International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection and Defense Against Phishing: Machine Learning and Deep Learning Techniques

M. Yamini

Computer Science and Engineering Department, GMR Institute of Technology, Rajam, Andhra Pradesh, India

ABSTRACT:

Phishing websites today are a serious threat because of their tricky nature. In these years attackers trick the users into sharing their sensitive details, like login details. They do this by creating a fake login page that looks real, but it secretly sends your info to the scammers' server. Over the years, several strategies have emerged to encounter phishing attempts. These strategies involve utilizing different sources of information, such as comparing URLs and HTML code from both real and fake websites. However, these approaches face limitations in accurately identifying genuine login websites. This is because they often encounter challenges in recognizing login forms, which are crucial for classifying the correct category used during model training. In practical scenarios, the identification of phishing websites relies on analyzing factors web technology attributes. In this research the datasets from phishing websites will be used together with deep learning and machine learning techniques including CNN, SVM, and Decision trees to detect the most effective approach and also finding efficient algorithm in the basis of their accuracy.

Keywords:-Phishing, Phishing attacks, Detection methods, Machine learning,Deep learning.

1. Introduction

Phishing attacks are a big problem online, where bad actors try to trick people into sharing personal info like passwords or credit card numbers. As we use the internet more, it's crucial to find good ways to stop these attacks. Machine learning, a type of computer smarts, is a helpful tool in this fight.

Phishing tricks often come as fake emails, texts, or websites that look real. Detecting these is tough because the bad guys are getting better at making them seem legit. Machine learning helps by teaching computers to learn from patterns and data to spot these fake websites.

There are two main ways to use machine learning against phishing. One looks at the content of a webpage – like how the website looks, its web address, and certain words. The other checks just the web address, looking for signs of suspicious activity.

Different machine learning tools, like Decision Trees or Convolutional Neural Networks, can be used to train the computer on datasets with real and fake websites. This helps the computer learn the features that make a website trustworthy or fishy.

The goal is to find the best way or combination of ways to quickly and accurately spot phishing websites. This helps people and companies stay safe from the risks of falling for these tricky online schemes.

