# An Enhanced Deep Ensemble Learning Approach for Accurate Phishing Website Detection

International Conference on Cyber Warfare, Security & Space Computing
SpacSec 2026
Paper ID - 66

Jenefa J, Mary Anita E A, Vinodha D, Maria Lapina, and Vitalii Lapin

# Agenda

- Abstract
- Introduction
- Related Works
- Proposed Work
- Performance Analysis
- Conclusion
- References

# Abstract


- Phishing websites are a crucial threat that mimic legitimate websites to deceive people and gain access to sensitive information.
- Though there are many traditional approaches used to detect phishing attacks, it is difficult to keep up with the changes in patterns in phishing URLs.
- To learn more about these complex patterns, **deep learning approaches** are used.
- Hence, in this paper, we proposed an enhanced deep learning ensemble approach for accurate phishing website detection.
- An ensemble approach with four deep learning models: **Feedforward Neural Network (FNN), Deep Neural Network (DNN), Wide & Deep, and TabNet, is proposed,** which combines the strengths of individual models.

# Introduction

- Phishing attacks are one of the most common and crucial attacks in today's digital era.
- Deep learning techniques can handle complex and non-linear patterns efficiently, and hence they can be used in phishing detection.
- Structured features like URLs and metadata of domains can be processed to identify patterns that can be used for the identification of phishing websites.

# Related Works

- Adeyemo et al. [1] proposed an ensemble-based logistic model tree for phishing website detection.
- Innab et al. [2] compare the six machine learning models for phishing detection.
- Bhogesh et al. [3] proposed a comparative study on different machine learning approaches used for phishing detection.
- Setu et al. [18] proposed a phishing detection mechanism using a novel feature selection algorithm.
- Chiew et al. [19] proposed a hybrid feature selection approach using machine learning models for phishing detection.

# Proposed Work

- An ensemble approach of four deep learning models: **FNN, DNN, Wide & Deep, and TabNet** is proposed.
- The models are stacked and built using the **meta-classifier, XGBoost**, which integrates the predictions of these base models.
- The models are trained with the **top 20 most significant features**.
- These prominent features are extracted using permutation importance.
- The unbalanced nature of the dataset is addressed by using SMOTE.
- The models are trained using the processed dataset and evaluated against individual models and other ensemble combinations.

# Dataset Description

- **Mendeley Phishing Feature Dataset (2020)** is used to train the model. It has real entries of phishing and legitimate websites.
- It has **88,647 samples and 112 handcrafted features** in total, in which each entry represents a website, and the features are based on lexical, structural, and behavioral characteristics of the URL.
- The target attribute is a binary attribute with 1 indicating a phishing website and 0 indicating a legitimate website.
- The class distribution of the dataset is skewed, with 58,000 samples of legitimate websites and 30,647 samples of phishing websites.

