Phishing detection using Machine learning and Deep learning techniques: A review

Maral Saleh[1*], Seda Şahin[2]

[1]Engineering Faculty, Kirkuk University, Street, Kirkuk, Iraq, Email: maralismail92@gmail
[2]Engineering Faculty, University, Uluyazı, Çankırı, Türkiye
* Corresponding Author

Received: 10.07.2024 Revised: 13.08.2024 Accepted: 07.09.2024

ABSTRACT
Cyber criminals have not ceased to develop new ways of attacking users of the cyberspace, this is sees in the continuous style of phishing attacks. This paper aims at providing a comprehensive overview of the latest state-of-art in the development and implementation of phishing detection system with particular emphasis on the use of machine learning and deep learning methods. The review includes a wide variety of studies, comparing the characteristics of various system architectures, algorithms, applications, and implementation tools used in the development of recommender systems. Discovery includes massive trends in development of new combined deep learning models, an identification of feature engineering and data preprocessing as significant steps in developing DL models, and identification of brand-new approaches to ease the learning of deep learning such as visual similarity analysis, and semantic analysis of texts. The review also notes that heightened user awareness and education helps make users who are targeted by phishing more resilient while pointing out that the benefits of blockchain technology in improving cyberspace security and meaningfully advancing cooperation continue to remain enormous. This paper offers an extensive assessment of the existing literature on phishing detection, as well as sheds light on possible future research directions that are more effective in fight against this increasingly looming menace.

Keyword: Phishing Detection, Machine Learning, Deep Learning, Cybersecurity, Feature Engineering

1. INTRODUCTION

Phishing stands as the most obvious threat to the web since its creation, remaining a constant risk due to growing dependence on technology. It involves a cybercrime that deceives users so as to gain confidential data bearing information about credit cards, bank accounts, and account logins. The phishing email is enhanced via fraudulent technologies, including forming fake sites or mails that appear legitimate and request for details from users regarding user login procedures [Almomani et al., 2013]. Even as various initiatives to create anti-phishing and enhance the security of web browsers are underway, such measures have had limited success, and phishing still remains a great danger both to individuals and institutions.

The growth of phishing which was noted earlier is partly due to the increasing adoption of new technology and the internet space monetization. With many businesses including individual clients engaging in online shopping and saving most of their personal details online, the chances of falling victims of phishing attacks also escalate. Not to mention, criminals are always coming up with new ways to stay undercover and misuse existing systems even when security measures have been imposed [Patel et al., 2022].

Evidence suggests traditional types of phishing have increased, and they are now surpassing what was witnessed back in as far as older phishing tactics. For example, the Anti-Phishing Working Group (APWG) noted in its 2010 Phishing Activity Trends Report that this trend was already noticeable when attacks targeting both individuals and businesses started increasing [APWG 2010]. Such statistics underline the necessity for other complementary or standalone phishing detection and preventative measures.

More recently, more attention has been oriented to incorporating supporting computational intelligence methods, like machine learning (ML) or deep learning (DL), to enhance existing phishing detection and classification systems. These methods make use of the capacity of algorithms to extract useful and engaging data which is capable of revealing the fishing intent. For instance, Jain and Richariya [Jain and Richariya 2011] suggested that phishing detection mechanisms could be incorporated in the web browsers to improve the security of the users from the attacks. On the other side, Shahrivari et al.

