Int. J. Sci. R. Tech., 2025 2(11)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

A Review of Machine Learning Techniques for Phishing Detection

Teja Ravi Hulse*

Department of Computer Science and Engineering, SDM College of Engineering and Technology, Dharwad, India

ABSTRACT

Phishing is one of the major online threats today. It tricks users into revealing sensitive information such as personal data, passwords, and financial details. Traditional detection techniques are often unable to keep track of evolving attack patterns since phishing has become increasing rapidly. Machine Learning (ML) has become a strong tool for detecting phishing techniques. It analyses and learns from the pattern to get the results. This paper presents a review of various ML techniques applied to detect the threats. Considering algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, Naïve Bayes, and deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). The paper also contains datasets such as PhishTank and UCI Phishing Websites Dataset. It compares model performances and tracks the trends. In the future, researchers must focus on hybrid models, real-time detection mechanisms, and strong protection against new phishing techniques.

Keywords: Machine Learning, Phishing Detection, PhishTank, UCI dataset, Hybrid Deep Learning

INTRODUCTION

Phishing is a social engineering attack that tricks users into revealing confidential information such as passwords and financial details by pretending to be a trusted source [1]. These attacks typically occur through fake emails, websites or SMS messages that appear real to users. *Email phishing* is the most common type. Attackers send messages having malicious links. *Website phishing* creates duplicate pages that mimic genuine sites. *SMS phishing* (or smishing) uses text messages to deceive victims [2]. Detecting phishing is difficult due to the **constantly evolving nature of attack patterns**. Attackers often change URLs, website content, and hosting patterns to evade traditional rule-based or blacklist-based detection systems [3]. The large amount of data generated from online transactions also makes manual detection hard. This leads to automated and intelligent solutions. ML has appeared as a powerful approach for phishing detection. It learns from data patterns and classifies websites as safe or malicious [4]. ML-based models can analyze multiple features such as URL length, domain age, SSL certificate status, and website behavior. Algorithms like **Support Vector Machine (SVM)**, **Random Forest**, **Decision Tree**, and **Naïve Bayes** have shown good results in phishing detecting tasks [5]. Recently, **Deep Learning** models, including **Convolutional Neural Networks (CNN)** and **Long Short-Term Memory (LSTM)** networks, have been used to handle the complex, nonlinear data [6]. This paper reviews recent ML-based phishing detecting techniques. It analyzes the methods, datasets, and performance trends used in current research. The review also highlights commonly used datasets such as **PhishTank** and **UCI Phishing Websites Dataset**. The aim is to find the most effective ML techniques and discuss their challenges, limitations, and future research in this field.

BACKGROUND

Machine Learning (ML) techniques have become increasingly important for phishing detection. It can learn patterns and classify it as safe or malicious [4], [5]. Different ML algorithms use different features from URLs and websites to make proper decisions. The main algorithms used for phishing detection are explained below.

2.1 Machine Learning Algorithms

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.1.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised learning algorithm that separates phishing and legitimate data using hyperplane [4]. It is particularly effective in high-dimensional feature spaces. It performs well even with limited data. SVMs have been widely used in URL-based phishing detection because it provides strong accuracy in binary classification tasks.

2.1.2 Decision Tree

Decision Tree algorithm splits the dataset into smaller subsets based on feature values. Each node in a tree has a simple rule [5]. They are easy to understand and implement. However, they can overfit if not pruned correctly. In phishing detection, they are often used as base learners for group models such as Random Forest.

2.1.3 Random Forest

Random Forest is a group of Decision Trees. It combining their results improves prediction accuracy and stability [5]. This method reduces overfitting and performs well with large datasets. It is widely used in phishing detection for its ability to handle noisy data.

2.1.4 Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It assumes that all features are independent [5]. This method performs well in text and email classification tasks. In phishing detection, it helps identify suspicious patterns in emails or webpages.

