Leveraging Machine Learning and Deep Learning for Phishing Website Prediction

Yoon-Teck Bau1, a) and Yuet-Xin Lam1, b)

1Centre for Advanced Analytics, CoE for Artificial Intelligence, Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysiab) Corresponding author: yuetxinflhcpm91@gmail.com  
a)ytbau@mmu.edu.my

**Abstract.** In this digital age, rapid technological advancements have changed the way we live. However, they have also introduced a big danger posed as a threat in the cyber world, one of the most prominent being phishing. Phishing websites are designed to imitate the real site that the average user may fail to recognize the deception. These fraudulent websites manipulate users into unknowingly providing sensitive information, putting their personal data at risk. As algorithms keep evolving regarding phishing aspects, traditional rule-based and heuristic predictions are becoming less effective. This research aims to use four machine learning algorithms to predict phishing websites. These algorithms are decision trees, random forests, artificial neural networks, and convolutional neural networks. The widely available dataset named phishing websites is from UCI online machine learning repository that has been emphasized on thorough data preprocessing, data analysis and algorithms evaluation. Data preprocessing makes sure the dataset contains no null and duplicated values. Data analysis shows what are two important feature predictors. All the algorithms will be evaluated using four metrics which are accuracy score, precision, recall and F1-score hence making comparison in determining the most effective algorithm. From the experimental results, random forest performed the best among four algorithms achieving 96.31% accuracy, 95.96% precision, 97.66% recall, and a 96.80% F1-score. Notably, all algorithms performed more than 90% accuracy.

# INTRODUCTION

Cyber threats have become a serious issue in this digital age. Phishing attacks are one of the common cybersecurity issues among technology users that can have a devastating impact on individuals and businesses. Phishing websites are designed by malicious actors for them to get financial gain or illegal purposes. Scammers create false websites or false links that look like genuine ones which cause people to get tricked by responding with sensitive information. Users might be unaware of the sign of a phishing attempt and are falling victim to such scams.

However, cybersecurity measures need to constantly evolve and up-to-date because the threat landscape is rapidly changing with new attack algorithms. So that traditional security method may not be effective against modern cyberattacks. Machine learning is a powerful method that can help in traditional rule-based systems.

Even though cybersecurity tools are getting better, the prediction of phishing websites remains a challenging task. Attackers keep changing their tricks, and their websites often look very real. This makes it hard for old systems to catch them. Furthermore, many new websites generated daily make manual monitoring and analysis infeasible.

This research focuses on addressing the limitations of existing systems by implementing machine learning algorithms to predict phishing websites. The goal is to make systems that are correct and can learn to catch new types of phishing. By investigating and implementing machine learning approaches, this study aims to contribute to the development of a reliable, efficient system for predicting phishing websites, ultimately enhancing cybersecurity measures in an increasingly connected world.

# literature review

Phishing website prediction has been reviewed using different machine learning algorithms. These different approaches show different levels of success. However, despite these advancements, significant challenges remain in terms of dataset diversity and algorithms comparison. These problems match the research objectives in this study.

Adeyemi Onih [1] developed a phishing prediction algorithm using random forest and support vector machine. This was to improve classification results. Their study used the PhishTank dataset, achieving 94.2% accuracy with random forest and 92.8% accuracy with support vector machine. Nevertheless, the study observed that computer-based prediction methods, such as artificial neural network would require greater computational resources and vulnerable to overfitting. Alshingiti et al.[2] designed a deep learning based phishing prediction system using convolutional neural network (CNN), long short-term memory (LSTM), and a hybrid LSTM-CNN algorithm. Experimental results proved that CNN gives the maximum accuracy followed by LSTM-CNN and LSTM. In this characteristic, the study highlighted that LSTM and hybrid LSTM-CNN algorithms require longer time training and computation resources than CNN.

Bahaghighat et al.[3] analyzed dataset bias in phishing prediction algorithms and its impact on accuracy. Their study demonstrated that algorithms trained on globally sourced datasets may not generalize well to local phishing attacks due to cultural and linguistic differences. The dataset used in their study contained phishing websites from various countries, and the best performing algorithm, random forest, achieved 94.3% accuracy. Future research is needed to investigate localized datasets that account for regional variations in prediction algorithm of phishing websites. Hannousse and Yahiouche [4] focused on phishing website prediction by comparing random forest, decision tree, and support vector machine. Their study relied on a benchmark phishing dataset and demonstrated that random forest outperformed other algorithms with an accuracy of 92.1%. However, the authors noted that their dataset was highly based on specific website features.