[Shahrivari et al. 2020] examined and compared some machine learning algorithms used for phishing detection and within the paper, the efficacy of these technologies was demonstrated.
Phishing attacks are one of the online threats for which deep learning, a form of machine learning, has considerable potential. Arfus and co-authors [Ariyadasa et al. 2020] introduced a hybrid convolutional neural network & LSTM model which can be used for the detection of phish at very high accuracy. Because of this advanced mathematical background, these models have the capability of capturing and modeling nonlinear, dynamic relationships and complex data patterns and features that cannot be easily achieved with conventional methods.
In addition, it has been reported that there have been efforts in phishing detection systems focused on the adoption of data mining approaches. Lungu and Tbus¸c [Lungu and Tbus¸c 2010] suggested the creation of versatile systems for phishing detection incorporating information retrieval and data analysis from trusted web resources. Alanezi [Alanezi 2021] presented a systematic review of the existing online phishing detection systems and pointed out the possible use of machine learning to address the improvement of their effectiveness.
The textual content of the phishing attempts has also undergone a Natural Language Processing (NLP) approach. An advanced work to implement NLP based text identification and semantic analysis for phishing detection was done by Peng et al. [Peng et al. 2018], they proved their work by detecting phishing emails. Dey [Dey 2020] analyzed the effectiveness of employing both a neural network and PCA in selecting the features used in the solution of the phishing detection problem through the evaluation of two models based on accuracy and the time required for computation.
Other specific research contributions include Zamir et al.'s [Zamir et al. 2020] approach to identifying stacking patterns in phishing websites' misuse and Khan et al's [Khan et al. 2020] review of multiple datasets for the machine learning algorithms of phishing detection. Rashid et al. [Rashid et al. 2020] addressed machine learning based phishing detection methods and Suryan et al. [Suryan et al. 2020] proposed a learning model for detecting phishing web sites using thirty features. Another paper by Alam et al., [Alam et al. 2020] worked on a technique in a Machine Learning algorithm like Random Forest, Decision Tree, and PCA to detect phishing.
Apart from the above stated research, there has been progress in the training algorithms of deep learning. For example, the Levenberg-Marquardt (LM) algorithm is employed as an enhanced option to back-propagation (BP) algorithm used for the training of artificial neural networks (ANNs). Ibrahim et al [Ibrahim et al. 2017] have made a comparison between the BP ad LM algorithm in the process of defining the right structure of ANNs in different classification problems. According to them, their study showed that the accuracy and the convergence speed of the BP algorithm was better.
Also, the combination of using the blockchain and smart contracts in multi-cloud systems for security purposes has demonstrated an ability to mitigate security issues such as DDoS attacks. In their study, Hamodi et al. [Hamodi et al. 2022] developed a layer for using machine learning and smart contract to detect and response to DDoS attacks adopting apop blockchain-based architecture. Their work builds upon this feature to provide high-security features to all the stakeholders and enable a decentralized system.
The goal of this study is to present a comprehensive survey of methods based on machine learning and, in particular, deep learning, developed for the purpose of phishing detection. As a result of critically evaluating those approaches, it is possible to uncover promising tracks for further research and development in this essential field of cybersecurity.

2. Review Process

This section details the methodology that was followed in this review of different phishing detection techniques. The paper explains how papers were included, further selected, and analyzed in detail with regard to the system architecture, algorithms involved, application categories, and implementation tools used within the reviewed literature.

2.1 Inclusion Criteria

The review sought research work about the detection and categorization of phishing attacks. Papers to be selected for inclusion were included based on the following:

- **Focus on Phishing Detection:** The research work should be concerned with phishing detection problems, such as phishing website detection, phishing email detection, and other related online fraud issues.
- **Methodology and Approach:** The methodologies and approaches used in developing systems for phishing detection must be clearly identified; this will involve information about algorithms, datasets, and metrics that were used to evaluate the performance.

- **Technical Depth:** These papers are at a level that allows the understanding of the principles and the mechanisms involved in the proposed solutions, including the feature extraction information, classification techniques, and systems architecture.
- **Diversity of Techniques:** The aim here has been to cover, in this review, a wide spectrum of phishing detection techniques. Thus, it has included all those techniques based on machine learning, deep learning, data mining, and other relevant techniques.

These include several research papers, some review articles that have targeted discussions on specific sub-areas in phishing detection, and book chapters discussing broader concepts of cybersecurity and machine learning.

2.2 Classification of Papers

The surveyed papers have been classified into distinct groups according to their main theme and the type of phishing detection system being targeted. The main classes include the following:

2.2.1 System Structure Approaches

This class includes those contributions where work has been carried out to formulate a complete system for detecting as well as categorizing any phishing attempt. The general architecture of such systems consists of implementing a number of methods and components together to detect and thwart phishing attacks.