2.1.5 Deep Learning Models (CNN, LSTM)

Deep Learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) perform well with complex phishing detection tasks [6]. CNNs are used for analyzing webpage screenshots or raw text data. LSTMs are well known for working with sequential data, such as URLs and email messages [6]. These models can learn features automatically, reducing the need for manual feature generation.

2.2 Key phishing features

ML algorithms rely on carefully selected features that distinguish phishing from legitimate websites [3], [4]. Commonly used features include are mentioned below.

Table 1: Key Features Used in Phishing Detection

Feature Type	Description	Purpose in Detection
URL-based features	URL length, presence of specific characters such as "@" or "-," use of IP address instead of domain name.	Name suspicious URL structures commonly used in phishing websites.
Domain-based Features	Domain age, WHOIS registration validity, DNS record, hosting location.	Detect newly registered or short-lived domains used by attackers.
Content-based features	Forms requesting personal data, fake logos, HTML/JavaScript redirects.	Name deceptive website content is meant to mimic legitimate websites.
Network-based features	Server response time, IP reputation, geolocation, and DNS behavior.	Name malicious hosting services or IP patterns.
Behavioral Features	User interaction data, click behavior, session duration.	Analyze user-side behavior that may show phishing attempts.

2.3 Commonly used datasets

Researchers rely on benchmark datasets for training and evaluating phishing detection models [5], [6].

The most common datasets are listed below.

2.3.1 PhishTank Dataset

PhishTank [7] is publicly available and community-driven dataset having verified phishing URLs. It provides labeled data that helps researchers evaluate ML models on real-world phishing samples. This dataset is commonly used for both training and evaluation.

2.3.2 UCI Phishing Websites Dataset

The UCI Phishing websites Dataset [8] is one of the most widely used repositories. It has thirty features extracted from website URLs, each marked as legitimate or phishing. Researchers commonly use it for evaluating traditional ML models such as SVM, Decision Tree, and Random Forest.

2.3.3 Mendeley Phishing Dataset

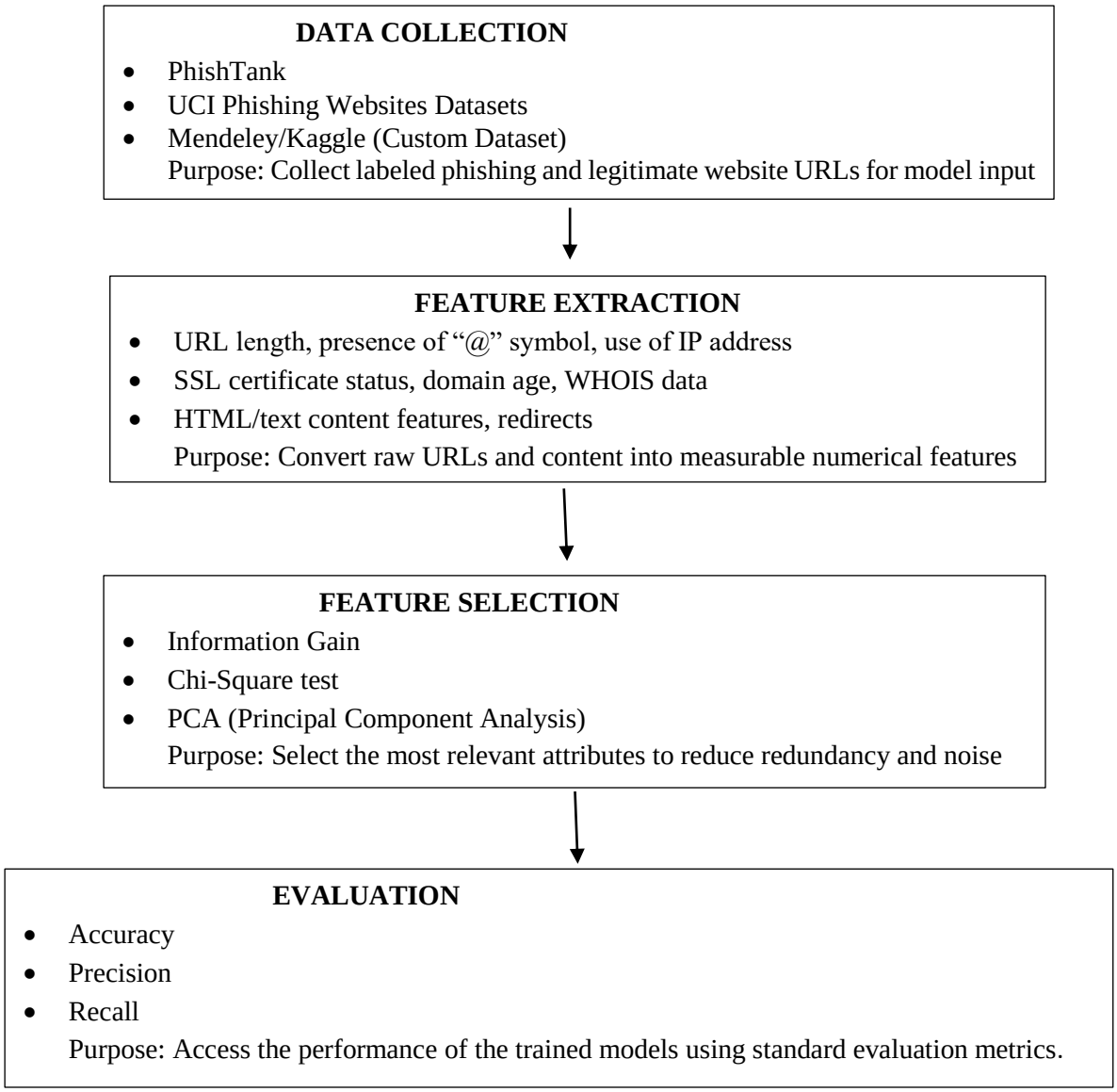
The Mendeley Phishing Dataset [9] includes both legitimate and phishing URLs collected from different sources. It includes a mix of lexical, host-based, and content-based features. It is suitable for ensemble and deep learning methods.