Hussain et al.[5] implemented a hybrid deep learning approach combining convolutional neural network and long short-term memory to predict phishing websites. The algorithm was trained on a phishing dataset and achieved an accuracy of 96.5%. Despite high accuracy, their study noted that convolutional long short-term memory algorithms require significant computational power, and future work will focus on optimizing the algorithm. Jayaprakash et al.[6] tested gradient boosting decision trees for phishing prediction. Their study used a balanced dataset from Kaggle and attained 95.3% accuracy. The research indicated that gradient boosting decision trees algorithms require a lot of feature engineering, which may not work well in real-time, and proposed automating feature extraction using neural networks.

Karim et al.[7] developed an explainable artificial intelligence algorithm using random forest for phishing website prediction with Shapley values. Their study analyzed a large-scale phishing dataset from multiple sources, achieved 96.1% accuracy. The algorithm required extensive feature computation, prompting future research into reducing computational overhead while maintaining transparency. Majgave and Gavankar[8]explored the optimization of convolutional neural network algorithm for phishing prediction. Using a dataset comprising web links and hypertext markup language content, their study achieved an accuracy of 96.4%. However, the computational overhead of convolutional neural network models and using web links instead of website features were a major drawback, making them unsuitable.

Nadar et al.[9] implemented the prediction of phishing websites using multiple machine learning algorithms such as logistic regression, Adaboost, and gradient boost. They also used a hybrid stacking classifier algorithm for this purpose. They found that the stacking classifier achieved the highest accuracy of 85.6% for phishing website prediction, whereas other classifiers could not match its higher efficiency. Ogundairo and Broklyn[10] introduced a deep belief network for phishing website classification, comparing it with machine learning algorithms like naive Bayes and decision tree. Their experiments on the OpenPhish dataset yielded a deep belief network accuracy of 95.9%, outperforming non-neural network classifiers.

Omari[11] examined the effectiveness of artificial neural networks in phishing prediction. Their study used a dataset compiled from multiple phishing sources, with artificial neural networks achieving an accuracy of 91.2%. Future work is expected to explore convolutional neural networks deep learning based algorithm to enhance robustness of phishing prediction systems. Rabbi et al.[12] explored a hybrid random forest and naive Bayes model for phishing website prediction. The algorithm was evaluated on the University of California Irvine phishing dataset, where it reached 94.7% accuracy. Their findings revealed that hybrid algorithms increase computational complexity, and future research will focus on reducing processing time without sacrificing accuracy.

Rao et al.[13] investigated how phishing attacks affect vulnerable populations, such as the elderly and less technologically aware users. Their study used a phishing awareness dataset and applied naive Bayes and logistic regression algorithms, achieving a maximum accuracy of 89.4%. However, their work primarily focuses on awareness campaigns. Safi and Singh[14] emphasized the lack of benchmark datasets for phishing prediction, which affects the fair evaluation of different algorithms. Their study tested multiple machine learning algorithms, including random forest, decision trees, and support vector machine, using an imbalanced phishing dataset. The results indicated that gradient boosting achieved the highest accuracy of 95.7% but the algorithm suffered from bias due to dataset imbalance. Future research aims to develop standardized benchmarking datasets to ensure consistent algorithm evaluation and fair comparisons. The authors suggested enhancing prediction algorithms by incorporating deep learning based algorithm.

Siddiq et al.[15] implemented two phishing websites prediction algorithms using artificial neural networks and convolutional neural network. Their study utilized the UCI online machine learning repository dataset and achieved 94.8% accuracy using artificial neural networks and 93.6% using convolutional neural network, respectively. However, the authors noted that convolutional neural network training was time consuming and required significant computational resources. Zara et al.[16] implemented a graph based machine learning algorithm to predict phishing domains by analyzing domain name system query patterns. Their algorithm was trained on a custom built phishing domain dataset and achieved an accuracy of 94.5%. However, they noted that domain name system based prediction alone is insufficient against website features.

By reviewing these literature papers, it is evident that while machine learning has improved phishing prediction, existing research still faces limitations regarding dataset diversity and application of various algorithms [17]. This study seeks to address these gaps by acquiring a dataset that contains multiple data sources and comparing a broader range of machine learning and deep learning algorithms for phishing prediction.

# machine learning and deep learning algorithms

The broader comparison in this research provides a more comprehensive evaluation of phishing prediction algorithms. Addressing these gaps will further enhance phishing prediction results and improve cybersecurity resilience. This section is organized as follows, starting from dataset, followed by data preprocessing, data analysis, and four algorithms including both machine learning and deep learning. The machine learning models decision tree, random forest and artificial neural network were implemented using the Scikit-learn library, while the convolutional neural network model was developed using Keras library.