- **Integration with Web Browser:** Modern web browsers are used as the interface for accessing most online applications; hence, some researchers have tried to integrate phishing detection mechanisms within browsers themselves. Jain and Richariya [Jain and Richariya 2011] proposed a phishing detection technique integrated with a web browser for real-time protection of users.
- **AI-Powered Systems:** In order to have such a sophisticated phishing detection system that performs deep analysis of complex patterns of data in order to find these subtle indications of malicious activity, fast development related to artificial intelligence, particularly machine learning and deep learning, is required. Many of them use AI structures for real-time detection and response. Osho et al. [Osho et al. 2019] compared 35 deep learning algorithms for classification, highlighting those with the highest accuracy in phishing detection.
- **Data Mining and Feature Selection:** The phishing detection system is effective if the system could identify relevant features that will distinguish malicious websites and emails from legitimate ones. Most of the time, data mining techniques are used like clustering to extract key features and reduce the dimensionality of the data. Dey [Dey 2020] used data mining techniques with clustering to select essential features to improve the efficiency of their proposed phishing detection system.
- **Visual Similarity Analysis:** Most phishing websites attempt to look similar to their targeted legitimate websites from a visual point of view. Jain and Gupta [Jain and Gupta 2017] developed a method of phishing detection based on the visual similarity analysis between websites. Their approach takes advantage of the technique of computer vision, which identifies suspicious similarities in the website design and layout.
- **Hybrid Deep Learning Models:** Many research works explored the integration of various deep learning models into hybrid architectures to enhance the accuracy and robustness of the models. Ariyadasa et al. [Ariyadasa et al. 2020] proposed a hybrid deep learning model for the detection of phishing by effectively integrating CNNs with LSTM networks. While CNNs are efficient in the extraction of spatial features, LSTMs are much stronger in modeling temporal dependencies, and hence a combination of both would be immensely effective for the analysis of complex web pages and emails.
- **Data Analysis and Feature Engineering:** In general, the quality and nature of the data used to train the models are highly critical in determining their performance in machine learning-based phishing detection systems. Shahrivari et al. [Shahrivari et al. 2020] emphasized dataset description and analysis. For example, they showed that based on the presence of certain symbols in URLs, one can effectively classify phishing attempts.
- **Traditional vs. Non-Traditional Systems:** Alanezi [Alanezi 2021] classified phishing detection systems into traditional and non-traditional techniques. Traditional techniques normally depend on the user's awareness and education by providing guidelines and best practices that can help a user in identifying a phishing attempt. The non-traditional techniques depend on automated techniques such as machine learning, deep learning, and fuzzy rule-based systems for detecting phishing attacks without user intervention.
- **Semantic Analysis and NLP:** Semantic meaning understanding of text is an important aspect in spotting phishing emails that generally use social engineering methods. Peng et al. [Peng et al. 2018]

conduct the NLP based text analysis of email focusing most on the semantic features along with malicious command-item pairs identification. This explains that in finding out the phishing emails, NLP will give minor clues as in the form of subtle linguistic features.

- **Comparative Studies:** The studies conducted by the researchers will be able to compare the effectiveness of the various approaches. Dey [Dey 2020] compares two neural network-based models for phishing detection with and without PCA for feature selection, which might result in respective accuracy and computational time.

Table 1: Comparison between Models

Model	Dataset	Accuracy
Random Forest	UCI	97.3%
Decision Tree	Phishtank	89.40%
Heuristic Features	Phishtank	97%
Word Embedding	Antiphishing in China	90%
ANN	UCI	44.27%
ANN-PCA	UCI	54.27%
CNN-LSTM	Web Pages	96.20%
CNN	HTML Features	88.67%

This table presents a concise overview of the models discussed in the reviewed literature, highlighting their respective datasets and accuracy results.

2.2.2 Algorithms

This section will present selected algorithms implemented in the described systems and analyze their principles, advantages, and limitations in terms of phishing detection. These will range from traditional artificial intelligence to the state-of-the-art deep learning techniques.

