

A Review of Phishing Detection Techniques Using Machine Learning and Natural Language Processing

Dr. Sunita Dixit

Professor

Department of Computer Science & Engineering in St. Andrews Institute of Technology & Management

Gurgaon

bhardwajsunita23@gmail.com

Corresponding Author: *Sunita Dixit (bhardwajsunita23@gmail.com)

Received: 10 September, 2026 | Revised: 15 September, 2026 | Accepted: 28 September, 2026 |

Abstract—The goal of phishing attacks, which rank high among cyber risks, is to steal sensitive information (such as login passwords, financial data, and more) from unsuspecting people and companies. Blacklist, whitelist, rule, heuristic, and signature-based detection methods are no longer suitable due to the rising sophistication of phishing schemes. As a result, ML, DL, and NLP are rapidly becoming into effective instruments for detecting both well-known and brand-new forms of phishing. Background on phishing assaults, old ways of detecting them, and recent advances in intelligent phishing detection are all included in this article. It delves into several machine learning algorithms such as decision tree, LSTM, ANN, CNN, and RNN, as well as deep learning models like KNN, NB, and SVM. Also, natural language processing methods including word embedding, sentiment analysis, text categorization, and tokenization are explored with the aim of detecting phishing material. Presenting a comprehensive, precise, and up-to-the-minute phishing detection system is the ultimate goal of this study, which focuses on machine learning frameworks, performance indicators, present obstacles, and future research directions.

Keywords—Phishing Attacks, Machine Learning, Deep Learning, Natural Language Processing, URL Classification.

I. INTRODUCTION

A high degree of security is required even for the most cutting-edge technologies, such as cloud computing, mobile computing, e-commerce, and online banking. It is now imperative that these technologies be secure since they store sensitive personal information [1]. It is crucial for the security and economic well-being of every nation to strengthen cyber defenses and safeguard vital information infrastructures. New service development and government legislation are increasingly focused on making the Internet safer (and safeguarding Internet users). Cybercrime prevention efforts must be more thorough and secure. It is crucial that law enforcement authorities are granted the authority to investigate and punish cybercrime efficiently, as technological safeguards cannot prevent all crimes on their own. To stop the leak of sensitive data, some countries' governments are passing stringent cybersecurity regulations.

To protect themselves from these ever-increasing cybercrimes, everyone should get cybersecurity training [2].

Email phishing refers to the practice of pretending to be a legitimate business or organization in order to trick recipients into divulging sensitive information (such as login credentials or financial data). Scammers frequently employ communications that seem to originate from well-known websites in order to trick the unwary. These websites may be social media, auction houses, online payment processors, or even IT administrators. Links in phishing emails might lead to malicious websites [3]. One form of social engineering is phishing. In email phishing, the hacker sends a link via email requesting sensitive information, such as a user's bank details or personal information. The user then follows the link and enters all of their details, allowing the hacker to obtain all of their data. Phishing works in this way. The first step in a phishing assault is sending an email or other form of contact with the intent to trick the target. It appears as though the communication is being sent via a reliable source. The victim may end up giving their personal information to a spam website if it manages to trick them [4]. Malware is sometimes downloaded into the PC of the victim as well [5].

ML has been employed to combat some of the most significant cyberthreats. ML is rapidly revolutionizing every industry, and it is one of the most recent innovations in the informatics area. Since ML was introduced to the cybersecurity field, a lot of work has been mechanized in an effort to find automated solutions to various questions. Also, compared to what's already available on the market, several of these options are cheaper [6]. Security automation isn't going to solve all of cybersecurity's problems. Nevertheless, several advances in natural language processing (NLP) within ML have occurred in recent years. In order to assist computers, understand human languages, this field focuses on text analysis and processing, recognize the potential of NLP to learn how to distinguish phishing texts from legitimate ones, which could have far-reaching implications for cybersecurity [7][8].

The five parts of this article are as follows: Section II lays out the groundwork for a phishing assault. Chapter III:

Detecting Phishing with Machine Learning and Natural Language Processing. After a brief overview of relevant literature in Section IV, the article draws to a close in Section V, which also suggests avenues for further study.

II. PHISHING ATTACKS DETECTION TECHNIQUES

Phishing attacks, a type of cybersecurity threat related to network attacks, pose as reputable companies or people to coerce victims into disclosing personal information[9][10]. This technique entails contacting a large number of people at random with phony emails, phone calls, or SMS messages asking them to divulge their private information. They pretend to be from a reputable organization or entity in order to defraud individuals of their password and other personal information. From a direct message to a fake website, whatever share ends up in a scammer's pocket and may be used.

A. Types of Phishing Attacks

There are two types of phishing attacks depending on the communication method and the cybercriminal attack method. Usual examples are spear phishing, whaling and vishing (voice phishing). Each kind aims to deceive the user into disclosing personal information, such as login credentials, financial information, or personal details, by posing as a reliable person or organization. The increasingly sophisticated nature of phishing attacks calls for sophisticated methods of detection and user awareness, otherwise the risk of successful attacks increases.