## Dataset

The acquired dataset is a table that includes details about both genuine and phishing websites. It was acquired from UCI online machine learning repository. One of the most well-known public repositories offers an enormous number of datasets that may be used to train machine learning algorithms. This dataset is collected mainly from PhishTank archive, MillerSmiles archive and Google searching operators.

The dataset used contains a total of 11,055 rows and 32 columns. It is a labeled dataset designed for supervised machine learning tasks, where the objective is to classify websites as either phishing (-1) or legitimate (1). Among the columns, 31 represent features that describe various characteristics of websites, while the “Result” column serves as the target variable. The features are highly relevant to predict phishing behavior, incorporating structural, content-based, and reputation-based attributes.

Key features of the phishing dataset that include the presence of IP addresses in web links (IP\_having\_IP\_Address), web link length (URL\_Length), the use of web link shortening services (Shortining\_Service), and the presence of special characters such as "@" (having\_At\_Symbol). Other features like double\_slash\_redirecting, Prefix\_Suffix, and having\_Sub\_Domain examine web link structures that may indicate phishing. The dataset also evaluates website security and reputation with attributes like SSLfinal\_State, Domain\_registeration\_length, age\_of\_domain, and web\_traffic. Additional indicators include the use of external favicons, server form handling issues (SFH), and backlink information (Links\_pointing\_to\_page). Features such as Page\_Rank, Google\_Index, and Statistical\_report assess the website's credibility based on search engine and statistical metrics.

## Dataset Preprocessing

Data preprocessing is an important activity at which you prepare the dataset for data analysis and training machine learning algorithms. The dataset has a number of features, none of them have missing and duplicated values. Thus, making it complete and free from the need for imputation. All features are integer-encoded (int64) with values such as -1, 0, and 1, to represent phishing, suspicious and legitimate websites, respectively.

## Dataset Analysis

Data analysis is the most essential step in understanding dataset for learning their structure, characteristics, and patterns that will aid machine learning algorithm development. All features are integer encoded (int64) and without missing values. The dataset has a total of 11,055 samples and 32 features. Each sample represents a single website, and the features are corresponding attributes by which the websites are classified as phishing or legitimate. Since the features in the dataset are big enough, it will support training machine learning algorithms, providing a good amount of data to learn patterns and relationships effectively. With 32 features, this dataset is going to capture a wide variety of attributes on websites, ensuring that the algorithm will make predictions with great confidence on the phishing websites. The Result attribute, which serves as the target feature, classifies websites as phishing (-1) or legitimate (1). It gives the fact that 44% are defined and tagged as phishing websites in this dataset, while 56% are defined as legitimate websites. It slightly imbalanced dataset and incorporated legitimate websites as the majority class. Nevertheless, a near to balanced distribution means that machine learning algorithms trained on this dataset will effectively learn from phishing and legitimate websites.

Among those highly positive correlated feature pairs is SSLfinal\_State (0.71) with the Result, indicating that sites with valid SSL certificates are more accessible to be termed legitimate. Likewise, URL\_of\_Anchor (0.69) is also highly correlated with the Result displays quite positive associations to highlight their predictive capacities toward legitimate websites. On the other side, some features, such as Domain\_registration\_length (-0.23) and Abnormal\_URL (-0.6), show negative correlations to the Result feature, indicating that shorter domain registration length and abnormal web link are linked to phishing websites.

## Decision Tree

A decision tree is an algorithm used widely in classification problems to show the potential outcomes. It tries to predict whether a given website is phishing or legitimate in the cybersecurity domain. A decision tree acts in a prescriptive way through decision-making processes using different tree like structures for making conclusions. For phishing prediction, it would take in characteristics of a website as inputs and would work by evaluating each feature and going towards the branches corresponding to the particular feature values for decision-making. This systematic way breaks up complex problems and displays an accurate output.

## Random Forest

Random forest follows an ensemble learning methodology, with several decision trees constructed during the training phase. The very last decision is based on the majority voting of predictions from all the trees. The algorithm is favored for its strength and accuracy, especially in complex classification tasks such as prediction of phishing websites. In this way, it reduces overfitting and increases generalization to unseen data. The random forest algorithm has always been recognized as a robust and accurate one for complex classification tasks.

## Artificial Neural Network

An artificial neural network is a machine learning algorithm inspired by the working of the human brain, consisting of interconnected neurons distinguished into three layers which are input layer, hidden layer, and output layer. In phishing website prediction, the artificial neural network processes the features to differentiate between phishing and legitimate sites. The network learns patterns whereby weights are adjusted to minimize error. Furthermore, an artificial neural network is powerful in capturing non-linear relationships and handling high-dimensional data.