# Dataset Description

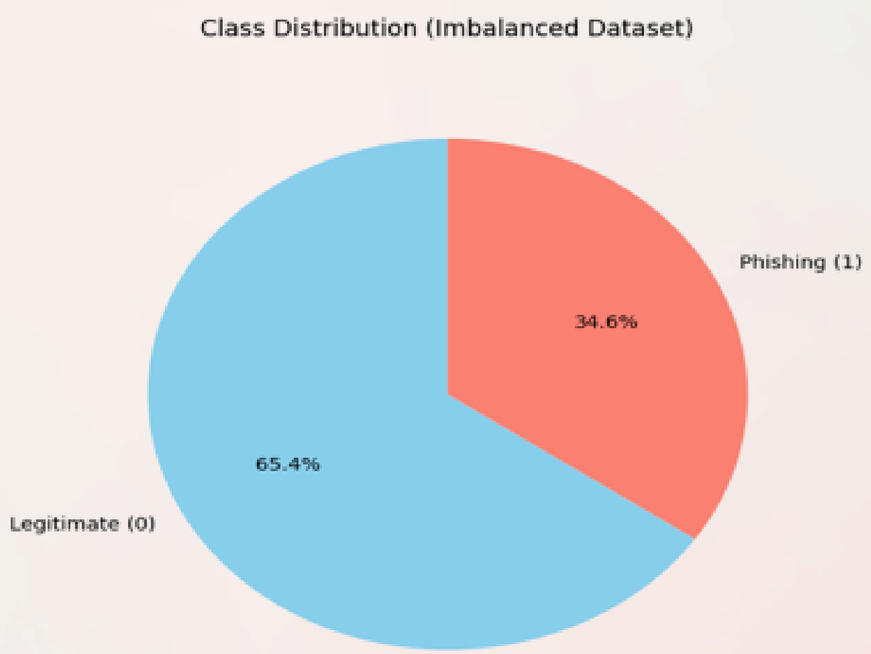



Figure 1: Dataset Class Distribution

# Dataset Preprocessing and Feature Engineering

- **Removal of Defective and Non-informative Features:** Constant Columns, Binary Placeholder Columns, and Handling -1 Values.
- **Feature Scaling:** Min-Max Normalization.
- **Correlation-Based Feature Reduction:** Pearson correlation coefficient.
- **Class Balancing with Smote Tomek:** Dataset is balanced with 57,642 samples for both classes.
- **Feature Selection using Permutation Importance:** Top 20 features are extracted.