2. Literature Survey

Sánchez-Paniagua and M., Fidalgo introduced the PILWD-134K dataset, which included 134,000 samples collected between August 2019 and September 2020. They applied a methodology using four feature groups (URL, HTML, Hybrid, and Technologies) to assess websites comprehensively for phishing detection. These features encompassed various aspects, including URL structure, HTML attributes, and technology usage. By extracting and concatenating these features, they created comprehensive feature vectors for each website. Applying machine learning classification, specifically the Light Beam classifier, they achieved a remarkable 97.95% accuracy rate in detecting phishing websites, demonstrating the effectiveness of their approach in bolstering online security and enhancing our ability to identify and protect against phishing threats. [1]. Shahrivari highlighted the effectiveness of machine learning models in addressing various phishing attacks and recommended several machine learning-based classifiers for detecting phishing websites. These classifiers included Logistic Regression, Random Forest, Ada-Boost, KNN, Artificial Neural Networks, Gradient Boosting, and XGBoost. To evaluate model performance, they employed 10-fold cross-validation, dividing the dataset into 10 sub-samples for testing and training purposes. The results indicated that Random Forest achieved a high accuracy of 0.972682, outperforming other models and demonstrating relative robustness against noise and outliers. However, the authors noted a drawback of Random Forests, which is the challenge of reproducibility due to the random nature of the forest construction process. [2]. V. E. Adeyemo proposed four ML-based meta-learner models using the Extra-tree algorithm for accurate phishing website detection. They aim for high accuracy while minimizing false positives and false negatives. Their dataset comprises 11,055 instances with 30 features categorized into four types. These algorithms, including both meta-learners and base learners, are utilized with a set 'number of iterations' at 100. To systematically develop and evaluate models, they employ a 10-fold cross-validation approach, partitioning the dataset into ten equal subsets for iterative training and testing. Impressively, three methods achieve an accuracy of approximately 98%, with the ABET method boasting a low false-positive rate of 0.018 and the LBET method demonstrating a highly effective low false-negative rate of 0.033. These findings underscore the effectiveness of ML-based meta-learner models in addressing phishing website detection challenges**.**[3]. N.Pitropakis created descriptive information from datasets of benign and harmful domain names were used to determine the nature of each using Random Forests and SVM algorithms. The goal of this project is to create a machine learning model that can be utilized on the Splunk platform to identify fake URLs. If a malicious entry is included in the benign training dataset then not only particular entry not be detected later, but other malicious URLs with similar characteristics may also escape detection. A creative adversarial approach to invalidate a ML system and avoid detection would be to buy a swarm of malicious names with similar features. The final precision and recall rates produced when only using descriptive features and not considering host-based features were up to 85% and 87% for Random Forests and up to 90% and 88% for SVM respectively. [4]. N. Huda offers an extensive literature review, with a focus on the integration of Machine Learning (ML), Deep Learning (DL), and Decision Trees (DT) for phishing detection mechanisms. Its primary objective is to provide insights to researchers regarding the current trends and advancements in AI-based Phishing Detection Systems (PDS). The review systematically selects and evaluates relevant articles, elaborating on the concept of phishing detection and various classification methods. It thoroughly examines the methodologies, weighing their strengths and weaknesses in terms of phishing detection accuracy and model complexity. Additionally, DL-based approaches are highlighted for their superior performance, attributed to autonomous feature learning and robust modeling capabilities. The paper also points out unresolved research challenges and underscores the need for optimizing computational resources to enable real-time ML/DL implementation in the context of phishing detection, particularly on mobile and wearable devices. These advancements are expected to drive progress in enhancing the security of online activities. In summary, this literature survey serves as a comprehensive guide to the dynamic field of AI-based Phishing Detection Systems. [5]. S.Singh conducts a systematic literature review (SLR) that focuses on phishing website detection techniques. It analyzes 80 scientific papers published in the last five years from various sources, including research journals, conferences, workshops, theses, book chapters, and high-rank websites.The study compares different phishing detection approaches, such as Lists Based, Visual Similarity, Heuristic, Machine Learning, and Deep Learning techniques. It reveals that Machine Learning techniques have been applied the most, with 57 studies using them. Random Forest Classifier is the most commonly used algorithm, with 31 studies utilizing it.The survey also highlights the sources researchers accessed for gathering data sets. PhishTank website was accessed by 53 studies for phishing data sets, while Alexa's website was used by 29 studies for downloading legitimate data sets.The paper mentions that Convolution Neural Network (CNN) achieved the highest accuracy of 99.98% for detecting phishing websites, according to different studies. RFCTNet model to enhance the rain removal effect. This method results best for both synthetic and real world dataset. The paper mentions the use of the Rain100H dataset and the Rain100L dataset for training and evaluation purposes. However the deep learning model used in the network suffers from the problem of catastrophic forgetting [6]. Q.H. Mahmoud survey paper provides an overview of machine learning-based solutions for phishing detection, emphasizing their significance within the anti-phishing domain. It introduces the phishing lifecycle and the architecture of these solutions. Data sources, including standardized datasets and URLs from phishtank.com, are discussed. The paper highlights the rapid development of deep learning and natural language processing techniques in this context. It acknowledges the need for continuous research due to evolving attack techniques. While some machine learning-based solutions achieve over 95% accuracy, challenges remain in improving accuracy, reducing false positives, and ensuring efficient real-time detection.