- **Text Classification:** Jain and Richariya [Jain and Richariya 2011] They have made used of deep learning-based text classification on URL for detecting phishing where the structure of the URL to be parsed and matched with the suspicious structure. Feature extraction in this approach includes several items such as number of visible and hidden links, discrepancy between the visible URL and actual URL and so on to label a site as a phishing attempt.
- **Machine Learning Techniques:** Shahrivari et al. [Shahrivari et al. 2020] also explained a number of machine learning approaches reported to be used for the detection of phishing. The experimental methodologies include features like Supervised Learning algorithms which are Logistical Regression, Support Vector Machine, Decision Trees, AdaBoost, XGBoost, and Random Forest and more. The advantages of integrating multiple models were explained to work towards enhanced accuracy and model broadness.
- **Deep Learning Algorithms:** Due to the rise in research interest in deep learning, some powerful algorithms have emerged for analyzing complex data patterns, which is especially essential in phishing sites and emails. Osho et al. [Osho et al. 2019] did the review of the performance of different deep learning techniques for phishing detection, such as CNNs, Recurrent Neural Networks RNNs, and their hybrid variants. They also pointed out how this algorithm will easily adapt to the change in phishing tactics and can handle a wide range of data types.
- **Feature Selection and Dimensionality Reduction:** The feature selection algorithms pick the most informative features to provide the best phishing detection with an enhancement of efficiency and a reduction in computation complexity. Dey [Dey 2020] used the Principal Component Analysis (PCA) method to reduce the number of features for higher performance of the neural network-based phishing detection system.
- **Ensemble Methods and Stacking:** Model averaging techniques for example bagging and stacking improves the result of prediction by combining different models. Stacking and combining decision trees models was tested by Zamir et.al [Zamir et al. 2020] with Random Forest, Neural Networks and KNN among others.

2.2.3 Application Categories

Various categories of online applications and platforms can utilize the phishing detection system in order to protect users against various types of phishing attacks. The following section addresses categories of specific applications where research has been reviewed.

- **Weak Authentication Schemes:** Most phishing attacks are carried out when vulnerabilities are found in a weak authentication system. This sort of phishing attack is aimed at a web platform that has very poor security measures regarding authentication of the user. As Jain and Richariya [Jain and Richariya 2011] pointed out, strong authentication mechanisms can prevent the attacker from accessing any random user account.
- **Browser Vulnerabilities:** The exploitation of various vulnerabilities that might exist in web browsers is one of the usual tactics of phishing attackers. Researchers have focused on the development of techniques for the identification and mitigation of browser vulnerabilities while improving its security against phishing attacks.
- **Social Engineering:** The most frequent phishing attacks are those employing social engineering techniques that force the users to disclose credentials or to take some actions resulting in the compromise of their security. The identification and addressing of social engineering techniques used by phishing attackers were emphasized in the work of Shahrivari et al.
- **Link Manipulation:** Attackers in phishing usually tamper with URLs to redirect users to a malicious site. Researchers have developed various algorithms that analyze the structure of URLs for inconsistencies that, if subtle, might indicate a phishing attempt.
- **Filter Evasion:** The attackers keep developing techniques that will help them get away from security filters. The detection systems against phishing must change and evolve to act effectively against such evasive techniques.
- **HTML Feature Analysis:** Much information about probable phishing attempts can be known through the structure of the web pages in HTML. Ariyadasa et al. [Ariyadasa et al. 2020] discussed how to analyze various features of HTML by utilizing machine learning techniques to determine patterns that are indicative of phishing websites.
- **User Awareness and Education:** Most traditional phishing detection methods are based on user education regarding phishing risks and measures to identify suspicious emails and websites. The approach is based on the empowerment of users to make informed decisions and avoid falling prey to phishing attacks.
- **Intelligent Toolbars and Extensions:** There is the proposition of researchers to incorporate intelligent toolbars and extensions into web browsers for providing real-time phishing detection and warnings to the users. Such extensions can be used to analyze a website, URLs, and user interactions to find out possible threats and warn users.
- **Email Security:** It is among the most prevalent attack vectors as it usually contains an attachment with malware and/or phishing links in messages. For an incoming email, the phishing detection system can parse its content, structure, and sender, analyze whether it is a suspicious message for protection from such threats.
- **Blockchain-Based DDoS Mitigation:** The emerging technologies of blockchain and smart contracts have addressed security issues and have proved their potential for adaptation in rushing against DDOS attacks, which are considered an important aspect of phishing attackers to disrupt services and enable them to carry out their attacks Hamodi et al. Based on the use of the distributed and immutable properties of blockchain to enhance security and collaboration among various actors [Hamodi et al. 2022] proposed a conceptual approach of implementing blockchain architecture to detect and mitigate DDoS attacks.

2.2.4 Implementing Programs

This section covers the software tools and programming languages applied for the development and implementation of different reviewed systems of phishing detection. The choice of either is very much dependent on the applied techniques and the requirements of the system in view.