1) Deceptive Phishing

This is the most prevalent kind of phishing, when the attackers pretend to be a reputable business in an attempt to get people's login credentials or personal information. The users are subsequently coerced into complying with the hacker's demands.

2) Spear Phishing

An intrusion detection and prevention system that is based on wireless technology examines wireless protocol behaviors in order to identify potential threats and implements countermeasures. It can identify if someone is using an illegal wireless local area network [5]. Application, transport, and protocol layer suspicious behavior cannot be detected.

3) Clone Phishing

This kind of phishing attempt involves using an already-existing email with the same or similar attachments and links, as well as the recipient's address book, to craft a new email. A malicious version of the attachment or link inside the email is delivered to the victim using an email address that is spoofing to make it look to have originated from the original sender.

4) Whaling

This is one kind of phishing; in it, the perpetrator targets the victim's high social standing in an effort to get access to sensitive information through various online platforms, such as social media accounts.

5) Link Manipulation

Phishing techniques include "Link Manipulation," in which the perpetrator spoofs or otherwise compromises a target's website by sending a link to it. In reality, rather of directing the visitor to the advertised website, clicking on that link actually accesses the phisher's website. Voice PhishingThe term "voice phishing" describes a specific type of phone criminal assault in which the perpetrator uses social

engineering techniques in conjunction with the telephone network to get access to victims' sensitive financial and personal data.

B. Phishing attack Lifecycle

The phishing attack lifecycle is a series of steps that attackers take to target individuals, create a phishing kit, send a scam message, deceive the user into clicking on a link, and remove evidence. Recognizing these stages helps in detecting, preventing, and mitigating phishing attacks effectively. A typical phishing attempt goes through several stages, as illustrated in Fig. 1 below [11].

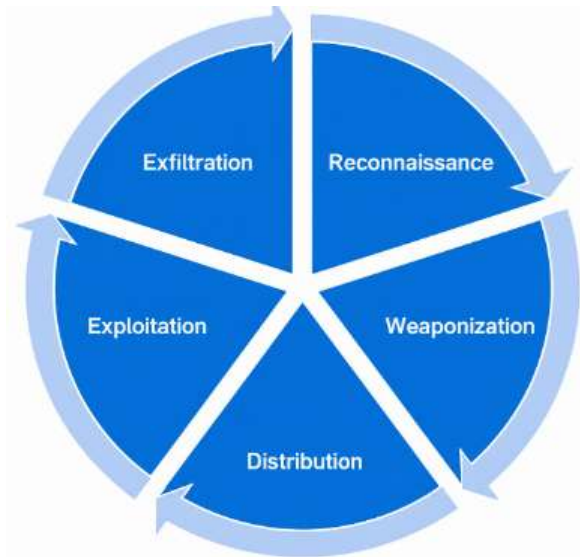


Fig. 1. Phishing Attack Lifecycle

- **Reconnaissance:** Reconnaissance, also referred to as the planning phase, includes the selection of the attack method, definition of attack goals, and collection of information of potential attack targets [12].
- **Weaponization:** The weaponization or preparatory phase is when an attacker finds potential weak spots in order to launch an attack by sending phishing emails to unsuspecting victims [13]. Either human hands or automated tools can do this step.
- **Distribution:** This phase involves the phishers sending out the bait, including phishing lures, botnets, or malware, to the victims and waiting for them to respond [14].
- **Exploitation:** The exploitation/penetration phase is largely about playing with the psychology of people. The malicious actions, such as data collection or illegal access to the target device, are subsequently carried out by the phishers [15].
- **Exfiltration:** The final stage of a phishing attack is for the perpetrator to attempt erasing duplicate websites and rerouting victims to the legitimate one [16].

C. The Evolution of Phishing Attack

Phishing attacks have matured from the basic mass email scams to sophisticated and targeted cyber-attacks. These days, attackers employ social engineering, AI and custom information to create realistic phishing attempts in emails, SMS, social media and phone calls. This rapid evolution has made phishing more difficult to detect, emphasizing the need

for advanced detection techniques and continuous user awareness.

- **Early Phishing Attacks (1995-2005)** The first ten years of Phishing (1995-2005) were relatively simple in their approach. The initial attacks used email as the primary weapon, and the attackers' strategy involved sending mass spam to the effect that they hope a few recipients fall for it. This was an early stage which gave rise to a new type of threat: phishing, which would become a major cybersecurity problem.
- **Model Training on Labeled Threat Data:** Supervised learning models learn from historical data by
- **Evolution of Sophisticated Techniques (2005-2015):** Phishing techniques began to become more sophisticated during the second decade, as technology matured and more people began to use the Internet [17]. This was a pivotal shift from being solely a mass mailing environment to a more sophisticated and targeted environment, altering the threat landscape in a fundamental way.
- **Modern Phishing Landscape (2015-Present):** These are the most modern times of phishing, when highly sophisticated attacks use advanced technologies and social engineering techniques. Phishing has evolved into a complex assault that utilizes a combination of psychology and technology, making it more difficult to identify and stop.

