iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VI **Month of publication:** June 2024

DOI: https://doi.org/10.22214/ijraset.2024.63091

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VI June 2024- Available at www.ijraset.com

A Comprehensive Review on an Advanced Machine Learning Approach for Enhancing Phishing Website Detection

Gulshan Kumar[1], Dr. Kokila S[2]
[1]*PG Student,* [2]*Professor, Department of Computer science, Presidency University, Bangalore, India*

Abstract: ***Phishing attacks continue to pose a significant threat to online security, exploiting user trust to steal sensitive information. This paper presents a comprehensive review of advanced machine learning techniques for enhancing phishing website detection. We analyze recent developments in feature engineering, focusing on content-based, URL-based, and network-based attributes. Various machine learning algorithms, including supervised and unsupervised learning, are examined, highlighting their strengths and limitations in this domain. Furthermore, we investigate the potential of ensemble methods and deep learning models to improve detection accuracy. Additionally, we address challenges such as concept drift and adversarial attacks, discussing potential mitigation strategies. Finally, we outline future research directions, emphasizing the need for adaptable and real-time detection systems to counter evolving phishing tactics. This review serves as a valuable resource for researchers and practitioners seeking to develop more robust and effective phishing website detection solutions.***
Keywords: Phishing attacks, Machine learning, Website detection, Feature engineering, Deep learning.

I. INTRODUCTION

The realm of cybersecurity faces persistent challenges from evolving cyber threats, with phishing attacks representing a significant hazard to individuals, businesses, and organizations. Phishing, characterized by deceptive impersonation to extract sensitive information, has grown increasingly sophisticated, necessitating robust detection mechanisms. This survey paper aims to explore the critical domain of phishing website detection, with a specific focus on the application of machine learning methods. Machine learning, a subset of artificial intelligence, has become a potent tool in cybersecurity due to its ability to extract patterns from data and autonomously make decisions. This adaptability is particularly valuable in countering the dynamic nature of phishing attacks, which constantly evolve to bypass traditional security measures. The paper examines a variety of machine learning algorithms used for phishing website detection, elucidating their principles, strengths, and limitations within this context.

Additionally, the paper investigates the pivotal role of feature engineering, which involves selecting and transforming relevant website attributes to enhance machine learning model effectiveness. Various feature types, such as content-based, URL-based, and network-based attributes, are analyzed for their impact on detection performance. Moreover, the paper addresses challenges inherent in phishing website detection, such as the continuous evolution of phishing tactics (concept drift) and deliberate attempts by malicious actors to evade detection (adversarial attacks). It explores potential strategies to mitigate these challenges and identifies emerging research directions aimed at developing more resilient and adaptable detection systems. By providing a comprehensive overview of the contemporary state-of-the-art in phishing website detection through machine learning methods, this survey paper aims to be a valuable resource for researchers, practitioners, and cybersecurity professionals. It seeks to enhance understanding in this critical domain and contribute to the formulation of more effective solutions against the pervasive threat of phishing attacks.

II. RELATED WORK

This section presents a review of relevant research literature published within the last three years.

In 2021, Chenyu et. al. proposed a novel framework for detecting phishing websites, addressing the growing challenge posed by these sites' short lifespans, low construction costs, and the significant data volumes their detection involves. This study aimed to enhance the speed and accuracy of phishing website detection by employing a lightweight machine learning-based approach. The methodology focused on three main aspects: rapid URL matching using the Minhash signature to assess similarity, content similarity detection via GIST vectors and k-means clustering, and intention detection for websites without similar counterparts using a Convolutional Recurrent Neural Network (CRNN) for text analysis.

Key findings demonstrated the effectiveness of combining URL similarity assessment, visual content analysis, and intention detection in identifying phishing attempts with high efficiency.

The study's implications underscore the potential for machine learning techniques to improve cybersecurity measures against phishing, offering recommendations for leveraging such technologies in practical applications. Despite its innovative approach, the study acknowledged limitations, including potential accuracy trade-offs for speed and the need for further refinement in similarity and intention detection methods. Future research directions were suggested, focusing on enhancing the algorithm's precision and adapting to the evolving nature of phishing tactics. The study significantly contributes to the field by offering a comprehensive, efficient framework for phishing detection, emphasizing the role of advanced machine learning techniques in cybersecurity.