2.3.4 Custom Datasets (Kaggle and others)

In addition to standard datasets, researchers build custom datasets from sources like Kaggle [10]. These datasets may include domain-based or content-based features extracted from phishing databases. Custom datasets are useful for creating hybrid or real-time phishing detection models.

2.4 General framework of Machine Learning for phishing detection

The overall process typically involves key stages. Beginning with data collection and followed by feature extraction, feature selection, and model evaluation. Each stage plays an important role in building an efficient system. The general workflow followed is illustrated in Figure 1 below.



LITERATURE REVIEW

Machine Learning (ML) and Deep Learning techniques have been widely studied for phishing detection. It offers significant improvements over traditional rule-based or blacklist-based approaches. Recent studies have focused on finding effective algorithms, datasets, and evaluation methods for enhancing phishing detection accuracy. Table 2 summarizes the most relevant recent works from 2023-2024, highlighting the algorithms, datasets, and reported performance results. Recent studies show that hybrid deep learning architecture such as CNN-LSTM and transformer-based models outperform classical algorithms like Support Vector Machine and Random Forest when trained on large and diverse datasets. Traditional ML models are still valuable due to their lower computational cost and interpretability. Researchers have also noted that dataset quality, feature selection, and preprocessing significantly affect model accuracy and generalizability [13], [18], [19]. Overall, the literature reflects a clear evolution from rule-based systems towards **data-driven, hybrid, and transformer-based approaches** for phishing detection [12], [16], [20].

Table 2-Summary of Recent ML-Based Phishing Detection Works (2023-2024)

Year	Algorithm/Model	Datasets	Accuracy/ Performance	Key Findings
2023	LSTM, CNN, Hybrid LSTM-CNN	PhishTank/Mixed URL sets	High	Hybrid LSTM-CNN models improve detection accuracy on URL data. [11]
2023	PhishTransformer (Transformer-based)	Custom (URLs content)	>97%	Transformer models using both URL and content achieve superior accuracy. [12]
2024	1D-CNN (Lightweight model)	PhishTank/URL Datasets	96-98%	Lightweight CNN is suitable for real-time phishing detection. [13]
2024	CNN+LSTM (Hybrid)	Alexa+ PhishTank	97%	Large-scale study (500k URLs) showing effectiveness of CNN/LSTM. [14]
2023	CNN-Fusion (Character-level CNN)	Public URL datasets	High accuracy	Character-level CNN reduces preprocessing needs. [15]
2024	Transformer-based (BERT, DistilBERT)	Phishing Email datasets	98%	Fine-tuned transformer models outperform classical ML on email phishing. [16]
2024	IPSDM (Improved BERT variant)	Email datasets (Phish/ham)	>97%	Improve fine-tuning techniques enhance transformer performance. [17]
2023	Random Forest, SVM, XGBoost	UCI/ PhishTank	93-96%	Ensembles and tree-based models stay effective with feature selection. [18]
2024	CNN (Deep Learning)	Mixed datasets	99.2%	CNN achieves high accuracy; results vary by dataset quality. [19]
2024	Survey / Review papers	_	_	Research trend shifting from classical ML to hybrid and transformer models.

4. Comprehensive Analysis

The comprehensive analysis of existing studies reveals significant differences in the performance of traditional Machine Learning and Deep Learning techniques for phishing detection. Traditional ML algorithms continue to prove competitive performance in URL-based phishing detection due to their simplicity, interpretability, and efficiency on smaller datasets [18]. These models are particularly effective when combined with well-engineered lexical and domain-based features extracted from datasets such as UCI Phishing Websites and PhishTank [7], [8]. However, their performance tends to degrade when faced with unstructured data, such as website content or email text [11], [12]. In contrast, Deep Learning (DL) models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have proven superior capability for capturing complex patterns and feature relationships [11], [13], [14]. CNN models are highly effective for analyzing URL sequences and visual webpage representations. LSTM-based models are better suited for handling sequential or contextual data, such as email messages or character-level URL strings. Hybrid architectures, such as CNN-LSTM

and transformer-based approaches like Phish Transformer and BERT fine-tuning, have achieved ultramodern performance with reported accuracies exceeding 97-99% [12], [16], [17], [19]. These models also require minimal manual feature engineering because they can automatically learn discriminative representations from raw data. The observation also shows that dataset selection plays a key role in model performance. Most studies rely on PhishTank and UCI datasets, while others use custom URL-based corpora or email datasets for specific use cases [7], [8], [12]. Studies using larger and more diverse datasets tend to report higher accuracy, often above 97%. Smaller or imbalanced datasets show reduced generalization. Accuracy trends from recent studies suggest a clear improvement from around 90-93 % with traditional ML approaches to 98-99% with deep and hybrid models [13], [19]. However, Deep Learning models usually require more computational resources, including longer training time, and larger datasets to achieve stable performance.

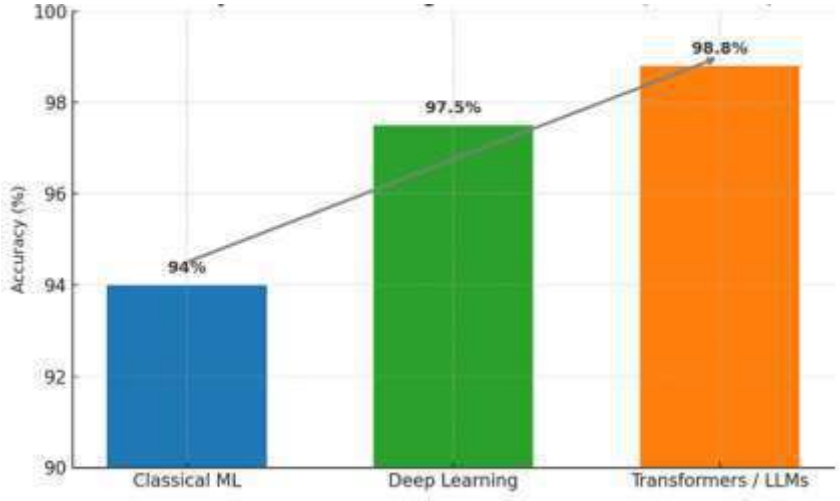


Figure 2: Accuracy Trends of Machine-Learning and Deep-Learning Models in Phishing Detection (2021–2024). Source: Created by the author based on literature [11]– [20].