D. Traditional Phishing Detection Techniques

Phishing detection methods are based on rules, blacklists, whitelists, signatures and heuristic analysis of phishing web sites and e-mails. These methods are effective for detecting known phishing attacks and require relatively low computational resources. However, they have limited capability in detecting new or evolving phishing attacks, making them less effective against zero-day threats.

1) List-Based Detection Techniques

In order to shield the user from an online phishing scam, list-based detection evaluates the URL's validity using lists. According to this method, a list is kept as a database of URLs. There are often URLs, IP addresses, and keywords stored there. A whitelist, consisting of links to valid websites, is kept by certain researchers.

2) Heuristic Rule-Based Detection Techniques

The suspicious URL is subject to thumb rules used by heuristic rule-based detection systems in order to ensure security. Most of the time, there are such few changes between real and phishing site addresses that users are unable to tell them apart [18]. Criminals put a lot of effort into creating fake websites so that people won't believe they're real.

3) Signature-Based Detection Techniques

An intrusion detection system's signature-based detection mechanism looks for suspicious behavior by comparing incoming network traffic to a database of previously identified attack signatures [19]. Each signature corresponds to a specific type of threat, capturing its unique characteristics.

III. MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING BASED PHISHING DETECTION

ML is a subfield of AI concerned with developing systems and algorithms that can autonomously learn from data,

generate predictions, and make judgments [20]. Large datasets are used to train ML algorithms [21]. used for phishing identification, it incorporates characteristics from both popular phishing sites and legal ones. In order to accurately categorize questionable websites, these algorithms learn to spot patterns, outliers, and telltale indicators of phishing attempts.

1) Supervised Learning

In order to train a classification model that can assign each input data point to one of a predetermined number of classes, supervised ML algorithms need massive labeled datasets as resources [22]. A ML pipeline for training supervised models to identify phishing attempts. As there are just two possible outcomes in a phishing webpage detection problem—phishing or benign—the trained models are binary classifiers.

2) Unsupervised Learning

Unsupervised learning is used when data is only provided as an input and there is no corresponding output variable. Such algorithms model the underlying patterns in the data to learn more about its properties.

3) Semi-supervised Learning

This is a cross between supervised and unsupervised learning methods [23]. These algorithms are trained on a combination of labeled and unlabeled data. In a typical context, there is a small amount of labeled data and a big number of unlabeled data.

4) Reinforcement Learning

RL is used when the objective at hand is to make a series of decisions leading to a final reward. During the learning process, an artificial agent is rewarded or penalized for acts it takes. The aim is to maximize the overall prize. Examples include teaching agents to play computer games or doing robotics tasks with an end objective.

B. Emerging Machine Learning Algorithms for Phishing Detection

Emerging ML algorithms have significantly improved phishing detection by learning complex patterns from URLs, emails, and website content. Ensemble learning, GNNs, transformers, and XAI are among the advanced techniques that deliver greater accuracy in phishing detection, adaptability to new phishing tactics, and a lower false positive rate. The algorithms facilitate the identification of sophisticated and never-observed phishing attacks in real time.

1) Logistic Regression (LR)

LR is one of the most used ML algorithms for binary classification. This is because it is a basic calculation that performs very well on a wide scope of issues. The logistic regression model takes genuine, esteemed sources of information and makes an estimate of the likelihood that the information belongs to the default class.

2) K-Nearest Neighbor (KNN)

Predictions for another event (x) are generated using the whole training set for the K most relevant models (neighbors), and the output variable for those K events is summarized. This might be the mean output variable in regression, or it could be the median class value in characterization. A distance metric is used to determine which of the K examples in the test dataset are the most similar to each other.

3) *Random Forest (RF) Classifier*

RF is a supervised ML technique effective for classification and regression [24]. A RF algorithm creates DTs from the training dataset provided and then predicts each output from each decision tree and selected the tree with the best solution by voting and hence is a very efficient and useful algorithm with a wide range of applications.

4) *Support Vector Machine (SVM)*

The main aim of the SVM is to decide on the best line or choice cutoff that can separate n-dimensional space into classes so can put the new data point in the correct classification later on. This aids simpler prediction for new data. This best line cutoff is known as a hyperplane.

C. *Phishing Detection Framework Using Machine Learning*

A framework for detecting phishing using ML usually includes data collection, feature extraction, model training, classification, and performance evaluation. The framework analyzes features from URLs, emails, or web pages to distinguish phishing attempts from legitimate ones, enabling accurate and real-time threat detection with continuous model improvement.

