bhoyar, G. P. Selvarajan, and M. Malaga, "Transforming Decision-Making with Big Data Analytics: Advanced Approaches to Real-Time Insights, Predictive Modeling, and Scalable Data Integration," *Int. J. Commun. Networks Inf. Secur.*, vol. 16, no. 5, pp. 506–519., 2024.

[47] J. Kumar Chaudhary, S. Tyagi, H. Prapan Sharma, S. Vaseem Akram, D. R. Sisodia, and D. Kapila, "Machine Learning Model-Based Financial Market Sentiment Prediction and Application," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, May 2023, pp. 1456–1459. doi: 10.1109/ICACITE57410.2023.10183344.

[48] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3557–3564, 2025.

[49] A. Balasubramanian, "Improving Air Quality Prediction Using Gradient Boosting," *Int. J. Sci. Technol.*, vol. 13, no. 2, pp. 1–9, 2022.

[50] N. Sharma, Anju, and A. Juneja, "Extreme gradient boosting with squared logistic loss function," in *Advances in Intelligent Systems and Computing*, 2019. doi: 10.1007/978-981-13-0923-6_27.

[51] E. Gandotra and D. Gupta, "Improving Spoofed Website Detection Using Machine Learning," *Cybern. Syst.*, 2021, doi: 10.1080/01969722.2020.1826659.

[52] K. T. Chu, H. T. Hsu, J. J. Sheu, W. P. Yang, and C. C. Lee, "Effective spam filter based on a hybrid method of header checking and content parsing," *IET Networks*, 2020, doi: 10.1049/iet-net.2019.0191.

[53] V. E. Adeyemo, A. O. Balogun, H. A. Mojeed, N. O. Akande, and K. S. Adewole, "Ensemble-Based Logistic Model Trees for Website Phishing Detection," in *Communications in Computer and Information Science*, 2021. doi: 10.1007/978-981-33-6835-4_41.

[54] Z. Wang, X. Sun, X. Li, and D. Zhang, "An Efficient SVM-Based Spam Filtering Algorithm," in *2006 International Conference on Machine Learning and Cybernetics*, IEEE, 2006, pp. 3682–3686. doi: 10.1109/ICMLC.2006.258626.

[55] Y. Han and Y. Shen, "Accurate spear phishing campaign attribution and early detection," in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, New York, NY, USA: ACM, Apr. 2016, pp. 2079–2086. doi: 10.1145/2851613.2851801.

[56] L. M. Form, K. L. Chiew, S. N. Sze, and W. K. Tiong, "Phishing email detection technique by using hybrid features," in *2015 9th International Conference on IT in Asia: Transforming Big Data into Knowledge, CITA 2015 - Proceedings*, 2015. doi: 10.1109/CITA.2015.7349818.

[57] S. Pandya, "Comparative Analysis of Large Language Models and Traditional Methods for Sentiment Analysis of Tweets Dataset," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 12, pp. 1647–1657, 2024, doi: 10.5281/zenodo.14575886.

[58] P. Choudhary, R. Choudhary, and S. Garaga, "Enhancing Training by Incorporating ChatGPT in Learning Modules: An Exploration of Benefits, Challenges, and Best Practices," *Int. J. Innov. Sci. Res. Technol.*, vol. 9, no. 11, 2024.

[59] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," 2021. doi: 10.1007/s11235-020-00733-2.

[60] S. Mathur and S. Gupta, "Classification and Detection of Automated Facial Mask to COVID-19 based on Deep CNN Model," in *2023 IEEE 7th Conference on Information and Communication Technology (CICT)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CICT59886.2023.10455699.

[61] M. K. Prabakaran, P. Meenakshi Sundaram, and A. D. Chandrasekar, "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Inf. Secur.*, vol. 17, no. 3, pp. 423–440, May 2023, doi: 10.1049/ise2.12106.

[62] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. P. Niyigena, "An effective phishing detection model based on character level convolutional neural network from URL," *Electron.*, 2020, doi: 10.3390/electronics9091514.

[63] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. A. Alzahrani, "A Deep Learning-Based Innovative Technique for Phishing Detection in Modern Security with Uniform Resource Locators," *Sensors*, vol. 23, no. 9, 2023, doi: 10.3390/s23094403.

[64] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Intelligent phishing detection scheme using deep learning algorithms," *J. Enterp. Inf. Manag.*, 2023, doi: 10.1108/JEIM-01-2020-0036.

[65] W. Wang, F. Zhang, X. Luo, and S. Zhang, "PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks," *Secur. Commun. Networks*, vol. 2019, 2019, doi: 10.1155/2019/2595794.

[66] W. Ali and A. A. Ahmed, "Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting," *IET Inf. Secur.*, 2019, doi: 10.1049/iet-ifs.2019.0006.

[67] A. Alhuzali, A. Alloqmani, M. Aljabri, and F. Alharbi, "In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets," *Appl. Sci.*, vol. 15, no. 6, pp. 1–30, 2025, doi: 10.3390/app15063396.

[68] P. Zhao and S. Jin, "Fewshing: A Few-Shot Learning Approach to Phishing Email Detection," in *2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI)*, IEEE, Jun. 2024, pp. 371–375. doi: 10.1109/SEAI62072.2024.10674290.

[69] Q. Qi, Z. Wang, Y. Xu, Y. Fang, and C. Wang, "Enhancing Phishing Email Detection through Ensemble Learning and Undersampling," *Appl. Sci.*, 2023, doi: 10.3390/app13158756.

[70] S. P. Ripa, F. Islam, and M. Arifuzzaman, "The emergence threat of phishing attack and the detection techniques using machine learning models," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021*, 2021. doi: 10.1109/ACMI53878.2021.9528204.

[71] X. Li, D. Zhang, and B. Wu, "Detection method of phishing email based on persuasion principle," in *Proceedings of 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2020*, 2020. doi: 10.1109/ITNEC48623.2020.9084766.

[72] P. K. K. Loh, A. Z. Y. Lee, and V. Balachandran, "Towards a Hybrid Security Framework for Phishing Awareness Education and Defense," *Futur. Internet*, vol. 16, no. 3, 2024, doi: 10.3390/fi16030086.