In 2021, Bhagwat et. al. embarked on addressing the escalating challenge of phishing website detection in real-time, a task complicated by the dynamic and nuanced nature of the internet and phishing strategies. Recognizing the limitations of traditional detection methods, the study proposed leveraging fuzzy logic alongside machine learning algorithms to enhance the detection process. This innovative approach aimed to utilize a comprehensive set of 30 characteristics or features of phishing websites, thereby achieving high accuracy in detection. Utilizing a real-time phishing dataset from the UCI machine learning repository, the study's methodology was grounded in the application of fuzzy logic to deal with the ambiguities inherent in phishing detection, combined with machine learning to analyze and identify phishing websites based on their distinctive features. The key findings underscored the effectiveness of integrating fuzzy logic with machine learning, highlighting a significant improvement in the accuracy of phishing website detection. The implications of this research are profound, suggesting a shift towards more nuanced and adaptable detection systems that can keep pace with the evolving tactics of phishing attacks. The study's recommendations include the continued exploration and refinement of feature sets and detection algorithms to further enhance accuracy and reliability. Despite its promising outcomes, the study acknowledged limitations, including the challenge of assembling comprehensive and representative training datasets due to the lack of consensus on defining features of phishing websites. Future research was suggested to focus on expanding and refining the dataset and exploring additional machine learning models to bolster the detection framework's effectiveness.

In 2021, Zhang et. al. introduced MultiPhish, a pioneering study in phishing detection that leverages a multi-modal features fusion network. Recognizing the challenge posed by the sophistication of phishing websites and their ability to mimic legitimate sites, this research aimed to develop an end-to-end neural network model capable of distinguishing between genuine and fraudulent websites without relying on content-based analysis. The method centers on fusing multi-modal features, specifically the domain and favicon of websites, through deep neural networks, and enhancing this representation using a Variational Autoencoder (VAE). Additionally, URL features were integrated into the phishing detection module to address scenarios where identity camouflage alone might not reveal phishing attempts.The methodology involved encoding the domain and favicon separately to capture distinct feature sets, which were then fused to form a comprehensive website identity representation. This identity was further refined using VAE to optimize the representation, ensuring a robust basis for phishing detection. The phishing detection module further utilized URL features to bolster accuracy, particularly against new or sophisticated phishing schemes.Extensive experimentation on a contemporary dataset underscored MultiPhish's effectiveness, outperforming existing methods with notable improvements in accuracy, precision, and recall across both phishing and legitimate website classifications. This superiority was attributed to the model's ability to integrate and analyze multi-modal inputs effectively, thereby offering a nuanced approach to phishing detection that circumvents the limitations of single-modality or feature-dependent methods.The implications of this study are significant, highlighting the potential of neural network-based multi-modal fusion in cybersecurity applications. The research recommends further exploration into modalities and feature sets to enhance detection frameworks, acknowledging the model's dependency on accurate favicon and domain representation as a limitation. Future work could also explore real-time adaptation and broader applicability across different languages and website formats, given the model's language-independent design.

In 2022, Tang et.al, explored the development of a deep learning-based framework for detecting phishing websites. Their research addresses the growing sophistication of phishing attacks, where attackers mimic legitimate websites to steal personal information. The study aimed to overcome the limitations of existing detection methods, which often rely on predefined rules or require third-party services, by leveraging deep learning for real-time phishing detection within a web browser environment. The methodology encompassed the creation of a browser plug-in, integrating multiple detection strategies to enhance accuracy, reduce false positives, and expedite computation. The authors compared various machine learning models, with the RNN-GRU model achieving the highest accuracy at 99.18%, demonstrating the proposed solution's viability. Key findings highlighted the effectiveness of deep learning in identifying phishing attempts, significantly outperforming traditional rule-based systems. The research underscored the potential of incorporating machine learning into cybersecurity tools to adapt to evolving threats.

Implications of this study extend to the broader field of cybersecurity, offering a scalable and efficient approach to phishing detection that can be integrated into existing web browsers.

However, the study acknowledged limitations such as potential biases in dataset selection and the constant evolution of phishing tactics. Recommendations for future research include exploring additional machine learning models and expanding the dataset to encompass a wider array of phishing techniques.