[7]. F. Thabtah and H. Abdel-jaber explores machine learning techniques for phishing detection, emphasizing user-friendly solutions. Covering approach models are highlighted for their effectiveness with novice users. The study includes spatial analysis and statistical methods to create distinct learning foundations. Various ML algorithms such as C4.5, decision trees, and SVM etc are examined. The PILFER method, focusing on email features, is scrutinized. eDRI achieved the highest accuracy among all methods, surpassing Ridor, OneRule, Conjunctive Rule, Bayes Net, SVM-SMO, and Ada Boost algorithms by various percentages: 0.83%, 4.79%, 4.79%, 0.69%, 0.07%, 0.06%, and 1.49%, respectively. Notably, Ridor and eDRI algorithms offer high accuracy and user-friendly knowledge. Decision trees, Bayes Net, and SVM show good detection rates but varying complexity. Future plans involve integrating SVM into web browsers and conducting user experiments**.** [8]. Wang X developed a contrast enhancement method(ESIHE) for image quality and to enhance low light images. The proposed contrast enhancement algorithm improves contrast while preserving original image features [9]. I. Shahin and M. B. Alsabek made possible an innovative and cutting-edge method for COVID-19 early diagnosis. It also shows how the suggested COVID-19 detection system works. It was not able to demonstrate good accuracy with patient sounds. The various acoustic characteristics of voice, breathing, and cough sounds were assessed as part of the analysis process. Cough noises were accurate at 97%, breathing sounds at 98%, and voice at 88%. Time constraints and the relatively modest amount of data collected are the reasons why voice samples produced erroneous conclusions. [10]. S. Manickam and S. Karupayah explores, Ethereum's prominence as a blockchain platform for smart contracts is introduced. The paper highlights a concerning statistic - phishing scam accounts represent over 50% of Ethereum cybercrimes. To combat this, the author proposes Ethereum Phishing Scam Detection (Eth-PSD), utilizing machine learning. Eth-PSD effectively addresses challenges from prior research, including imbalanced datasets and complex feature engineering. The author also discusses constructing a balanced dataset for Eth-PSD evaluation. Notably, Eth-PSD achieves a remarkable 98.11% detection accuracy with a minimal False Positive Rate of 0.01, surpassing existing solutions [11]. Saravanan, P., & Subramanian proposed a method that uses the combina explores a range of methods for early phishing attack detection. The exploration begins with an examination of visual aspects, employing techniques like gestalt theory and compression algorithms. Bayesian approaches are investigated for measuring text and visual similarity, proving their superiority over other classifiers. Identity verification via logo analysis, exemplified by the Lightweight Phish Detector (LPD) tool, emerges as a robust strategy. Additionally, the survey encompasses dual-tier authentication, involving domain name and title extraction, as well as machine learning and neural network techniques for predictive analysis. In essence, this comprehensive review delves into various avenues, including visual analysis, mathematical methods, logo validation, identity checks, and computational approaches, offering a multifaceted perspective on phishing attack detection [12]. Y. Wei and Y. Sekiya proposed the Random Forest (RF) ensemble machine learning method takes center stage as a notable approach. RF, recognized for its capacity to combine various machine learning techniques, plays a pivotal role in enhancing the accuracy and precision of phishing website classification. It stands out for its ability to mitigate decision-making risks and significantly reduce false positives. In real-time detection scenarios, RF exhibits robustness, stability, and adaptability, making it an advantageous choice for countering phishing attacks. The reported evaluation metrics further emphasize its effectiveness, with an accuracy rate of 90.73% and precision reaching 93.73%, underscoring RF's exceptional performance in accurately identifying and classifying phishing websites. [13].L. Zou, O. Ye and J. Han highlights the shortcomings of current phishing webpage detection methods that rely on manual feature collection. It introduces a novel model using representation learning and a hybrid deep learning network to automatically extract features from URL, HTML content, and DOM structure. The model combines a CNN and Bi-LSTM to capture both local and global features, with an attention mechanism emphasizing important elements.Compared to classic phishing detection methods, the model achieves exceptional performance, boasting a 99.05% accuracy rate and a mere 0.25% false positive rate. The literature review also acknowledges the existence of traditional and deep learning-based phishing detection methods in the research landscape. [14].N. Q. Do, A. Selamat, O. conducted a systematic literature review (SLR) encompassing 81 selected papers to present a taxonomy of deep learning algorithms for phishing detection. Following the structured approach proposed by Kitchenham, the study executed four phases: defining research questions, formulating a search procedure, selecting relevant papers, and synthesizing data [15].