## Convolutional Neural Network

A convolutional neural network is a deep learning algorithm that automatically extracts patterns from input data and so can work well in phishing website prediction in cybersecurity. While web link structures, domain features, and content features are being processed, the convolutional neural networks will identify very slight and complicated patterns that will distinguish phishing websites from legitimate ones. Convolutional neural networks have more than three layers for feature extraction, dimensionality reduction, and classification; hence, they can operate well with high-dimensional data and learn hierarchical features without any manual engineering.

# results and discussion

To evaluate the effectiveness of different machine learning and deep learning algorithms in predicting phishing websites, four algorithms were implemented and assessed based on four key performance metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's prediction capabilities, especially in binary classification problems where both false positives and false negatives are critical. Table 1 shows performance comparison of four binary classification algorithms based on accuracy, precision, recall, and F1-score in percent.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TABLE 1.** Performance comparison of four binary classification algorithms based on accuracy, precision, recall, and F1-score in percent. The best results are in bold | | | | |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| Artificial Neural Network | 91.86 | 90.70 | 95.57 | 93.07 |
| Decision Tree | 95.30 | 95.49 | 96.33 | 95.91 |
| Convolutional Neural Network | 95.73 | **96.27** | 96.27 | 96.27 |
| Random Forest | **96.31** | 95.96 | **97.66** | **96.80** |

Artificial neural network had 91.86 percent accuracy, 90.70 percent precision, 95.57 percent recall, and 93.07 percent F1-score. It is the lowest among four algorithms for all four-performance metrics. This indicates the algorithm’s weakness in correctly identifying phishing websites.

Decision tree classifier achieved an accuracy of 95.30 percent, with a precision of 95.49 percent, a recall of 96.33 percent, and an F1-score of 95.91 percent. This indicates that the decision tree model was third effective in identifying phishing websites, offering a balance between correctly identifying true positives and minimizing false positives. However, decision trees are known to overfit on training data and may not generalize well if not tuned properly.

The random forest outperformed all other algorithms across all four-performance metrics. It achieved the highest accuracy of 96.31 percent, recall of 97.66 percent and an F1-score of 96.80 percent. Meanwhile random forest is placed second with the precision of 95.96 percent. Three highest performance metrics reflected by random forest indicates it is the strongest algorithm as overall performance in predicting phishing websites.

Figure 1 highlights the best algorithm which is random forest has been selected for further study to find high impact features. Two high impact features discovered are websites with SSL state (0.2933) and URL anchors (0.2434) that link to its own websites. The SSL state plays a crucial role, as legitimate websites typically use valid and trusted SSL state that denoted by "https", which ensure secure data transmission. In contrast, phishing websites often lack valid SSL certificates or use self-signed ones, making this a strong indicator of malicious intent. Similarly, URL anchors links within its own pages offer important insight. Legitimate websites usually have internal anchors that link to their own domain. However, phishing websites often include anchors that redirect users to external or unrelated domains, a common tactic used to deceive users and collect sensitive information. These two features significantly enhance the algorithm’s ability to distinguish between legitimate and phishing websites.

A graph with numbers and a curve

AI-generated content may be incorrect.

**Figure 1.** High impact features in predicting phishing websites using random forest

Deep learning based convolutional neural network achieved an accuracy of 95.73 percent, precision of 96.27 percent, recall of 96.27 percent, and F1-score of 96.27 percent. Overall, convolutional neural network ranked second position because it achieved second highest positions in two performance metrics which are accuracy and F1-score metrics. In addition, convolutional neural network achieved the highest performance for precision metric. Convolutional neural network required more computational power and longer training times which can be the factors in producing better results for predicting phishing websites.

Based on the evaluation results obtained, random forest algorithm demonstrated the most superior performance across multiple performance metrics. In terms of accuracy, recall and F1-score, random forest outperforms decision tree and both artificial neural network and deep learning i.e. convolutional neural network. Random forest is the most superior in accuracy to indicate the best overall prediction correctness due to its ensemble approach which combines multiple decision trees to reduce variance and improve generalization. This makes random forest less prone to overfitting compared to decision tree and neural network models. Additionally, random forest’s recall score highlights its ability to capture nearly all positive cases and reduce false negatives. The highest F1-score further confirms that random forest is balance in performance for both recall and precision, even though convolutional neural network having the best precision but falls short in recall, suggesting convolutional neural network may have more false negatives.