1) *URL-Based Phishing Detection*

The development of a ML-based approach to identify phishing and malware URLs is one example. The suggested classifier [25] is trained using data that the user has themselves acquired. Virus Total URLs included three basic categories of features: domain, hosting, and lexical. Because phishing URLs often use lexical qualities that imitate well-known companies, their research shows that these features are crucial to the task at hand. In contrast, malware URLs seldom use such features.

2) *Email-Based Phishing Detection*

The objective of sending phishing emails is to obtain sensitive information from the receivers. Most people fell prey to phishing assaults because they were negligent when surfing the internet. Some spam filters use a plethora of characteristics to identify and block fraudulent emails.

3) *Threat Intelligence Assisted Phishing Detection*

The policy rules are designed to replicate SOC operations by using a simulated threat intelligence feed similar to Phish Tank, which allows suspicious URLs to be verified against live indications. This dual-layer validation enhances zero-day resilience and enables future integration with SIEM or SOAR systems.

4) *Feature Engineering Phishing Detection*

Feature engineering yielded a complete collection of URL-level characteristics, including domain structure, subdomain count, length, special character frequency, IP address presence, HTTPS flag, and suspicious keywords [26]. Following cleaning and normalization, the dataset was split into training (80%) and testing (20%) subsets to retain the class distribution.

5) *Ensemble Learning Phishing Detection*

The ensemble layer combines the basis models using hard and soft voting procedures, improving resilience to feature variation and dataset imbalance.

D. *Deep Learning Phishing Detection*

DL is an algorithm or subset of ML. It mostly applies to deep ANNs and deep RL, both of which utilize the term "deep" to denote the number of layers in the system [27]. There are numerous ways (such as CNNs, RNNs, LSTMs, and auto-encoders) to apply DL, and these approaches rely on the use cases, such as the sort of data that has to be processed.

1) *Convolutional Neural Networks:*

The modified URL matrix is used to create and train an enhanced CNN network using the character embedding approach [28]. The URL training set is used to train the CNN network, and the back-propagation approach updates various model parameters on a continuous basis. Following the model training, multilayer URL features are retrieved from the CNN network.

2) *Recurrent Neural Networks*

The RNN training for the phishing detection model employs a structured algorithmic method [29]. The method begins with initializing the RNN parameters, which include the network's layer weights and biases. These settings are assigned to random values to disrupt the symmetry and assist learning.

3) *Long Short-Term Memory*

In DL deep analysis, the LSTM approach stands out from the competition thanks to its advanced features that allow it to identify unexpected assaults [30]. Problems with anomaly detection are common on the deep web. The LSTM's long-term memory is able to identify network-based Tor phishing and website fingerprinting assaults by utilizing several cell topologies, including forget, input, and cell state.

E. *Natural Language Processing in Phishing Detection*

Phishing detection based on NLP combines the linguistic analysis and the ML to detect bad messages, by relying on the content and structure of the message rather than on a signature-based approach [31][32]. The NLP models can detect small indications of phishing in the email body, subject line, URL and sender. Signs of this may be tone changes, urgency markers, grammatical mistakes, or odd word orders. This technique helps reduce the chances of missing a threat by understanding the intent and significance of communication.

F. *NLP Techniques in Phishing Detection*

NLP techniques are able to identify text-based patterns of email and message communications and websites that correspond to phishing activity. These techniques include tokenization, word embeddings, TF-IDF, and transformer-based models like BERT, which are effective for identifying deceptive language, malicious intent, and social engineering techniques. Table I compares the working principles of phishing detection ML algorithms, their advantages, disadvantages and some common applications to find malicious URLs, websites and phishing emails.

1) *Tokenization*

Breaking text into smaller pieces (words, phrases, sentences). Tokenization allows a machine to analyze text more effectively by breaking it down into digestible components for further analysis.

2) *Part-of-Speech Tagging (POS)*

This is the method used to determine the grammatical class of words (noun, verb, adjective etc.). By analyzing how different parts of speech are used in a text, NLP algorithms

can detect irregular patterns that may indicate phishing attempts, such as unusual word choices or the overuse of certain phrases.

3) *Named Entity Recognition (NER)*

NER finds and sorts things like names, dates, places, and organizations in the text. In phishing emails, attackers may use fake names or misrepresent legitimate entities. NLP can help detect discrepancies by recognizing and verifying these named entities.

4) *Sentiment Analysis*

This technique evaluates the subjective tone of the text. Emails designed to deceive recipients into making a quick decision often include emotionally charged wording [33]. Sentiment analysis can help detect suspicious emotional appeals, such as urgency, fear or excitement, which are common in phishing attempts.

5) *Word Embeddings*

Word embeddings, such as Word2Vec or Glove, represent words in a high dimensional vector space, capturing semantic relationships between words. These embeddings allow NLP models to understand the context and meaning behind words, helping to identify suspicious language or patterns in phishing emails that might otherwise go unnoticed.