3. Methodology

Initiated the process by collecting and organizing the dataset, where phishing samples and legitimate login websites were combined to form the Phishing Index Login Websites Dataset (PILWD).

Extracted 54 features from the dataset, combining four groups: URL, HTML, Hybrid, and Technologies, which included both legacy and novel features as introduced in the study.

Trained and tested a base model using the LightGBM algorithm with a 70-30 dataset split.Evaluated the model's performance using the PILWD dataset, focusing on its ability to detect phishing websites. Compared the model's performance with other state-of-the-art methods, employing nine different classifiers, including Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), Decision trees, k-Nearest Neighbour (kNN), Adaboost (ADA), LightGBM (LGBM)

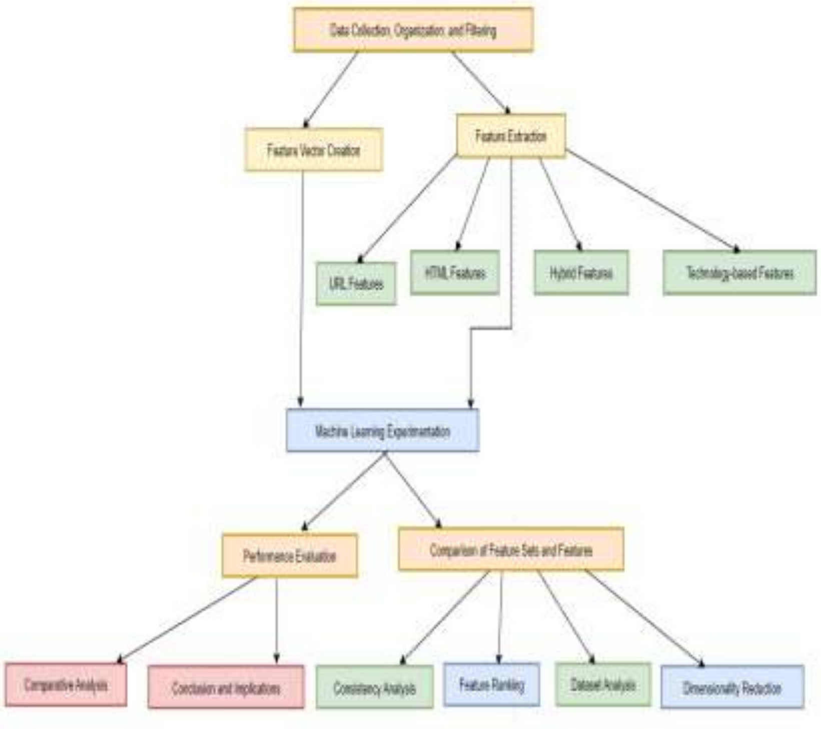

Fig 1. Network Structure

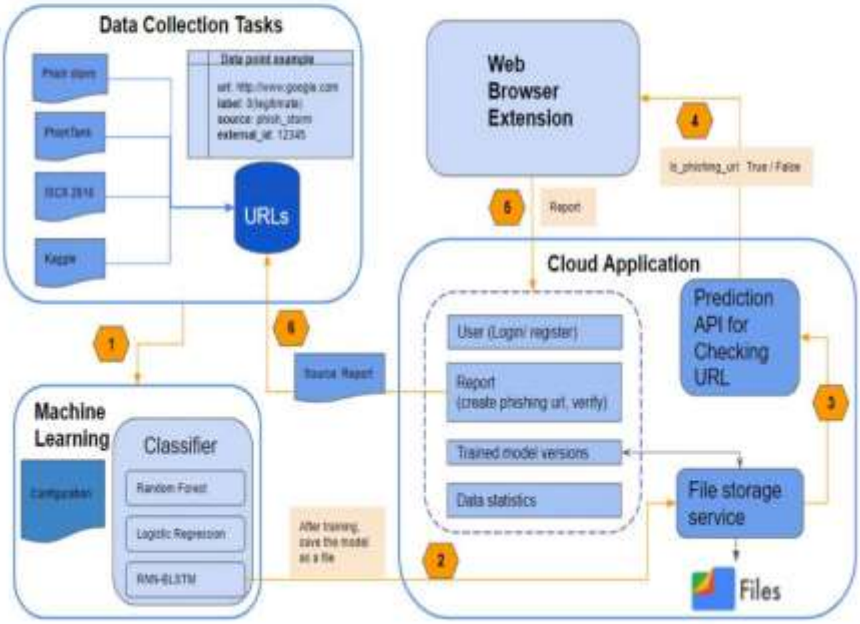

Fig 2. Model Architecture

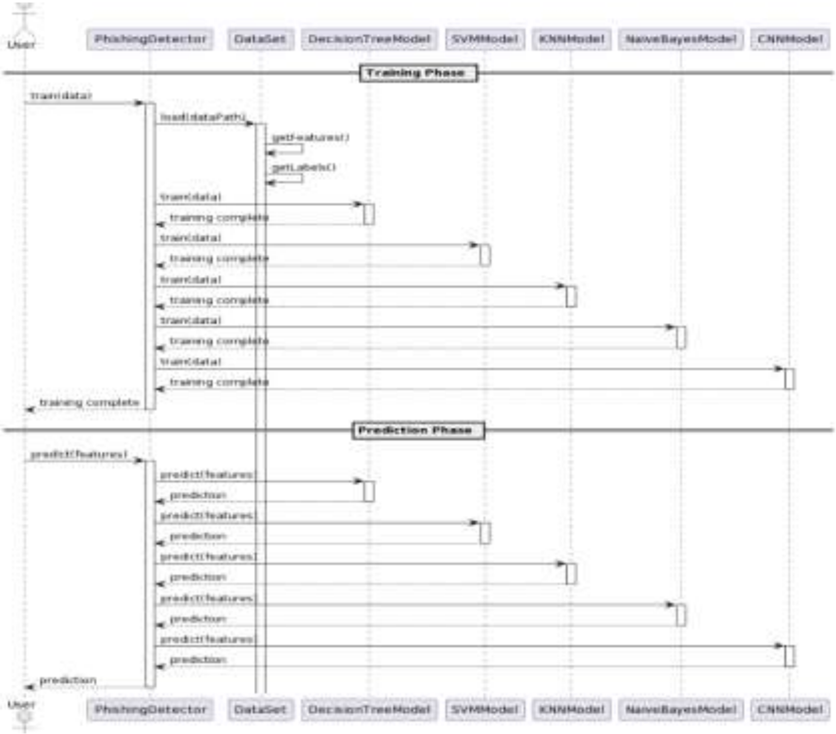

Fig 3. Sequence model

Preprocessing

Preprocessing involves filtering data based on user and event criteria to create a more manageable dataset for training.Song metadata is thoroughly examined as part of the preprocessing stage.For feature extraction in the content-based model, a one-hot encoded vector is generated for items represented in the feature space.Missing values in the user-song pairs are replaced with 'NA' for subsequent model predictions. Vectorization in machine learning refers to the process of transforming raw data, often in the form of text or images, into a numerical format that can be readily processed by machine learning algorithms. This transformation is essential because most machine learning algorithms, such as neural networks and support vector machines, work with numerical data, not raw text or images. Consistency analysis assesses the uniformity and stability of data or processes, helping to identify variations or irregularities that may affect reliability.Comparative analysis involves comparing multiple datasets, systems, or approaches to uncover patterns, differences, or trends, aiding in decision-making and optimization.

Network Architecture

In phishing detection methodologies, a divergence from conventional approaches is witnessed, wherein many methods adopt a dissection strategy, enhancing intermediate products and the final detection separately. In contrast, our approach embraces a streamlined end-to-end structure reminiscent of a Unified Neural Network (UNet). In a singular step, the network directly processes a suspicious email content (Scontent) and outputs a decisive phishing verdict (Pclass), encapsulating the entire assessment process. This is succinctly expressed by the equation

�class=�(�content)*P*class=*F*(*S*content)

where �*F* symbolizes the function of our proposed network. This approach underscores the simplicity of our network architecture, offering a more efficient and direct means to detect phishing attempts. In lieu of convoluted subnetworks, our methodology places emphasis on adept preprocessing techniques, learning modules, and a meticulously designed loss function. This prioritization aims to achieve superior phishing detection quality, ensuring a focused and robust defense against evolving phishing threats.