- **Secure Communication Protocols:** First, secure communication protocols like SSL and EV SSL have to be implemented so that the sensitive information of the users in a transaction over the Web could be safely guarded. Lungu and Tbus¸c [Lungu and Tbus¸c 2010] identified that there has to be more awareness among users to protect their security for doing online transactions.
- **Web Browser Integration:** Jain and Richariya [Jain and Richariya 2011] have implemented their phishing detection system using the C# programming language. To provide protection in real time, this system has been integrated with a web browser.
- **HTML and DOM Analysis Tools:** Specialized tools that can parse the structure and the content of web pages manipulate the Document Object Model are needed. Ariyadasa et al. [Ariyadasa et al. 2020] used various tools to analyze HTML pages, CSS, and DOM elements to extract features relevant for phishing detection.

- **OCR:** This technology picks up the text from an image that might be embedded in the web page. That can be useful in identifying phishing attempts that try to disguise malicious content within images.
- **Natural Language Processing Libraries:** These libraries avail the functionality and tools needed to process and analyze text data for the implementation of the phishing detection systems, such as the semantic analysis-based one. Peng et al. [Peng et al. 2018] resorted to the use of Stanford dependencies to analyze the textual content of emails for malicious command-item pairs.
- **Python Programming Language:** Python is one of the most known programming languages that have vast libraries and frameworks for machine learning and deep learning. Dey [Dey 2020] implemented their neural network-based phishing detection system using Python.

Therefore, the choice of incorporating different programs and tools reflects the great diversity of techniques and methodologies in developing systems for phishing detection. Such languages and tools provide support to the researcher in the analysis of data, modeling, and introducing effective solutions that cater to the increase in phishing attacks.

3. RESULTS

This review encompassed 14 research papers to identify the common patterns and trends of the phishing detection systems using the machine learning and deep learning methodologies. Emphasis has been given to various preprocessing techniques used in analyzing the HTML pages, the DOM structures, and the URLs, thereby underlining the use of hybrid models and original models like CNN, LSTMs, and Random Forests. Algorithms such as KNN, AdaBoost, ANN, NN, Gradient Boosting, XGBoost, MLP, Naïve Bayes, XCS, and CNNs variable convolutional layers are implemented.

3.1 System Structure Approaches

Most of the reviewed papers were based on more than one machine learning technique; further, the performance of different techniques was compared in order to get the best results. Similarly, various techniques in deep learning research were proposed, such as those that are feature-based and with modified preprocessing steps. Researchers also updated the architectures, tried different parameters, and combined different model structures in order to enhance previous models and propose new systems customized according to specific aims in research. More specifically.

3.2 Evaluation of Algorithms

Most of the surveyed studies employed machine learning and deep learning algorithms in detecting phishing. It is observed from Table 2 that clustering techniques have widely been used to find and extract the most effective features to classify accurately. Examples are as follows:

- **Jain and Richariya [Jain and Richariya 2011]:** This study work introduces a deep learning paradigm for classifying URL texts. It uses several parameters of the URL, for instance hidden links and the difference between the displayed and actual URL to detect a phishing attack.
- **Shahrivari et al. [Shahrivari et al. 2020]:** presented a survey of the machine learning method to detect phishing attacks. These papers indicated that machine learning algorithms will be able to adapt with the rapidity of the change in the phishing tactics and for handling different types of data. Many algorithms were compared, such as Logistic Regression, KNN, SVM, Decision Tree, AdaBoost, XGBoost, and Random Forest. Random Forest performed the best among all the models studied.
- **Osho et al. [Osho et al. 2019]:** It deals with the performance evaluation of deep learning algorithms for the purpose of phishing detection. The authors compared different deep learning methods, specifically CNN, RNN, and hybrid models, underlining the possibility of processing complex data patterns and detecting minimal signs of phishing.
- **Dey [Dey 2020]:** This paper proposes a neural network-based phishing detection system using PCA for feature selection so as to reduce the dimensions and make it more efficient. The authors used accuracy and computational time as criteria for the comparison between the models – with and without PCA.
- **Zamir et al. [Zamir et al. 2020]:** It presents, for the first time, research in the field of stacking for phishing detection. The authors used an ensemble of several machine learning methods such as Random Forest, Neural Networks, and KNN for feature improvement. This study also presented the advantageous points of using ensembles in different aspects: enhancing predictive accuracy and making models more robust.