TABLE I. COMPARISON OF MACHINE LEARNING ALGORITHM FOR PHISHING DETECTION

Machine Learning Algorithm	Working Principle	Advantages	Limitations	Common Applications in Phishing Detection
Random Forest (RF)	Combines multiple decision trees and uses majority voting for classification.	High accuracy, robust to overfitting, handles large feature sets.	Higher computational cost and lower interpretability.	URL classification, email phishing detection, website analysis.
Support Vector Machine (SVM)	Finds the optimal hyperplane that separates phishing and legitimate samples.	Effective in high-dimensional data, good generalization.	Slow for very large datasets and requires careful parameter tuning.	Phishing website detection, malicious URL classification.
K-Nearest Neighbors (KNN)	Classifies samples based on the majority class of the nearest neighbours.	Simple implementation, no training phase required.	High prediction time and sensitive to irrelevant features.	URL classification, phishing website detection.
Logistic Regression (LR)	Estimates the probability of phishing using a logistic function.	Simple, interpretable, efficient for binary classification.	Limited in capturing complex non-linear relationships.	Binary phishing detection, risk prediction.

IV. LITERATURE REVIEW

Research pertaining to ML, DL, and NLP-based phishing detection is summarized here. It focuses on the main methods and how they helped improve the accuracy of phishing detection.

T. Ngakpal, A. Malik, and I. Batra (2026) discusses the use of artificial intelligence and ML to identify and thwart social engineering assaults in response to these concerns. Using AI and ML models provide the ability to recognize intelligent behavior, detect anomalies, and quickly change due to new threats, all of which greatly improve accuracy in detection and efficiency in response. The research also highlights the need for the use of these intelligent security frameworks in healthcare, where hacker attacks to obtain PHI present great risks to vulnerable populations (i.e., patients who are on the Autism spectrum and other vulnerable patients) [34].

R. Arora and N. Kaur (2025) These attacks pose a huge challenge to social engineering due to the proliferation of online platforms and the activities that are integral to everyday life. Financial gain, identity concealment, fame, and notoriety are the principal goals of an attacker who launches a phishing assault. There have been several advancements in phishing detection and prevention methods in the past few years. Website URLs, HTML structures, and page layouts—from real and false websites alike—are the primary targets of the strategies. This SLR takes a look at a variety of phishing detection methods, including those that rely on whitelists and blacklists, visual site similarity, heuristic analysis, and sophisticated approaches in the form of ML and DL models [35].

M. S, S. Haniah et al. (2024) explores several methods that may be used to employ artificial intelligence, including ML and NLP, to aid in the real-time detection and mitigation of phishing assaults. Take a look at the latest research on DL-based phishing detection methods, cyber security awareness training, AI-enabled detection mechanisms, and phishing categorization strategies. To improve cyber defenses against new phishing attacks on online trading systems, it is necessary to conduct a thorough literature review, investigate possible issues with data availability, model interpretation, and scalability, and propose ways to proceed with future research. Improving user education and awareness is also important in the fight against phishing [36].

K. Sharma, P. Rai, and J. Chandel (2023) ML is an AI system that can identify malicious websites that attempt to trick users into giving their personal information. To train ML systems to distinguish between legitimate and fraudulent websites, large datasets of both types of websites may be used. These algorithms may be trained to instantly identify phishing and legitimate websites. This study discusses the current state of the art in identifying phishing websites using ML. This study makes use of many ML techniques, each of which is detailed in the article along with the attributes that were utilized to train these algorithms. In addition to discussing the potential of this study, the essay delves into the constraints of real-time phishing website recognition [37].

S. Salloum et al. (2022) investigate important topics in NLP for the purpose of detecting phishing emails, ML techniques for this purpose, text characteristics of these emails, datasets and resources utilized for this purpose, and assessment standards. Included in the results are techniques for categorizing and improving the detection of phishing

emails, with feature extraction and selection following as the primary study focus in phishing detection studies. Phishing email detection makes extensive use of SVMs, one of several categorization techniques. Word embeddings and TF-IDF are the two most popular natural language processing methods. The Nazario phishing corpus is also one of the most used datasets for testing different approaches to detecting phishing emails [38].

A. Villanueva et al. (2022) processed URLs using ML and DL models to identify phishing attempts. For the purpose of classifying URLs as either valid or phishing, ML algorithms like LR and Multi-Naive Bayes were employed. In addition,

DL models such as Long Short-Term Memory, Gated Recurrent Units, and Bidirectional Recurrent Neural Networks were utilized. The Long Term Short Memory and Gated Recurrent Units models, which are two of the models employed, have very good scores in training and validation, and they achieve an overall accuracy of 95%. The accuracy rate of the Bidirectional Recurrent Neural Net employing LSTM and Gated Recurrent Units is 97% [39].