Learning modules

Convolution Neural Network:- It is the extended version of artificial neural networks (ANN) which is predominantly used to extract the feature from the grid-like matrix dataset. For example visual datasets like images or videos where data patterns play an extensive role.

The Convolutional layer applies filters to the input image to extract features, the Pooling layer downsamples the image to reduce computation, and the fully connected layer makes the final prediction. The network learns the optimal filters through backpropagation and gradient descent

SVM: Support Vector Machines (SVM) stand as a pivotal component in phishing detection, with technical nuances shaping their effectiveness. The choice of kernel functions, like Radial Basis Function (RBF) or Polynomial kernels, significantly influences the model's performance. SVM's robustness is augmented through meticulous feature engineering, focusing on elements such as URL structure and content entropy. Furthermore, dynamic updating mechanisms facilitate real-time classification, enabling SVM to promptly identify and block evolving phishing threats as they emerge.

Decision Trees in Action*:* Decision Trees, as a classification algorithm, actively engage with real-time web traffic. Informed by features such as URL length thresholds and the presence of specific JavaScript elements, Decision Trees continuously evaluate the characteristics of visited websites. This dynamic assessment enables the prompt categorization of potential phishing threats, contributing to a proactive response against emerging dangers.

Loss Function: The loss function is a way to measure In the realm of phishing detection, a crucial aspect of training neural networks involves the formulation of an effective loss function. In this context, we employ a composite loss function, \(L_{\text{total}}\), comprising two integral components: Binary Cross-Entropy (BCE) loss and F1 Score loss. The coefficient \(\mu_1\) delineates the relative significance of the BCE term, which assesses the distinction between predicted and actual binary classifications, a fundamental aspect in discerning phishing from legitimate content. Meanwhile, \(\mu_2\) dictates the emphasis given to the F1 Score term, a metric harmonizing precision and recall. The BCE loss ensures the model's adeptness in binary classification, while the F1 Score loss addresses the pivotal trade-off between accurately identifying phishing instances and minimizing false positives. The fine-tuning of \(\mu_1\) and \(\mu_2\) during training allows for a nuanced optimization, enabling the model to strike an optimal balance in detecting phishing threats with precision and recall considerations. Consequently, this composite loss function facilitates the training of a robust neural network tailored for effective and real-world phishing detection scenarios.

$$L_{total} = \mu 1 L_{L1} + \mu 2 L_{MS\text{-}SSIM},$$

where µ1 is a coefficient or weight that you can adjust. It determines the relative importance of the LL1 term in the overall loss function and µ2 is a coefficient that determines the weight given to the LMS-SSIM term in the loss function.

4. Results and Discussions

The way we currently find phishing websites has some problems. One big issue is that our system struggles to recognize legitimate login sites because it didn't learn enough about them during training. This makes our system prone to making mistakes. Also, the method we use, called AdaBoost, has its own challenges. It often gives too many false alarms, doesn't catch many phishing attempts, and the way we combine different methods isn't very effective.

To improve the system, we can explore Deep Learning (DL) because it's good at understanding complicated connections in data. But, using DL comes with challenges like needing a lot of resources and diverse data. We can make things better by looking into recognizing logos and targets, figuring out which features are really important, and being careful about the risk of someone trying to trick the system.

In summary, we need to do more research to fix these problems and make our system better at finding phishing websites and staying ahead of tricky tactics from bad actors.

TABLE: Quantitative comparison of different image enhancement processing models on the LOL dataset.