# CONCLUSION

This research began with a literature review focused on varying algorithms concentrating on the prediction of phishing websites. Major tasks in this study included data acquisition, data preprocessing, and data analysis to render the dataset ready for training and test the algorithms. Dataset is acquired from UCI online machine learning repository. Data preprocessing has been done to ensure no null values and no duplicated samples in the dataset. From data analysis, the number of samples between phishing websites and legitimate websites are quite balanced with legitimate websites forming a little more. Two high impact features that contribute to phishing prediction accuracy are identified which are https websites with secure sockets layer and internal links (anchors) embedded in a page point to the same domain as the website itself or not to suspicious external domains. This research has implemented four machine learning algorithms. Their performances were then evaluated in terms of accuracy, precision, recall, and F1-score. This will not only provide a comparison of effectiveness among the algorithms but also helps identify weak areas in predicting phishing websites. The experimental results show that random forest achieved the highest accuracy performance among four tested algorithms followed by convolutional neural network. Decision tree secured third position, while the artificial neural network showed the lowest performance among the four algorithms tested. For future work, this research can be extended. A bigger dataset for the deep learning algorithm which performed less well in this study, may further boost its performance.

# References

1. V. Adeyemi Onih, “Phishing Detection Using Machine Learning: A Model Development and Integration,” IJSMR **07**(04), 27–63 (2024).
2. Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q.E.U. Haq, K. Saleem, and M.H. Faheem, “A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN,” Electronics **12**(1), 232 (2023).
3. M. Bahaghighat, M. Ghasemi, and F. Ozen, “A high-accuracy phishing website detection method based on machine learning,” Journal of Information Security and Applications **77**, 103553 (2023).
4. A. Hannousse, and S. Yahiouche, “Towards benchmark datasets for machine learning based website phishing detection: An experimental study,” Engineering Applications of Artificial Intelligence **104**, 104347 (2021).
5. M. Hussain, C. Cheng, R. Xu, and M. Afzal, “CNN-Fusion: An effective and lightweight phishing detection method based on multi-variant ConvNet,” Information Sciences **631**, 328–345 (2023).
6. R. Jayaprakash, K. Natarajan, J.A. Daniel, C.V. Chinnappan, J. Giri, H. Qin, and S. Mallik, “Heuristic machine learning approaches for identifying phishing threats across web and email platforms,” Front. Artif. Intell. **7**, 1414122 (2024).
7. A. Karim, M. Shahroz, K. Mustofa, S.B. Belhaouari, and S.R.K. Joga, “Phishing Detection System Through Hybrid Machine Learning Based on URL,” IEEE Access **11**, 36805–36822 (2023).
8. A.B. Majgave, and N.L. Gavankar, “Automatic phishing website detection and prevention model using transformer deep belief network,” Computers & Security **147**, 104071 (2024).
9. V.K. Nadar, B. Patel, V. Devmane, and U. Bhave, “Detection of Phishing Websites Using Machine Learning Approach,” in *2021 2nd Global Conference for Advancement in Technology (GCAT)*, (IEEE, Bangalore, India, 2021), pp. 1–8.
10. O. Ogundairo and Peter Broklyn, *AI-Driven Phishing Detection Systems* (2024).
11. K. Omari, “Comparative Study of Machine Learning Algorithms for Phishing Website Detection,” IJACSA **14**(9), (2023).
12. Md.F. Rabbi, A.I. Champa, and M.F. Zibran, “Phishy? Detecting Phishing Emails Using Machine Learning and Natural Language Processing,” in *Software Engineering and Management: Theory and Application*, edited by R. Lee, (Springer Nature Switzerland, Cham, 2024), pp. 119–137.
13. R.S. Rao, A. Umarekar, and A.R. Pais, “Application of word embedding and machine learning in detecting phishing websites,” Telecommun Syst **79**(1), 33–45 (2022).
14. A. Safi, and S. Singh, “A systematic literature review on phishing website detection techniques,” Journal of King Saud University - Computer and Information Sciences **35**(2), 590–611 (2023).
15. Md.A.A. Siddiq, M. Arifuzzaman, and M.S. Islam, “Phishing Website Detection using Deep Learning,” in *Proceedings of the 2nd International Conference on Computing Advancements*, (ACM, Dhaka Bangladesh, 2022), pp. 83–88.
16. U. Zara, K. Ayyub, H. Ullah Khan, A. Daud, T. Alsahfi, and S. Gulzar Ahmad, “Phishing Website Detection Using Deep Learning Models,” IEEE Access **12**, 167072–167087 (2024).
17. M.A. Daniel, S.-C. Chong, L.-Y. Chong, and K.-K. Wee, “Optimising Phishing Detection: A Comparative Analysis of Machine Learning Methods with Feature Selection,” Journal of Informatics and Web Engineering **4**(1), 200–212 (2025).