3.3 Evaluation of Implementing Programs/Tools

Most of the programs used in the reviewed research were for data preparation tasks-cleaning, formatting, and features extraction. Others had special focuses, like the extraction of text from a URL and comparing that to a legitimate source. Some examples of such utilities and their functionality are as follows:

- **Secure Sockets Layer (SSL) and Extended Validation SSL (EV SSL):** These are protocols that secure users' sensitive information during an online purchase by encrypting the data so that it cannot be sniffed by any attacker. Lungu and Tbus¸c [Lungu and Tbus¸c 2010] gave emphasis on increasing security by promoting the necessary awareness among users for the adoption of the mentioned protocols.
- **Web Browser Integration and C#:** Jain and Richariya [Jain and Richariya 2011] built their phishing detection system in the C# programming language, integrating the solution directly into a web browser so that protection may be done in real time.
- **HTML/CSS/DOM Analysis Tools:** Ariyadasa et al. [Ariyadasa et al. 2020] used analyzers of HTML page, CSS stylesheet, and DOM structure to extract features relevant for the detection of phishing. Their work highlighted the necessity of structural and content analysis of web pages in order to identify suspicious patterns.
- **Optical Character Recognition (OCR):** With the use of this technology, it is able to extract text out of an image possibly embedded within a web page, revealing hidden phishing content.
- **Stanford Dependencies for NLP:** Peng et al. [Peng et al. 2018] utilized Stanford dependencies for analyzing textual content in emails, finding malicious command-item pairs by using NLP techniques.
- **Python and its Libraries:** Python was employed for developing the phishing detection systems because of the versatility of Python language and freely available rich set of libraries to support machine learning and deep learning. The neural network-based detection system Dey [Dey 2020] implemented for the classification of the said image is in Python.

Table 2: Comparison between Used Techniques for Phishing Detection and their Accuracy Results

Model	Accuracy
Logistic Regression	92%
Decision Tree	96%
Random Forest	97%
AdaBoost	93%
KNN	95%
Neural Network	96%
SVM sigmoid	82%
Gradient Boosting	94%
XGBoost	98%
Naïve Bayes	97.18%
CNN+RNN	97.9%
MLP	96.65%
XCS	98.39%
Random Forest with NL	97.98%
Monte Carlo	97.71%
LSTM - 1D Conv	92.79%
CNN	92.55%
LSTM	92.79%

This table provides a compact view of the various techniques used for phishing detection and their results in terms of reported accuracy. It shows that different machine learning and deep learning algorithms are receiving high accuracy rates in detecting phishing attempts.

This section consolidates the different approaches in the development of phishing detection systems by exploring novelty in the application of machine learning, deep learning, and data mining techniques. It has also highlighted how important feature engineering is, proper model selection, and the correct use of tools for model optimization.

4 CONCLUSION

This review has surveyed a wide variety of machine learning and deep learning techniques for phishing detection and demonstrated the innovative application of these methods in fighting this ever-evolving threat. In this regard, this analysis has underlined the following:

- **Algorithm Selection:** The chosen algorithms will determine largely the accuracy and efficiency of the Phishing Detection Systems. Classic machine learning algorithms such as Logistic Regression, SVMs, and Decision Trees, and advanced deep learning models including CNNs, RNNs, and hybrid architectures have been discussed. Other studies have shown that Random Forest and stacking have also produced good results on the classification accuracy.
- **Feature Engineering:** Filtering on the HTML and extracting relevant features on the websites, in emails and URLs are important for accurate identification of phishing. To extract features meaningful to decision making, the researchers have used data mining techniques such as clustering, as well as analyzing the patterns using visual similarity analyses techniques.
- **Data Preprocessing:** As it has been established, the training data in essence are the decisive factors which define quality of the model. Hence normalizing and feature scaling are major type of steps to be performed while preprocessing the data for making it fit for proper training of a model.
- **Tool Selection:** The researchers have employed different programs and languages when designing and applying the phishing detection systems. The availability of libraries to parse HTML, CSS or DOM structures and NLP tool to generate features from text and score semantic similarity have been most helpful for carrying out feature extraction and model building.
- **System Architecture:** Unfortunately, no universal model of a system that aims at detecting phishing sites exists due to differences in the various applications that use such a system. Some researchers have considered the idea of incorporating detection mechanisms within web browsers, others have work on systems using artificial intelligence for real-time detection and finally some have proposed using blockchain for both security and collaborative purposes.