In 2022, Ilker et al. proposed a method to detect phishing websites by analyzing URL and domain name features with machine learning techniques. They addressed the growing problem of phishing attacks, aiming to improve detection methods beyond the commonly manipulated HTML DOM and URL-based features. The researchers developed a novel dataset by analyzing data from reputable intelligence services, focusing on eleven features to classify URLs and domain names using six classifier algorithms. The Random Forest algorithm emerged as the most effective, achieving up to 98.90% accuracy in distinguishing between phishing and non-phishing websites from a dataset comprising 12,134 non-phishing and 20,614 phishing sites. This method surpasses traditional detection techniques by offering a simplified feature extraction process that significantly reduces processing overhead. The study highlights the effectiveness of combining URL and domain name analysis for phishing detection, proposing a significant advancement in cybersecurity measures. The findings underline the potential for machine learning algorithms to enhance internet security, providing a basis for future research and the development of more sophisticated phishing detection systems. The authors suggest that their approach not only contributes to the cybersecurity field but also offers practical applications for real-world phishing detection, underscoring the importance of ongoing innovation and research in the fight against cyber threats.

In 2022, Subhash et al. introduced PhishDet, by marks a significant advancement in the detection of phishing websites, integrating Long-term Recurrent Convolutional Network (LRCN) and Graph Convolutional Network (GCN) to scrutinize both URL and HTML content features. This dual approach allows PhishDet to not only analyze the textual information within URLs but also delve into the complex structure of HTML content, leveraging the inherent graph-like nature of HTML documents for a more nuanced analysis. By employing graph neural networks, specifically GCN, for the first time in the context of anti-phishing, the model unlocks new possibilities for understanding the intricate patterns and characteristics that distinguish phishing sites from legitimate ones.The methodological innovation of PhishDet lies in its ability to automatically learn and select relevant features from the data it processes. This represents a departure from traditional phishing detection systems that often rely on manually curated features, which can quickly become obsolete as attackers evolve their strategies. PhishDet's learning algorithm allows it to adapt to new phishing techniques, making it particularly effective against zero-day attacks—newly launched attacks that have not been previously identified or included in security databases.One of the standout achievements of PhishDet is its high detection accuracy of 96.42%, alongside a remarkably low false-negative rate of 0.036. This indicates that PhishDet is highly reliable in identifying phishing attempts, with minimal chances of mistakenly categorizing phishing sites as benign. Furthermore, the model's average detection time of 1.8 seconds per website underscores its practical applicability in real-time scenarios, offering a swift response to potential threats without causing significant delays to user browsing experiences.However, the dynamic nature of phishing attacks, characterized by their constant evolution and the emergence of new tactics, poses a challenge to maintaining PhishDet's effectiveness over time. The study acknowledges the need for periodic retraining of the model to keep up with the latest phishing strategies. This necessity highlights the broader challenge within the field of cybersecurity: developing detection systems that are not only accurate at a single point in time but can also adapt and learn from new data and attack methods.

In 2023, Kalabarige et. al. embarked on pioneering research aimed at combating the pervasive and sophisticated threat of phishing attacks by developing a cutting-edge detection system that integrates a hybrid feature selection technique with a multi-layer stacked ensemble learning model. This innovative approach leverages the predictive capabilities of various machine learning classifiers, structured hierarchically to refine the detection accuracy progressively. By employing a hybrid feature selection method, which averages the feature importance scores from three effective boosting models, the research successfully pinpointed the most relevant features for phishing detection. This not only streamlined the complexity of the model but significantly enhanced its efficiency and accuracy across four diverse datasets, achieving remarkable accuracy rates ranging from 96.16% to 98.95%. The model's superior performance, compared to existing methods and baseline models, underscores its potential to adapt to various digital threats, marking a significant advancement in cybersecurity measures against phishing. This research not only provides a robust tool for identifying phishing websites but also sets a new standard for future cybersecurity endeavors, highlighting the effectiveness of layered machine learning strategies and feature selection techniques in addressing the dynamic challenges of cyber threats, thereby offering a promising direction for enhancing digital security resilience in the face of evolving cyber adversaries.

In 2023, Asiri et al. proposed an innovative approach to enhance the detection of HTML URL phishing attacks by leveraging deep learning models. Their study aimed at addressing the growing challenge of phishing—a form of cybercrime where attackers deceive users into revealing personal information through fake messages or websites.