CHALLENGES AND FUTURE SCOPE

Despite the progress in phishing detection research, there are challenges and limitations. These issues hinder the development of real-time detection systems. One issue is **dataset imbalance.** Here, phishing samples significantly outnumber legitimate ones or vice versa [13], [18]. This imbalance can make models become biased, reducing accuracy and increasing false positive rates when applied to real-world data. Another significant challenge is **real-time detection**. Many deep learning models work well with labs but not with real-time datasets. Therefore, it is difficult to deploy in real-time scenarios such as browser extensions or email gateways [14], [19]. In addition, phishing attacks are increasingly appearing on social media platforms. These attackers use shortened URLs, deceptive posts, and cloned profiles that are difficult for traditional ML systems to find [12]. The growing sophistication of adversarial attacks where malicious actors intentionally manipulate URLs or website features to evade detection [16], [17]. In addition, phishing attacks are no longer limited to emails or banking fraud. Latest trends show a rise in employment-related fraud. For instance, users have reported receiving fake job offer emails claiming a high-paying work-from-home role. The sent link redirects them to fraudulent websites demanding payment for registration. Such attacks exploit user trust and urgency, making detection difficult for both individuals and automated systems [1], [20]. Looking ahead, researchers are developing hybrid models that can combine both traditional ML and Deep Learning [14], [16]. Such models could balance accuracy and efficiency, enabling real-world usage. Additionally, the integration of transformer architecture and Large Language Models (LLMs), such as fine-tuned BERT and GPT-based systems, can also help analyze both textual and visual content [17], [20]. Future work should focus on the development of real-time, browser-based extensions and lightweight ML frameworks that learn continuously from new phishing data. Another goal is to make models more explainable and resistant to adversarial attacks. Building adaptive and transparent

systems will help future generations of phishing detection.

CONCLUSION

Phishing remains one of the most common and evolving cyber threats today. It continues to affect individuals, organizations, and financial systems. This review clearly shows that Machine Learning (ML) plays an important role in detecting phishing automatically through data-driven analysis. Traditional ML algorithms such as SVM, Decision Tree, and Random Forest continue to perform well. They are simple with interpretable results for URL-based detection. However, Deep Learning models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), and Transformer-based architectures consistently outperform conventional techniques. These models can handle large-scale data. Even with these advancements, challenges remain. There is still a need for more robust and diverse datasets, improved handling of data imbalance, and better real-time detection capabilities. Future researchers should focus on hybrid and explainable AI models that combine the strengths of traditional ML and Deep Learning. Such models should be transparent, scalable, and adaptable. Building these systems will be a key to creating safer online environments.

Author Biography

Teja Hulse has completed her bachelor's degree in computer science and engineering from SDM College of Engineering and Technology, Dharwad, India. She maintains a strong foundation in programming, networking and data management. Teja has worked on multiple academic projects, including a Health-Record Vault DApp, a blockchain-based decentralized application designed for secure sharing of medical data. Another project includes a Crop Recommendation and Leaf Disease Detection System using ML techniques, collecting real-time datasets, and achieving an accuracy of 78%. She has earned professional certifications such as a Generative AI professional from Oracle, a Data Science certificate from Academor, reflecting her commitment to continuous learning. Her research interests include Machine Learning, Cybersecurity and Artificial Intelligence. She has worked on review-based studies exploring ML techniques for secure digital systems. Teja aims to contribute to the development of secure and intelligent digital technologies