REFERENCE

[1] A. ALmomani, B. B. Gupta, T.-C. Wan, A. Altaher, and S. Manickam, "Phishing dynamic evolving neural fuzzy framework for online detection 'Zero-day' phishing email," Indian Journal of Science and Technology, vol. 6, no. 1, pp. 1–5, 2013. doi: 10.17485/ijst/2013/v6i1.18.

[2] D. Patel, D. Patel, A. Patel, and J. Makadia, "All about phishing attack, danger and its prevention," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 10, no. VI, pp. 4908, Jun. 2022.

[3] Phishing website detection using neural network and PCA based on feature selection, International Journal of Recent Technology and Engineering (IJRTE), vol. 8, 2020.

[4] M. N. Alam, D. Sarma, F. F. Lima, I. Saha, and S. Hossain, "Phishing attacks detection using machine learning approach," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1173–1179, IEEE, 2020.

[5] M. Alanezi, "Phishing detection methods: A review," 2021.

[6] APWG et al., "Phishing activity trends report," http://www.antiphishing.org/APWG Phishing-Activity Report Jul 05.pdf, 2010.

[7] S. Ariyadasa, S. Fernando, and S. Fernando, "Detecting phishing attacks using a combined model of LSTM and CNN," International Journal of Advanced and Applied Sciences, vol. 7, no. 7, pp. 56–67, 2020.

[8] A. Jain and V. Richariya, "Implementing a web browser with phishing detection techniques," arXiv preprint arXiv:1110.0360, 2011.

[9] A. K. Jain and B. B. Gupta, "Phishing detection: analysis of visual similarity based approaches," Security and Communication Networks, 2017.

[10] S. A. Khan, W. Khan, and A. Hussain, "Phishing attacks and websites classification using machine learning and multiple datasets (a comparative analysis)," in International Conference on Intelligent Computing, pp. 301–313, Springer, 2020.

[11] I. Lungu and A. Tbuşc, "Optimizing anti-phishing solutions based on user awareness, education and the use of the latest web security solutions," Informatica Economica, vol. 14, no. 2, 2010.

[12] O. Osho, A. Oluyomi, S. Misra, R. Ahuja, R. Damasevicius, and R. Maskeliunas, "Comparative evaluation of techniques for detection of phishing URLs," in International Conference on Applied Informatics, pp. 385–394, Springer, 2019.

[13] T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in 2018 IEEE 12th International Conference on Semantic Computing (ICSC), pp. 300–301, IEEE, 2018.

[14] J. Rashid, T. Mahmood, M. W. Nisar, and T. Nazir, "Phishing detection using machine learning technique," in 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), pp. 43–46, IEEE, 2020.

[15] V. Shahrivari, M. M. Darabi, and M. Izadi, "Phishing detection using machine learning techniques," arXiv preprint arXiv:2009.11116, 2020.
[16] A. Suryan, C. Kumar, M. Mehta, R. Juneja, and A. Sinha, "Learning model for phishing website detection," EAI Endorsed Transactions on Scalable Information Systems, vol. 7, no. 27, pp. e6–e6, 2020.
[17] J. Yearwood, D. Webb, L. Ma, P. Vamplew, B. Ofoghi, and A. Kelarev, "Applying clustering and ensemble clustering approaches to phishing profiling," in Proceedings of the Eighth Australasian Data Mining Conference-Volume 101, pp. 25–34, 2009.
[18] A. Zamir, H. U. Khan, T. Iqbal, N. Yousaf, F. Aslam, A. Anjum, and M. Hamdani, "Phishing web site detection using diverse machine learning algorithms," The Electronic Library, vol. 38, no. 1, pp. 65–80, 2020.
[19] M. H. Ibrahim, K. H. Jihad, and L. L. Kamal, "Determining optimum structure for artificial neural network and comparison between back-propagation and Levenberg-Marquardt training algorithms," International Journal of Engineering Science and Computing, vol. 7, no. 9, pp. 14887–14890, Sep. 2017.
[20] Y. I. Aljanabi, A. A. Majeed, K. H. Jihad, and B. A. Qader, "Detect and mitigate blockchain-based DDoS attacks using machine learning and smart contracts," Informatica, vol. 46, no. 7, pp. 55–62, 2022. doi: 10.31449/inf.v46i7.4033.