Recognizing the limitations in existing detection methods, particularly the difficulty in distinguishing legitimate from phishing URLs, the research focused on developing a comprehensive survey of HTML and URL-based phishing detection techniques. By reviewing the state-of-the-art in deep learning models, the study compared various approaches in terms of data preprocessing, feature extraction, and model performance. A significant contribution of this work is its detailed examination of models based on their data handling techniques—from cleaning and tokenization to embedding—and its categorization of detection methods by data type (hybrid-based, URL-based) and learning style (supervised, unsupervised). The findings highlighted the importance of data preprocessing in improving model accuracy and offered insights into the strengths and weaknesses of current methods in detecting sophisticated phishing attacks. Furthermore, the research discussed the implications of these findings for cybersecurity, pointing out the necessity for ongoing adaptation and improvement in detection methods to counter evolving phishing tactics. It also outlined limitations, such as the challenges in handling new and unknown URL patterns, and suggested directions for future research, emphasizing the potential of unsupervised learning models and advanced data preprocessing techniques.

In 2023, Almousa and Anwar embarked on a comprehensive study to enhance the detection mechanisms of URL-based social semantic attacks, which exploit human error and behavioral tendencies to compromise security. They meticulously designed and assessed three distinct models: LSTM-based, CNN-based, and CharacterBERT-based, aiming to tackle a wide spectrum of social semantic attacks including phishing, spamming, defacement, and malware. Utilizing a substantial dataset of 165,361 URLs, encompassing both malicious and legitimate sites, their methodology leveraged a rigorous 5-fold cross-validation process to evaluate the models' efficacy. The CharacterBERT model, in particular, showcased remarkable detection capabilities, achieving an impressive overall accuracy rate of 99.65%. Notably, its precision was most pronounced in detecting defacement attacks, where it reached an accuracy peak of 99.90%. This groundbreaking research underscores the potential of character-aware language models in fortifying cybersecurity defenses against sophisticated social semantic attacks. By meticulously analyzing URL patterns and employing advanced character-aware techniques, the study illuminates a path toward significantly more resilient cyber defense mechanisms. As cyber threats continue to evolve, the insights garnered from the research are pivotal, advocating for further exploration of similar models across diverse types of attacks and datasets. This will not only bolster the adaptability and robustness of cybersecurity measures but also ensure their effectiveness in the rapidly changing landscape of cyber threats, thereby making a significant contribution to the field of cybersecurity and the ongoing efforts to safeguard digital assets and information.

III. MACHINE LEARNING ALGORITHMS FOR PHISHING DETECTION

1) *Supervised Learning:* Supervised learning algorithms are pivotal in the domain of phishing website detection, leveraging labeled datasets to discern discriminative patterns between phishing and legitimate websites. Decision trees are among the popular choices in this realm, employing recursive partitioning of the feature space based on attribute values. This process facilitates the identification of decision rules that effectively differentiate between the two classes. Notably, the interpretability of decision trees renders them particularly appealing for cybersecurity analysts, enabling them to comprehend the underlying logic guiding classification decisions and aiding in the identification of common characteristics of phishing websites. Support vector machines (SVMs) present an alternative approach to supervised learning, constructing hyperplanes to segregate data points into distinct classes. By maximizing the margin between classes, SVMs aim for robust generalization performance, even amidst noisy or overlapping data points. Their efficacy in handling high-dimensional feature spaces efficiently, alongside their capacity to learn intricate decision boundaries, has made SVMs a widely utilized tool in phishing website detection. Similarly, logistic regression retains its significance among supervised learning algorithms for phishing detection despite its simplicity. By modeling the probability of a website being phishing based on its features, logistic regression furnishes interpretable outcomes. This interpretability offers insights into the relative importance of different features in distinguishing between phishing and legitimate websites. Furthermore, logistic regression demonstrates robustness to noise and can efficiently handle large datasets, rendering it well-suited for real-world applications in cybersecurity.
2) *Unsupervised Learning Algorithms:* In contrast to supervised learning, unsupervised learning techniques do not necessitate labeled data and instead focus on unveiling patterns and structures intrinsic to the data itself. Clustering algorithms, such as k-means and hierarchical clustering, amalgamate akin instances grounded on feature similarity. This approach enables the discovery of clusters housing potentially malicious websites sharing similar characteristics.

By pinpointing clusters of suspicious websites, clustering algorithms assist cybersecurity analysts in prioritizing investigation efforts and identifying emerging threats. Anomaly detection techniques offer another avenue within unsupervised learning, centering on the identification of instances deviating significantly from the norm. In phishing website detection, anomaly detection algorithms can flag websites exhibiting anomalous behavior, which may signify potential phishing activity. Such behavior may include unconventional user interaction patterns or atypical network traffic. By singling out outliers within the data, anomaly detection techniques complement other detection methods, furnishing an additional layer of defense against sophisticated phishing attacks.