# Dataset Preprocessing and Feature Engineering

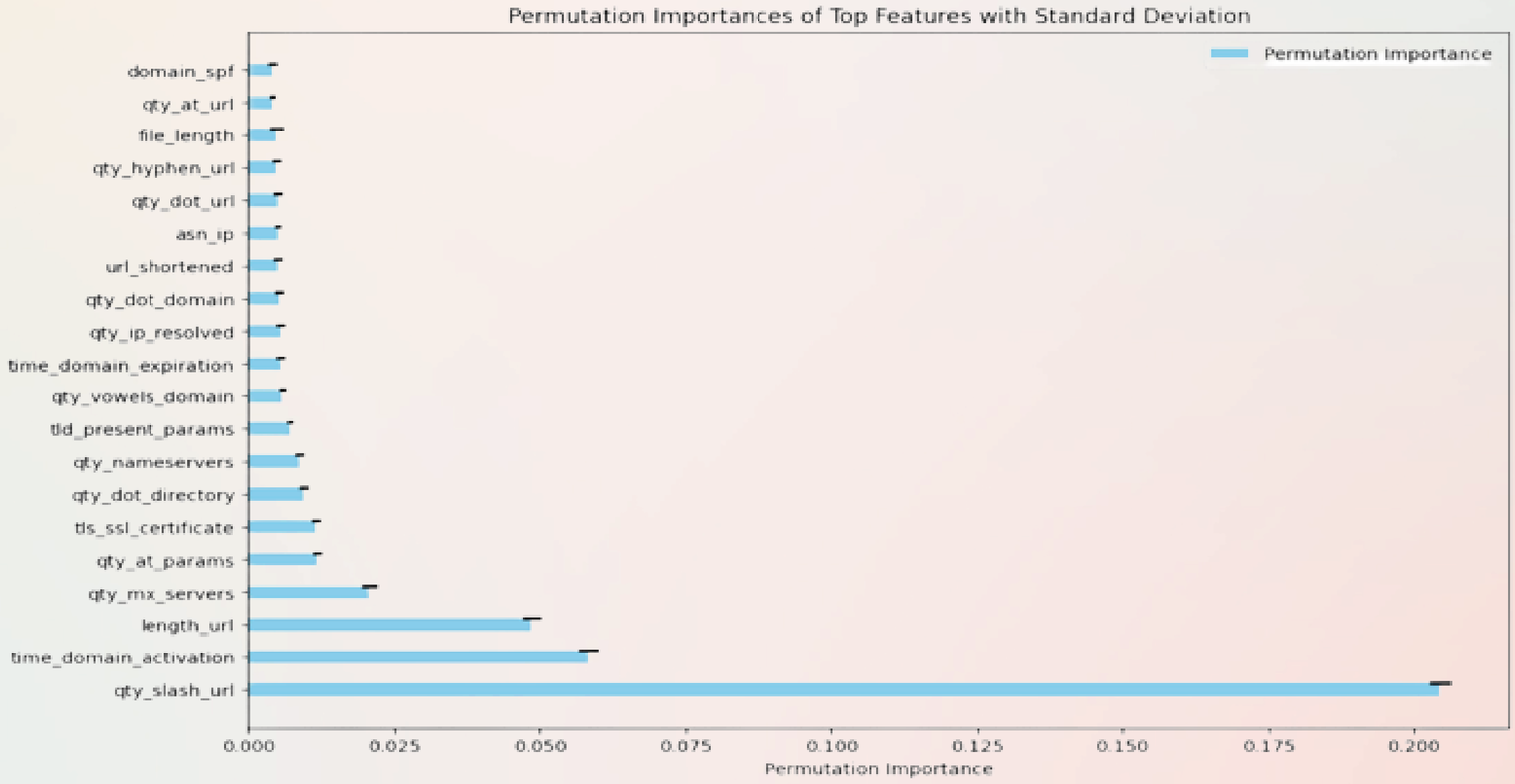



Figure 2: FNN Top 20 Features

# Dataset Preprocessing and Feature Engineering

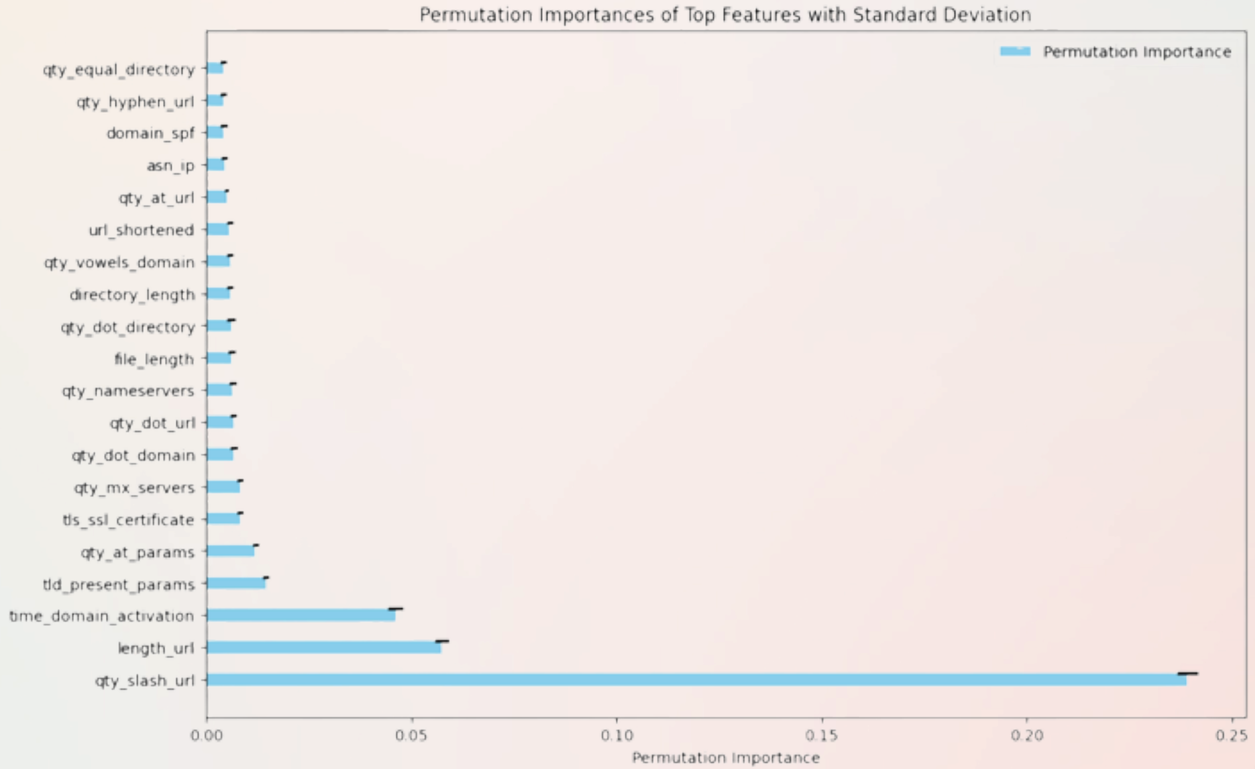



Figure 3: DNN Top 20 Features

# Dataset Preprocessing and Feature Engineering

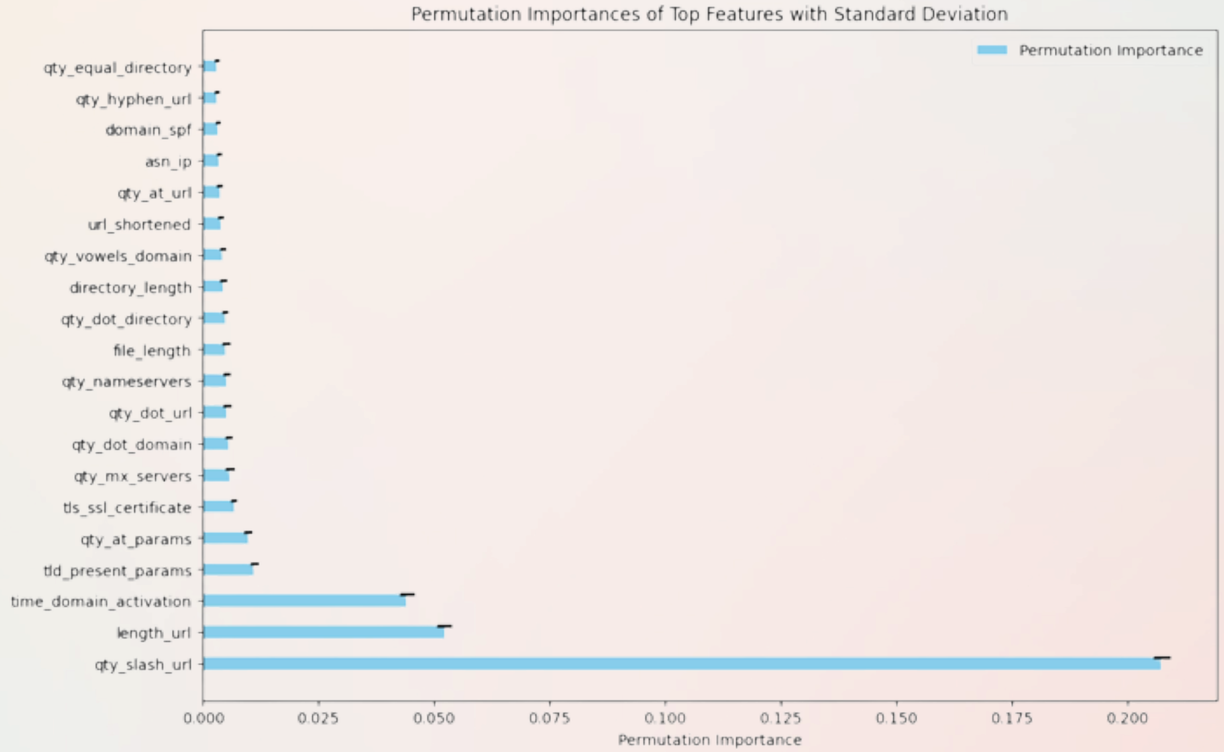



Figure 4: Wide and Deep Top 20 Features

# Dataset Preprocessing and Feature Engineering

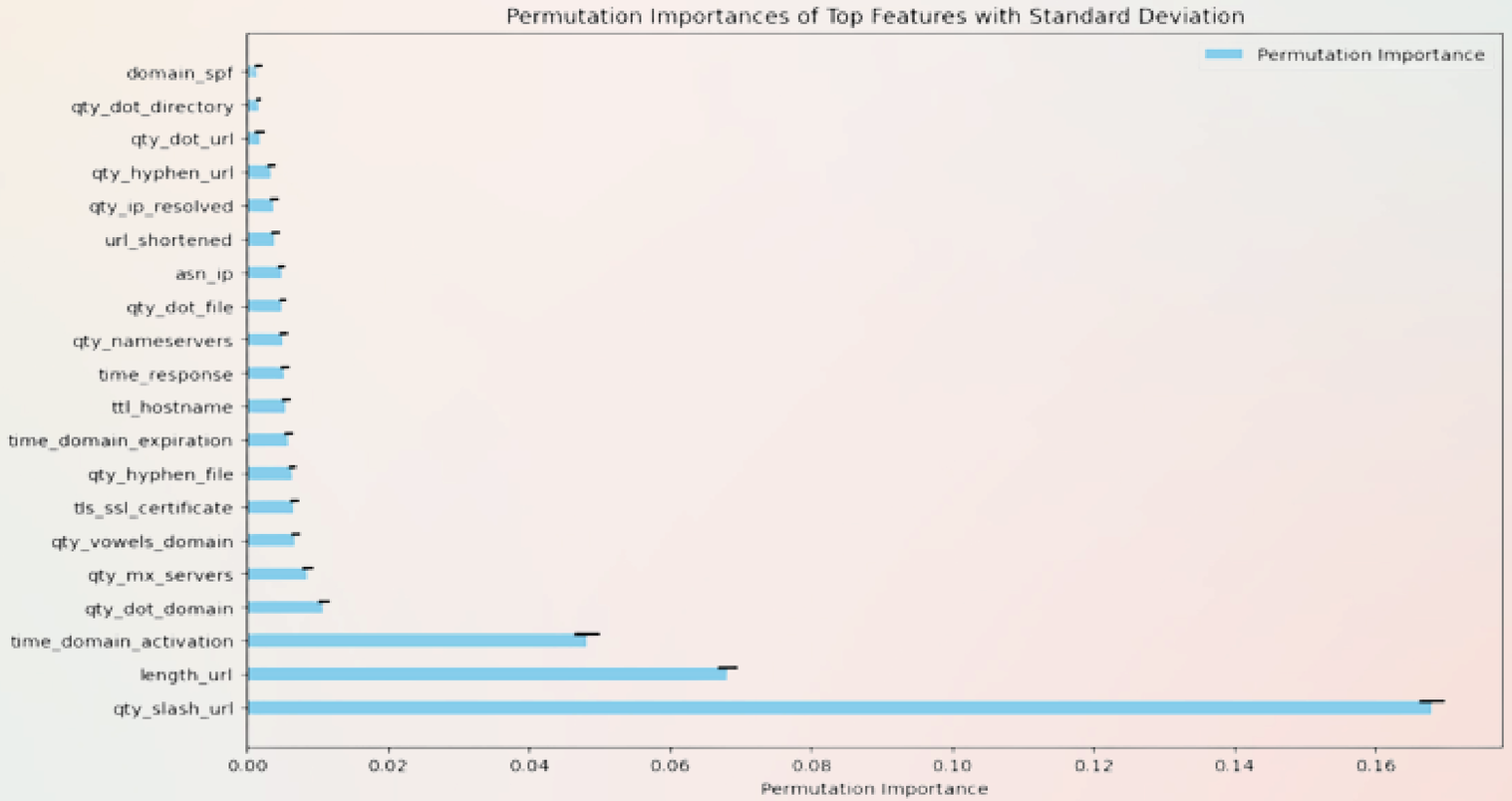



Figure 5: TabNet Top 20 Features

# Combined Feature Importance (Averaged across Models)

| Feature | Average Importance |
|---|---|
| qty_slash_url | 0.2048 |
| length_url | 0.0566 |
| time_domain_activation | 0.0492 |
| qty_mx_servers | 0.011 |
| qty_at_params | 0.0088 |
| tls_ssl_certificate | 0.008 |
| qty_dot_domain | 0.007 |
| qty_dot_directory | 0.0067 |
| qty_nameservers | 0.0065 |
| tld_present_params | 0.0063 |
| qty_vowels_domain | 0.0053 |
| url_shortened | 0.0048 |
| qty_dot_url | 0.0046 |
| asn_ip | 0.0043 |
| qty_hyphen_url | 0.004 |
| time_domain_expiration | 0.0038 |
| file_length | 0.0038 |
| directory_length | 0.0037 |
| qty_at_url | 0.0034 |
| domain_spf | 0.0033 |

Table 1. Averaged Top 20 Features

# Model Architectures and Training Procedures

- The models are trained with the extracted 20 prominent features independently.
- Evaluation is done through 5-fold stratified cross-validation for consistency.
- Each fold retains the original class distribution.
- For training four folds and one-fold for validation is used for each iteration.
- This is carried out for 5 iterations, hence in each fold, validation is done once and the final value is obtained by averaging the acquired values.