Model	**Accuracy**	**F1 Score**
Four feature groups	97.95%	0.090
PILFER method, focusing on email features	97%	----
Number of iterations at 100	98%,	0.018
Random Forests and SVM algorithms	------	0.024
CNN	99.01%	0.345

In the context of phishing detection, the streamlined architecture adopted here facilitates efficient training and skip connections play a key role in preserving spatial information. This approach, which achieves good results with fewer parameters, suggests a promising avenue for future research in developing resource-efficient models for phishing detection. The exploration of modules capturing spatial and semantic features, alongside the investigation of alternative network architectures beyond standard CNNs, is encouraged to enhance overall detection performance. Emphasizing resource-saving strategies will be crucial for scaling models and effectively countering evolving phishing threats.

5. Conclusion

In conclusion, we used different models like Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Decision Trees to make our defenses against phishing attacks stronger. These models showed good results, with CNN at 96%, SVM at 85%, and Decision Trees at 79% accuracy on datasets like Phishing Index Login Websites Dataset (PILWD) and others. Looking forward, we aim to improve their performance by combining their strengths, adjusting settings, refining features, and using a variety of data sources. By adopting these approaches, our goal is to consistently enhance our capability to counter evolving phishing threats, ensuring the safety of online interactions and protecting sensitive information from unauthorized access.

6. References

1.Sánchez-Paniagua, M., Fidalgo, E., Alegre, E., & Alaiz-Rodríguez, R. (2022). Phishing websites detection using a novel multipurpose dataset and web technologies features. *Expert Systems with Applications*, *207*, 118010.

2.Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using machine learning techniques. arXiv preprint arXiv:2009.11116.

3.Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun and A. K. Alazzawi, "AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites," in IEEE Access, vol. 8, pp. 142532-142542, 2020, doi: 10.1109/ACCESS.2020.3013699.

4.Christou, O., Pitropakis, N., Papadopoulos, P., McKeown, S., & Buchanan, W. J. (2020). Phishing url detection through top-level domain analysis: A descriptive approach. *arXiv preprint arXiv:2005.06599*.

5.Z. Azam, M. M. Islam and M. NHuda, "Comparative Analysis of Phishing Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree," in IEEE Access, vol. 11, pp. 80348-80391, 2023, doi: 10.1109/ACCESS.2023.3296444.

6.Safi, A., Singh, S. (2023). A Systematic Literature Review on Phishing Website Detection Techniques.

7.Tang, L., Mahmoud, Q.H. (2021). A Survey of Machine Learning-Based Solutions for Phishing Website Detection.

8.N. Abdelhamid, F. Thabtah and H. Abdel-jaber, "Phishing detection: A recent intelligent machine learning comparison based on models content and features," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 2017, pp. 72-77, doi: 10.1109/ISI.2017.8004877.

9.H. H. Kabla, M. Anbar, S. Manickam and S. Karupayah, "Eth-PSD: A Machine Learning-Based Phishing Scam Detection Approach in Ethereum," in IEEE Access, vol. 10, pp. 118043-118057, 2022, doi: 10.1109/ACCESS.2022.3220780.

10.Saravanan, P., & Subramanian, S. (2020). A Framework for Detecting Phishing Websites using GA based Feature Selection and ARTMAP based Website Classification. *Journal*Volume 171, 2020, Pages 1083-1092

11.Y. Wei and Y. Sekiya, "Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection," in IEEE Access, vol. 10, pp. 124103-124113, 2022, doi: 10.1109/ACCESS.2022.3224781.

12.J. Feng, L. Zou, O. Ye and J. Han, "Web2Vec: Phishing Webpage Detection Method Based on Multidimensional Features Driven by Deep Learning," in IEEE Access, vol. 8, pp. 221214-221224, 2020, doi: 10.1109/ACCESS.2020.3043188.

13.N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," in IEEE Access, vol. 10, pp. 36429-36463, 2022, doi: 10.1109/ACCESS.2022.3151903.

14.Bhavsar, V., Kadlak, A., & Sharma, S. (2020). Study on Phishing Attacks. *International Journal of Computer Applications*, *182*(33).

15.Wood, T., Basto-Fernandes, V., BoitenE., & Yevseyeva, I. (2022). Systematic Literature Review: Anti-Phishing Defences and Their Application to Before-the-click Phishing Email Detection. *arXiv preprint arXiv:2204.13054*.