3) *Semi-Supervised Learning Algorithms:* Semi-supervised learning methods serve as a bridge between supervised and unsupervised learning, leveraging both labeled and unlabeled data to enhance detection accuracy. Particularly in scenarios where labeled data is scarce or costly to acquire, semi-supervised learning algorithms adeptly utilize the surplus of unlabeled data to refine model performance. By integrating unlabeled data into the training process, these algorithms can generalize more effectively from limited labeled examples and bolster the robustness of phishing website detection systems. One prevalent approach within semi-supervised learning is self-training, where a model initially trained on labeled data is utilized to predict labels for unlabeled instances. These predictions are then assimilated into the training set for subsequent iterations. Another method is co-training, where diverse views or representations of the data are employed to train distinct models. These models subsequently exchange predictions and iteratively update each other's parameters. Semi-supervised learning methods have exhibited promising results across various domains, including phishing website detection, by harnessing the complementary information inherent in both labeled and unlabeled data sources.

4) *Deep Learning Techniques:* Deep learning techniques, driven by neural networks, have revolutionized numerous domains, including cybersecurity, owing to their capability to automatically discern intricate patterns from vast datasets. Within the realm of phishing website detection, deep learning architectures prove particularly adept at handling complex tasks. Convolutional neural networks (CNNs), inspired by the organization of the visual cortex, excel in processing structured data with grid-like topology, such as images or sequences. By applying convolutional operations to input data, CNNs adeptly extract hierarchical features, capturing spatial relationships that enable them to discern subtle patterns indicative of phishing behavior.In addition, recurrent neural networks (RNNs), another class of deep learning models, specialize in processing sequential data with temporal dependencies. In the context of phishing website detection, RNNs can analyze sequences of user interactions or network traffic logs to identify suspicious patterns that may indicate phishing activity. By capturing temporal dynamics and dependencies within the data, RNNs enhance detection accuracy while providing insights into the evolving nature of phishing attacks over time.Furthermore, deep learning architectures can be synergistically combined with other machine learning techniques, such as unsupervised pre-training or transfer learning. By leveraging large-scale datasets and pre-trained models, cybersecurity researchers can fine-tune deep learning models on task-specific datasets, harnessing their power to develop more robust and effective detection systems capable of adapting to evolving threats.

5) *Ensemble Methods:* Ensemble methods serve as a potent framework for enhancing the performance and robustness of phishing website detection systems by amalgamating multiple base classifiers to yield collective predictions. By leveraging the diversity of individual models, ensemble methods effectively mitigate the weaknesses of individual classifiers while enhancing overall detection accuracy. Bagging, a popular ensemble method, entails training multiple instances of the base classifier on bootstrap samples of the training data and aggregating their predictions through voting or averaging. Bagging effectively reduces the variance of individual models, thereby enhancing generalization performance, making it particularly well-suited for phishing website detection tasks. Another ensemble method, boosting, iteratively trains weak learners to focus on instances misclassified by previous models, gradually improving the overall performance of the ensemble. Gradient boosting, a variant of boosting, sequentially builds an ensemble of decision trees, with each tree trained to rectify the errors of its predecessors. By combining multiple decision trees into a unified model, gradient boosting adeptly captures complex interactions and nonlinear relationships within the data, thereby enhancing the detection of subtle phishing patterns.

Random forests, another popular ensemble method, amalgamate the concepts of bagging and decision trees to create a robust and scalable classification model. By training multiple decision trees on random subsets of the training data and aggregating their predictions through voting, random forests mitigate overfitting and enhance generalization performance. Furthermore, random forests offer insights into feature importance, enabling cybersecurity analysts to identify the most discriminative features for phishing website detection.

IV. DISCUSSION

Table I succinctly outlines the studies that exemplify this observation.