REFERENCE

1. APWG, "Phishing Activity Trends Report," 2024. [Online]. Available: https://apwg.org/
2. K. Hong, "The State of Phishing Attacks," Commun. ACM, vol. 55, no. 1, pp. 72–81, 2023.
3. A. Jain and B. Gupta, "Phishing Detection: Analysis of Attacks and Mitigation Strategies," Secure. Privacy, vol. 5, no. 4, pp. 1–12, 2022.
4. M. Basit, S. Zafar, and M. Qureshi, "A Machine Learning Approach for Phishing Detection Using URL Features," Comput. Secur., vol. 132, pp. 102948–102957, 2023.
5. S. Patil, A. Sharma, and K. Rao, "Comparative Study of ML Algorithms for Phishing Website Detection," Int. J. Recent Adv. Sci. Eng. Technol. (IJRASET), vol. 11, no. 6, pp. 45–52, 2023.
6. X. Li and J. Wang, "Deep Learning for Phishing Detection: A Survey," IEEE Access, vol. 12, pp. 78234–78249, 2024.
7. PhishTank, "PhishTank Dataset," [Online]. Available: https://www.phishtank.com/
8. D. Dua and E. Karra Taniskidou, "UCI Machine Learning Repository: Phishing Websites Data Set," Univ. California, Irvine, 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/phishing+websites
9. Mendeley Data Repository, "Phishing Dataset for Machine Learning Research," Elsevier, 2023. [Online]. Available: https://data.mendeley.com/
10. Kaggle, "Phishing URL Dataset," 2023. [Online]. Available: https://www.kaggle.com/datasets/
11. M. Alshingiti, N. Alotaibi, and M. Khan, "Three Deep Learning Methods for Phishing URL Detection," IEEE Access, vol. 11, pp. 87456–87467, 2023.
12. H. Asiri, F. Alharbi, and A. Alzahrani, "Phish Transformer: Transformer-Based Phishing Detection Model," Comput. Secur., vol. 131, pp. 103236–103245, 2023.
13. M. Haq, A. Alam, and R. Khan, "Lightweight 1D-CNN Model for URL-Based Phishing Detection," IEEE Access, vol. 12, pp. 102394–102406, 2024.
14. V. Ghalechyan, A. Petrosyan, and H. Hakobyan, "Large-Scale Phishing Detection Using CNN and

LSTM," in Proc. Springer Int. Conf. Inf. Syst. Secur., 2024, pp. 221–232.
15. A. Hussain, "CNN-Fusion: A Character-Level Deep Learning Model for Phishing URL Classification," Expert Syst. Appl., vol. 234, pp. 120944–120953, 2023.
16. S. Kumar, R. Sharma, and L. Gupta, "Transformer-Based Approaches for Email Phishing Detection," IEEE Access, vol. 12, pp. 109823–109836, 2024.
17. F. Jamal, "IPSDM: Improved BERT Fine-Tuning Model for Phishing Email Detection," J. Inf. Secur., vol. 15, no. 2, pp. 45–55, 2024.
18. R. Choudhary, N. Reddy, and A. Verma, "Comparative Study of ML Algorithms for Phishing Website Detection," Int. J. Recent Adv. Sci. Eng. Technol. (IJRASET), vol. 11, no. 5, pp. 58–65, 2023.
19. P. Singh, D. Malhotra, and R. Arora, "Deep CNN Model for Phishing Detection with High Accuracy," in Proc. Springer Int. Conf. Comput. Vis. Intell. Syst., 2024, pp. 312–321.
20. A. Sharma and N. Gupta, "A Comprehensive Survey on Machine Learning-Based Phishing Detection," ACM Comput. Surv., vol. 56, no. 4, pp. 1–28, 2024.

HOW TO CITE: Teja Ravi Hulse*, A Review of Machine Learning Techniques for Phishing Detection, Int. J. Sci. R. Tech., 2025, 2 (11), 616-622. https://doi.org/10.5281/zenodo.17668288