Table II presents a comparative analysis of phishing detection studies, highlighting their methodologies, key findings, limitations, and future directions across ML and NLP approaches.

TABLE II. COMPARATIVE ANALYSIS OF PHISHING DETECTION TECHNIQUES USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

Reference	Study Focus	Approach / Methodology	Key Findings	Challenges / Limitations	Future Directions
T. Nagpal, A. Malik, and I. Batra, (2026)	AI and ML for social engineering & phishing detection	AI- and ML-based anomaly detection and intelligent security framework	Improves phishing detection accuracy, anomaly detection, and healthcare cybersecurity resilience	Adapting to rapidly evolving phishing attacks; deployment complexity	Develop adaptive AI-based security frameworks and enhance protection of healthcare systems
R. Arora and N. Kaur, (2025)	Systematic review of phishing detection techniques	Review of blacklist, whitelist, heuristic, visual similarity, ML and DL approaches	Summarizes conventional and intelligent phishing detection methods	Traditional methods are ineffective against sophisticated phishing attacks	Develop hybrid ML/DL models with improved detection accuracy
M. S, S. Haniah et al. (2024)	AI-based phishing detection for online trading	Machine Learning, NLP, Deep Learning, and cybersecurity awareness	Enhances real-time phishing detection and user awareness	Data availability, model interpretability, and scalability	Improve explainable AI, scalable models, and collaboration among stakeholders
K. Sharma, P. Rai, and J. Chandel, (2023)	Machine learning-based phishing website detection	ML algorithms trained on phishing and legitimate websites	Effectively classifies phishing websites in real time	Limited real-time deployment and feature generalization	Develop robust real-time phishing detection systems with advanced feature engineering
S. Salloom et al. (2022)	NLP-based phishing email detection	Systematic review of NLP, TF-IDF, word embeddings, and ML classifiers	SVM, TF-IDF, and word embeddings achieve effective phishing email detection	Dependence on benchmark datasets and feature engineering	Explore transformer-based NLP models and multilingual phishing detection
A. Villanueva et al. (2022)	NLP-based phishing URL detection	Logistic Regression, Naïve Bayes, LSTM, GRU, and Bi-RNN models	Deep learning models achieved up to 97% accuracy in URL phishing detection	Uses only URL features; limited generalization	Integrate additional website features and advanced deep learning architectures to improve performance

V. CONCLUSION AND FUTURE WORK

Phishing attacks are still becoming more sophisticated, making them a threat to individuals, organizations and cybersecurity systems. This review considers the development of phishing attacks, the traditional phishing detection methods, and the latest developments in intelligent phishing detection. Blacklist-, whitelist-, heuristic-, rule- and signature-based solutions are still valuable but not very effective to identify emerging and zero-day phishing attacks. The classification accuracy of the ML algorithms such as LR, DT, RF, SVM, NB, and KNN has been found to be improved through learning discriminative patterns in phishing datasets. Furthermore, DL models, such as ANNs, CNNs, RNNs, and LSTM networks, provide enhanced capability for automatic feature extraction and detection of complex phishing behaviors. NLP techniques enhance phishing detection by leveraging textual analysis, semantic relationships within emails and web pages, and social engineering indicators.

The creation of real-time zero-day phishing attack detectors that are lightweight, explainable, and adaptable requires more research. Integrating transformer-based architectures, XAI, federated learning, and multimodal analysis can further improve detection accuracy, scalability,

privacy preservation, and robustness against rapidly evolving phishing techniques.

- **Funding:** This research received no external funding.
- **Institutional Review Board Statement:** Not applicable.
- **Informed Consent Statement:** Not applicable.
- **Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author upon reasonable request.
- **Conflicts of Interest:** The author declares no conflict of interest.

How to cite this article:
S. Dixit, “A Review of Phishing Detection Techniques Using Machine Learning and Natural Language Processing,” *Journal of Global Research in Multidisciplinary Studies (JGRMS)*, vol. 2, no. 9, pp. 15–21, Sep. 2026, Doi,XXX

REFERENCES

[1] D. Patel, “Integrating Price Elasticity and Reinforcement Learning: A Data-Driven Framework for Strategic E-commerce Pricing,” in *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, Houston, TX, USA: IEEE, 2026, pp. 1–6, February. doi: 10.1109/ICAIC67076.2026.11395747.

[2] N. R. Gade and U. Reddy, “A Study Of Cyber Security Challenges

- And Its Emerging Trends On Latest Technologies,” 2014.
- [3] S. Chatterjee, “Advanced Malware Detection in Operational Technology: Signature-Based Vs. Behaviour-Based Approaches,” *ESP J. Eng. Technol. Adv.*, vol. 1, 2021, doi: 10.56472/25832646/JETA-V1I2P128.
- [4] S. Pawar, G. Patil, K. Patel, P. Pawar, S. Khedkar, and B. More, “Falsified News Detection Using Deep Learning Approach,” in *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 2021, pp. 1–5, doi: 10.1109/ASIANCON51346.2021.9544585.
- [5] V. Bhavsar, A. Kadlak, and S. Sharma, “Study on Phishing Attacks,” *Int. J. Comput. Appl.*, vol. 182, pp. 27–29, 2018, doi: 10.5120/ijca2018918286.
- [6] M. R. C. Mukkolakkal, “1.5 Million Messages Per Second on 3 Machines: Benchmarking and Latency Optimization of Apache Pulsar at Enterprise Scale,” *Arxiv J.*, pp. 8–11, 2026, doi: 10.48550/arXiv.2603.29113.
- [7] R. Jonker, R. Poudel, T. Pedrosa, and R. Lopes, “Using Natural Language Processing for Phishing Detection,” in *Conference paper*, 2022, pp. 540–552, doi: 10.1007/978-3-030-91885-9_40.
- [8] C. P. Singh, “Enterprise-Grade AI-Driven Text Analytics and Insight Extraction Using Transformer-Based NLP Models,” in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, IEEE, Mar. 2026, pp. 1–6, doi: 10.1109/ISDFS69419.2026.11459059.
- [9] Sharjil Iqbal, “NETWORK ATTACK DETECTION BY GROUP SEARCH OPTIMIZATION USING CONVOLUTIONAL DEEP LEARNING MODEL,” *Int. J. Eng. Res. Sci. Technol.*, vol. 22, no. 1, pp. 579–585, February, Feb. 2026, doi: 10.62643/ijerst.2026.v22.n1.pp579-585.
- [10] K. Chetoui, B. Bah, A. Alami, and A. Bahnasse, “Overview of Social Engineering Attacks on Social Networks,” *Procedia Comput. Sci.*, pp. 656–661, 2022, doi: 10.1016/j.procs.2021.12.302.
- [11] E. Daoud, J. Garcia-Blas, S. Alawadi, and J. Carretero, “Phishing in the age of distributed intelligence: taxonomies, detection strategies, and the emerging role of federated learning,” *Prog. Artif. Intell.*, 2026, doi: 10.1007/s13748-026-00448-6.
- [12] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, “Phishing Attacks: A Recent Comprehensive Study and a New Anatomy,” *Front. Comput. Sci.*, vol. 3, no. March, pp. 1–23, 2021, doi: 10.3389/fcomp.2021.563060.
- [13] A. Aleroud and L. Zhou, “Phishing environments, techniques, and countermeasures: A survey,” *Comput. Secur.*, vol. 68, pp. 160–196, 2017, doi: 10.1016/j.cose.2017.04.006.
- [14] N. Goenka, R., Chawla, M., Tiwari, “A comprehensive survey of phishing: mediums, intended targets, attack and defence techniques and a novel taxonomy,” *Int. J. Inf. Secur.*, vol. 23, pp. 819–848, 2023.
- [15] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, “Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions,” *IEEE Access*, vol. 10, pp. 36429–36463, 2022, doi: 10.1109/ACCESS.2022.3151903.
- [16] R. Alabdan, “Phishing Attacks Survey: Types, Vectors, and Technical Approaches,” *Futur. Internet*, vol. 12, no. 10, p. 168, Sep. 2020, doi: 10.3390/fi12100168.
- [17] J. Osamor, M. Ashawa, A. Shahrabi, A. Philip, and C. Iwendu, “The Evolution of Phishing and Future Directions: A Review,” *Int. Conf. Cyber Warf. Secur.*, vol. 20, no. 1, pp. 361–368, Mar. 2025, doi: 10.34190/iccws.20.1.3366.
- [18] M. Vijayalakshmi, S. Mercy Shalinie, M. H. Yang, and R. M. U., “Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions,” *IET Networks*, vol. 9, no. 5, pp. 235–246, 2020, doi: 10.1049/iet-net.2020.0078.
- [19] W. Olabiyi, A. Chole, and Y. Themmah, “Signature-Based vs. Anomaly-Based Detection,” 2023.
- [20] V. Adeyemi Onih, “Phishing Detection Using Machine Learning: A Model Development and Integration,” *Int. J. Sci. Manag. Res.*, vol. 07, no. 04, pp. 27–63, 2024, doi: 10.37502/IJSMR.2024.7403.
- [21] V. Misal, V. P. Janeja, S. C. Pallaprolu, Y. Yesha, and R. Chintalapati, “Iterative unified clustering in big data,” in *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA: IEEE, Dec. 2016, pp. 3412–3421, doi: 10.1109/BigData.2016.7841002.