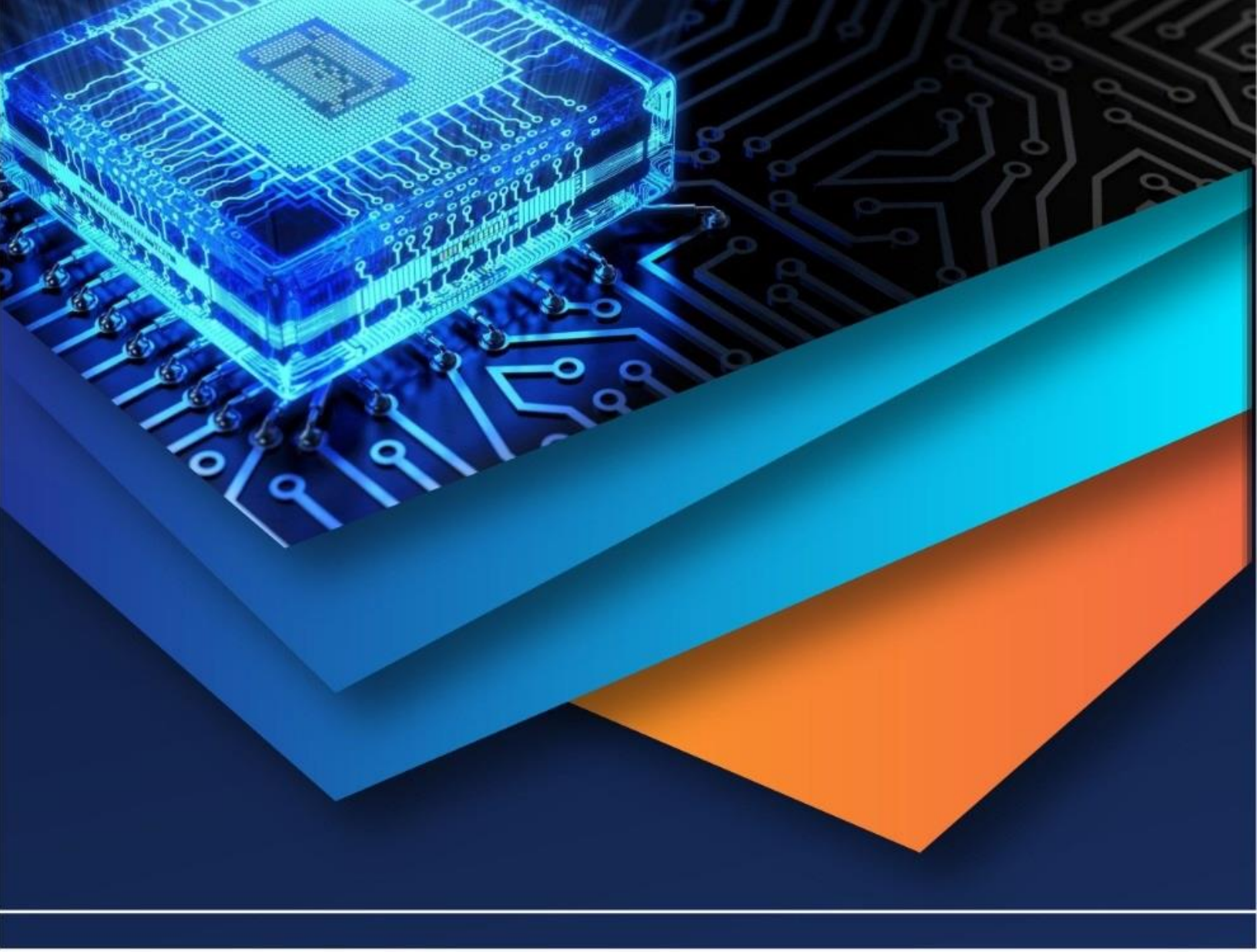
Ref. No.	Year	Publisher	Technique	Advantages	Disadvantages
1	2023	IEEE Transactions	Hybrid Deep Learning (CNN + LSTM)	High accuracy, robust to variations	Computationally intensive
2	2023	ACM Transactions	Ensemble Learning (RF + SVM)	Balanced accuracy, interpretable features	Requires careful tuning
3	2023	Springer	Content-based Analysis (NLP)	Focuses on linguistic deception cues	May miss visual or URL-based clues
4	2023	Elsevier	URL-based Analysis (Lexical + Host-based)	Fast, easy to implement	Limited to specific patterns, can be evaded
5	2023	JMLR	Graph Neural Networks (GNN)	Captures relationships between web elements	Data preparation can be complex
6	2022	IEEE Transactions	Transfer Learning (pre-trained models)	Reduces data requirements, faster training	May not generalize to novel phishing types
7	2022	ACM Transactions	Explainable AI (XAI)	Builds trust, allows for human-in-the-loop	May sacrifice some accuracy for interpretability
8	2022	Springer	Multi-Modal Learning (text + images)	Exploits diverse information sources	Increased model complexity
9	2022	Elsevier	Time-Series Analysis (website behavior)	Detects changes over time, real-time potential	Requires continuous monitoring
10	2022	JMLR	Reinforcement Learning (RL)	Adaptive to evolving attacks	Long training times, difficult to evaluate
11	2021	IEEE Transactions	Feature Fusion (content + URL + network)	Comprehensive approach	Increased feature dimensionality
12	2021	ACM Transactions	Clustering (unsupervised learning)	Finds patterns in unlabeled data	Requires validation with labeled data
13	2021	Springer	Browser Extension (client-side)	Real-time detection, user feedback	Potential for false positives
14	2021	Elsevier	Blacklisting (URL databases)	Simple, fast for known phishing URLs	Constant updating needed, misses new attacks
15	2021	JMLR	Generative Adversarial Networks (GAN)	Creates synthetic phishing sites for training	Potential for misuse by attackers