# Model Architectures and Training Procedures

- FNN comprises two hidden layers with 128 and 64 neurons, and ReLU activation functions are used.
- Similar architecture is used in the DNN model with two dense layers of 64 neurons to offer a better compact representation.
- The wide & Deep model combines memorization and generalization by processing the input using two parallel branches.
- Two shallow layers (32 and 16 neurons) are used in the wide component.
- TabNet is configured with a decision width (n_d) and attention width (n_a) of 8, across 3 decision steps.

# Ensemble Design and Prediction Alignment

Different ensemble combinations are tested to identify the best combination for phishing detection.

- FNN + DNN + Wide & Deep
- FNN + DNN + TabNet
- FNN + Wide & Deep + TabNet
- DNN + Wide & Deep + TabNet
- Full Ensemble: FNN + DNN + Wide & Deep + TabNet

# Performance Analysis

## Table 2. Balanced Dataset Performance

| Model | Accuracy | Precision | Recall | F1 Score | AUC-ROC |
|---|---|---|---|---|---|
| FNN | 95.39% | 93.57% | 97.48% | 95.48% | 0.9908 |
| DNN | 94.87% | 93.17% | 96.86% | 94.97% | 0.9888 |
| Wide & Deep | 94.86% | 92.92% | 97.12% | 94.97% | 0.9886 |
| TabNet | 95.18% | 93.35% | 97.29% | 95.28% | 0.9904 |
| FNN + DNN + Wide & Deep | 95.81% | 95.11% | 96.59% | 95.84% | 0.9916 |
| FNN + DNN + TabNet | 96.12% | 95.58% | 96.71% | 96.14% | 0.9927 |
| FNN + Wide & Deep + TabNet | 96.14% | 95.54% | 96.80% | 96.16% | 0.9927 |
| DNN + Wide & Deep + TabNet | 96.11% | 95.55% | 96.72% | 96.14% | 0.9926 |
| FNN + DNN + Wide & Deep + TabNet | 96.24% | 95.63% | 96.90% | 96.26% | 0.9928 |

# Performance Analysis

## Table 3. Imbalanced Dataset Performance

| Model | Accuracy | Precision | Recall | F1 Score | AUC-ROC |
|---|---|---|---|---|---|
| FNN | 95.44% | 93.39% | 93.42% | 93.40% | 0.9898 |
| DNN | 94.93% | 92.33% | 93.16% | 92.70% | 0.9886 |
| Wide & Deep | 95.10% | 92.79% | 93.11% | 92.93% | 0.9889 |
| TabNet | 94.96% | 91.93% | 93.68% | 92.78% | 0.9886 |
| FNN + DNN + Wide & Deep | 95.61% | 93.53% | 93.78% | 93.65% | 0.99 |
| FNN + DNN + TabNet | 95.68% | 93.81% | 93.68% | 93.74% | 0.9907 |
| FNN + Wide & Deep + TabNet | 95.69% | 93.83% | 93.70% | 93.77% | 0.9906 |
| DNN + Wide & Deep + TabNet | 95.57% | 93.68% | 93.49% | 93.59% | 0.9903 |
| FNN + DNN + Wide & Deep + TabNet | 95.73% | 93.95% | 93.67% | 93.81% | 0.9907 |

# Performance Analysis

## Table 4. Performance Comparison

| Scheme | Accuracy |
|---|---|
| Setu et al. [18] | 95.48% |
| Chiew et al. [19] | 94.60% |
| Proposed | 96.24% |

# Conclusion

- An enhanced deep ensemble learning approach for phishing detection is proposed in this paper.
- A comprehensive comparison of different individual models and ensemble approaches is conducted, which shows that the ensemble combination with FNN, DNN, Wide & Deep, and TabNet outperforms other individual models and ensemble combinations.
- It performs better with balanced and imbalanced datasets.
- Feature extraction also plays an important role in extracting the prominent features and reducing the execution time of the models.
- This proves that the ensemble approach combines the strengths of the individual models to better predict the phishing URLs.
- The proposed scheme is also compared with other existing solutions.

# References


1. Adeyemo, V.E., Balogun, A.O., Mojeed, H.A., Akande, N.O., Adewole, K.S. (2021). Ensemble-Based Logistic Model Trees for Website Phishing Detection. In: Anbar, M., Abdullah, N., Manickam, S. (eds) Advances in Cyber Security. ACeS 2020. Communications in Computer and Information Science, vol 1347. Springer, Singapore.
2. Nisreen Innab, Ahmed Abdelgader Fadol Osman, Mohammed Awad Mohammed Ataelfadiel, Marwan Abu-Zanona, Bassam Mohammad Elzaghmouri, Farah H. Zawaideh, and Mouiad Fadeil Alawneh, “Phishing Attacks Detection Using Ensemble Machine Learning Algorithms”, Computers, Materials and Continua, Volume 80, Issue 1, 18 July 2024, Pages 1325-1345.
3. Bhogesh Karthik Gontla, Priyanka Gundu, Padma Jyothi Uppalapati, Kandula Nara-simharao, and S Mahaboob Hussain, “A Machine Learning Approach to Identify Phish-ing Websites: A Comparative Study of Classification Models and Ensemble Learning Techniques”, EAI Endorsed Transactions on Scalable Information Systems, Volume 10, Issue 5, 2023.
4. J. O. Ajayi, A. O. Adetunmbi, T. A. Olowookere, and S. A. Sodiya, “Phishing Detec-tion: Performance Evaluation of Both Ensemble and Classical Machine Learning Mod-els”, International Journal of Information Security, Privacy and Digital Forensics. An International Journal of the Nigeria Computer Society (NCS), Vol. 6, No. 2, December 2022.
5. C. Singh and Meenu, "Phishing Website Detection Based on Machine Learning: A Sur-vey," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 398-404.

# References


6. Iman Akour, Noha Alnazzawi, Ahmad Aburayya, Raghad Alfaisal, and Said A. Salloum, “Using Classical Machine Learning for Phishing Websites Detection from URLs”, Journal of Management Information and Decision Sciences, Volume 24, Special Issue 6, 2021.
7. Kumar, Jitendra, et al. "Phishing website classification and detection using machine learning." 2020 international conference on computer communication and informatics (ICCCI). IEEE, 2020.
8. Alam, Mohammad Nazmul, et al. "Phishing attacks detection using machine learning approach." 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020.
9. Gururaj H L, Goutham BoreGowda. "Phishing website detection based on effective machine learning approach." Journal of Cyber Security Technology 5.1 (2021): 1-14.
10. Rashid, Junaid, et al. "Phishing detection using machine learning technique." 2020 first international conference of smart systems and emerging technologies (SMARTTECH). IEEE, 2020.
11. Desai, Priyanshu, and Mann Shah. "Phishing Website Detection Using Machine Learn-ing: A Comprehensive Study." Int J Multidiscip Res (2023).
12. Saha, Ishita, et al. "Phishing attacks detection using deep learning approach." 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2020.

# References


13. Basit, Abdul, et al. "A novel ensemble machine learning method to detect phishing at-tack." 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, 2020.
14. Dharani, M., et al. "Detection of phishing websites using ensemble machine learning approach." ITM Web of Conferences. Vol. 40. EDP Sciences, 2021.
15. Avula Vennela, Ramesh Babu Akarapu, Bairi Levi Rakshith, Lazarus Godson Asirvat-ham, G. Sunil, Intelligent cybersecurity systems for phishing attack detection - An overview, Computers and Electrical Engineering, Volume 130, 2026, 110829.
16. Mohammad Amaz Uddin, Md Mahiuddin, Iqbal H. Sarker, An explainable transformer-based model for phishing email detection: A large language model approach, Computer Networks, Volume 277, 2026, 112061.
17. Lapina, M., Gediev, K., Babenko, M., Tchernykh, A., Drozdov, A.Yu.: DNS Tunneling Detection Using Methods of Machine Learning. Program Comput Soft. 51, 727–751 (2025). https://doi.org/10.1134/S036176882570053 7.
18. J. H. Setu, N. Halder, A. Islam and M. A. Amin, "RSTHFS: A Rough Set Theory-Based Hybrid Feature Selection Method for Phishing Website Classification," in IEEE Access, vol. 13, pp. 68820-68830, 2025, doi: 10.1109/ACCESS.2025.3561237.
19. K. L. Chiew, C. L. Tan, K. Wong, K. S. C. Yong, and W. K. Tiong, “A new hybrid en-semble feature selection framework for machine learning-based phishing detection sys-tem,” Inf. Sci., vol. 484, pp. 153–166, May 2019.

# Thank you