- [22] D. M. Divakaran and A. Oest, “Phishing Detection Leveraging Machine Learning and Deep Learning: A Review,” *IEEE Secur. Priv.*, vol. 20, no. 5, pp. 86–95, 2022, doi: 10.1109/MSEC.2022.3175225.
- [23] S. Sah, “Machine Learning: A Review of Learning Types,” *Neural Networks 152*, vol. 7, no. 1, pp. 267–275, 2020, doi: 10.20944/preprints202007.0230.v1.
- [24] P. Kadiyala, S. S. KV, B. S. Shashank, and K. A. Kumar, “Phishing Website Detection using Emerging Machine Learning Models,” *SSRN Electron. J.*, no. Icsmdi, 2021, doi: 10.2139/ssrn.3853016.
- [25] J. D. Duarte et al., “Machine Learning for Early Detection of Phishing URLs in Parked Domains: An Approach Applied to a Financial Institution,” *IEEE Access*, vol. 13, no. June, pp. 145736–145753, 2025, doi: 10.1109/ACCESS.2025.3599454.
- [26] A. Davis, S. Abdelsalam, M. Ghaleb, M. S. M. Gismalla, E. I. Eltahir, and M. Hamdan, *A Multi-Layer Phishing Defense Framework for Trusted Cloud Environments*, vol. 1, no. 1. Association for Computing Machinery, 2025, doi: 10.1145/3773276.3774878.
- [27] O. Sahingoz, E. Buber, and E. Kugu, “DEPHIDES: Deep Learning Based Phishing Detection System,” *IEEE Access*, vol. 12, pp. 8052–8070, 2024, doi: 10.1109/ACCESS.2024.3352629.
- [28] 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*, vol. 21, no. 24, pp. 1–18, 2021, doi: 10.3390/s21248281.
- [29] B. B. Gupta, A. Gaurav, R. W. Attar, V. Arya, A. Alhomoud, and K. T. Chui, “Optimized Phishing Detection with Recurrent Neural Network and Whale Optimizer Algorithm,” *Comput. Mater. Contin.*, vol. 80, no. 3, pp. 4895–4916, 2024, doi: 10.32604/cmc.2024.050815.
- [30] R. Chanemougasoundaram and T. M., “The LSTM-Based Automated Phishing Detection Driven Model for Detecting Multiple Attacks on Tor Hidden Services,” 2022, doi: 10.21203/rs.3.rs-2180307/v1.
- [31] A. Ibrahim Sani, O. Onimole, A. Oludele Adebayo, N. Adeniyi Akande, and U. J. Eziokwu, “Phishing Detection Using Natural Language Processing and Behavioural Analysis: A Multi-Faceted Approach,” *Int. J. Sci. Manag. Res.*, vol. 08, no. 11, pp. 57–71, 2025, doi: 10.37502/ijsmr.2025.81105.
- [32] S. Devarakonda, R. Lingam, and V. Challa, “Confidence-Gated RAG for Adaptive Retrieval in Sequential Agents,” in *ICLR 2026 Workshop on Logical Reasoning of Large Language Models*, ICLR, 2026, pp. 01–10, Apr.
- [33] D. Oluremi, “Using Natural Language Processing (NLP) to Detect Phishing Emails in Web Applications,” 2017.
- [34] T. Ngakpal, A. Malik, and I. Batra, “Social Engineering Attack Detection Models for Individuals with Autism: A Systematic Review,” in *2026 International Conference on Signal, Systems, and Computing for Next-Gen Automation (ICSSCNA)*, 2026, pp. 1503–1507, doi: 10.1109/ICSSCNA68616.2026.11546772.
- [35] R. Arora and N. Kaur, “Phishing Detection Techniques Countermeasures and Challenges a Review,” in *2025 IEEE 1st International Conference on Smart Innovations in Systems, Infrastructure, Mechanical, Power, AI and Computing Technologies (SISIMPACT)*, 2025, pp. 1280–1284, doi: 10.1109/SISIMPACT67725.2025.11439257.
- [36] M. S. S. Haniah, S. M. Koti, S. Sahani, and S. S. N., “A Review on Phishing Threats and Data Security in Online Trading Systems using Artificial Intelligence Techniques,” in *2024 Second International Conference on Advances in Information Technology (ICAIT)*, 2024, pp. 1–6, doi: 10.1109/ICAIT61638.2024.10690690.
- [37] K. Sharma, P. Rai, and J. Chandel, “Review Paper Real-Time Phishing Website With Machine Learning,” in *2023 11th International Conference on Intelligent Systems and Embedded Design (ISED)*, 2023, pp. 1–5, doi: 10.1109/ISED59382.2023.10444574.
- [38] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, “A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques,” *IEEE Access*, vol. 10, pp. 65703–65727, 2022, doi: 10.1109/ACCESS.2022.3183083.

-
- [39] A. Villanueva, C. Atibagos, J. De Guzman, J. C. Dela Cruz, M. Rosales, and R. Francisco, "Application of Natural Language Processing for Phishing Detection Using Machine and Deep Learning Models," in *2022 International Conference on ICT for Smart Society (ICISS)*, 2022, pp. 1–6. doi: 10.1109/ICISS55894.2022.9915037.