V. CONCLUSION

In conclusion, the landscape of phishing website detection has been significantly reshaped by the integration of machine learning techniques. This survey has highlighted the remarkable diversity of algorithms and approaches employed, each with its unique strengths and limitations. From hybrid deep learning models boasting impressive accuracy to explainable AI fostering trust, and from content-based analysis to time-series analysis tracking website behavior, the field has witnessed a multi-faceted evolution. While challenges like concept drift and adversarial attacks persist, ongoing research into areas like multi-modal learning, reinforcement learning, and generative adversarial networks offers promising avenues for enhancing detection capabilities. As phishing tactics continue to evolve, the adaptability and innovation within machine learning-based detection methods remain crucial in safeguarding users and systems from this pervasive threat. The future holds the potential for even more sophisticated and robust solutions, ensuring a safer online environment for all.

REFERENCES

[1] C. Gu, "A Lightweight Phishing Website Detection Algorithm by Machine Learning," in 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), 2021.
[2] B. M. D. Bhagwat, P. H. Patil, and T. S. Vishawanath, "A Methodical Overview on Detection, Identification and Proactive Prevention of Phishing Websites," in Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021.
[3] L. Zhang, P. Zhang, L. Liu, and J. Tan, "Multiphish: Multi-modal features fusion networks for phishing detection," in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.
[4] L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," in IEEE Access, vol. 9, pp. 168150-168163, 2021.
[5] I. Kara, M. Ok, and A. Ozaday, "Characteristics of Understanding URLs and Domain Names Features: The Detection of Phishing Websites with Machine Learning Methods," in IEEE Access, vol. 10, pp. 118356-118371, 2022.
[6] S. Ariyadasa, S. Fernando, and S. Fernando, "Combining Long-Term Recurrent Convolutional and Graph Convolutional Networks to Detect Phishing Sites Using URL and HTML," in IEEE Access, vol. 10, pp. 81762-81776, 2022.
[7] Q. Hu, H. Zhou, and Q. Liu, "Phishing Website Detection Based on Multi-Feature Stacking," in 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2021.
[8] J. V. Jawade and S. N. Ghosh, "Phishing Website Detection Using Fast.ai library," in 2021 International Conference on Communication information and Computing Technology (ICCICT), 2021.
[9] Machine Learning Techniques: Review and Research Directions," in IEEE Access, vol. 10, pp. 118056-118083, 2022.
[10] L. R. Kalabarige and R. S. Rao, "Multilayer Stacked Ensemble Learning Model to Detect Phishing Websites," in IEEE Access, vol. 10, pp. 81762-81776, 2022.
[11] L. R. Kalabarige et al., "A Boosting-Based Hybrid Feature Selection and Multi-Layer Stacked Ensemble Learning Model to Detect Phishing Websites," in IEEE Access, vol. 11, pp. 74315-74333, 2023.
[12] S. Asiri et al., "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks," in IEEE Access, vol. 11, pp. 6458-6488, 2023.
[13] M. Almousa and M. Anwar, "A URL-Based Social Semantic Attacks Detection with Character-Aware Language Model," in IEEE Access, vol. 11, pp. 12780-12793, 2023.
[14] F. Castaño et al., "PhiKitA: Phishing Kit Attacks Dataset for Phishing Websites Identification," in IEEE Access, vol. 11, pp. 41053-41065, 2023.
[15] A. Karim et al., "Phishing Detection System Through Hybrid Machine Learning Based on URL," in IEEE Access, vol. 11, pp. 25425-25440, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY