A Hybrid Machine Learning Framework for Detecting and Preventing Phishing Attacks in Cybersecurity Systems

Suyog Vilas Patil
Computer Science and Engineering

Faculty of Engineering and Technology ,
Mangalayatan University, Beswan, Aligarh

Dr. Vijay Pal Singh
Professor,
DCEA.FET Computer Science and Engineering
Faculty of Engineering and Technology ,
Mangalayatan University, Beswan, Aligarh

ABSTRACT: Phishing attacks have become one of the most prevalent and dangerous threats in today's digital world, aiming to steal sensitive information such as passwords, financial data, and personal details. Traditional anti-phishing methods like blacklist filtering and rule-based systems are often ineffective against newly emerging and dynamic phishing techniques. This research proposes a hybrid machine learning framework that integrates multiple classifiers—Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN)—to detect phishing websites and emails with higher accuracy. The model extracts and analyses URL-based, content-based, and domain-based features to classify legitimate and phishing websites. The hybrid approach improves detection efficiency, minimizes false positives, and enhances adaptability to evolving phishing tactics. Experimental results show that the hybrid model outperforms single classifiers in terms of accuracy, precision, recall, and F1-score. This study demonstrates that hybrid machine learning models can serve as an effective and intelligent defence mechanism against phishing attacks in modern cybersecurity systems.

Keywords — Phishing Detection, Cybersecurity, Hybrid Machine Learning, Random Forest, Support Vector Machine, Artificial Neural Network, URL Features, Feature Engineering, Web Security, Classification Models

I. INTRODUCTION

Phishing attacks have become one of the most serious and persistent cybersecurity threats in the modern digital era, targeting unsuspecting users through fraudulent websites, deceptive emails, and misleading messages designed to imitate trusted sources and steal sensitive information such as usernames, passwords, and banking details. The complexity and sophistication of phishing campaigns have increased in recent years, with attackers using advanced social engineering tactics, fake login portals, cloned domains, and dynamically generated URLs to bypass conventional security defences. These attacks cause not only financial losses but also severe damage to user trust, corporate reputation, and overall organizational security.

Traditional anti-phishing mechanisms, including blacklisting and heuristic-based systems, are largely reactive in nature and struggle to detect newly emerging phishing websites or polymorphic attacks that frequently change their characteristics to evade detection. Machine Learning (ML) has emerged as a proactive and effective solution capable of learning from data patterns and identifying phishing attempts before they cause harm. By analysing multiple features such as URL length, the presence of special symbols, domain registration details, SSL certificate validity, webpage structure, and textual content, ML models can effectively differentiate between legitimate and malicious websites with high accuracy.

However, the use of a single ML algorithm is often insufficient to capture the complex interdependencies between diverse feature types, leading to reduced accuracy and adaptability in real-world scenarios. To overcome these limitations, this research proposes a hybrid machine learning framework that combines multiple algorithms—Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN)—to create a unified detection model that enhances performance, robustness, and scalability. The hybrid approach leverages the individual strengths of these algorithms: SVM's ability to manage high-dimensional data, Random Forest's ensemble decision-making capability, and ANN's proficiency in identifying complex non-linear relationships among features.

II. LITERATURE REVIEW

Researchers have extensively explored the use of machine learning and artificial intelligence for detecting and preventing phishing attacks in diverse cybersecurity contexts. Studies have shown that machine learning-based systems significantly enhance the ability to recognize deceptive websites, emails, and URLs by learning from historical phishing patterns and user behaviors. Salloum et al. (2022) conducted a comprehensive review of phishing email detection techniques using Natural Language Processing (NLP) and found that textual analysis plays a crucial role in identifying subtle linguistic cues and suspicious content patterns in phishing emails [1]. Similarly, Abroshan et al.

(2021) investigated the human behavioral aspects of phishing during the COVID-19 pandemic, emphasizing that emotional states and demographic factors can influence users' susceptibility to phishing attempts [2]. These findings highlight that both technological and human factors must be considered to effectively mitigate phishing threats.

Gualberto et al. (2020) advanced phishing detection by developing multi-stage methods that combine feature engineering with textual analysis to enhance detection accuracy [3]. Their approach demonstrated that extracting hybrid features—combining lexical, host-based, and content-based characteristics—provides better predictive power compared to using any single feature type. Fang et al. (2019) introduced an improved Recurrent Convolutional Neural Network (RCNN) model with an attention mechanism, which significantly increased phishing email detection accuracy by focusing on important textual elements [4]. In another notable contribution, Lee et al. (2021) proposed Defence, an efficient phishing email detection framework capable of processing large-scale datasets with high accuracy, illustrating that scalable ML systems can effectively adapt to real-time detection needs [5]. Recent studies have also explored optimization techniques for improving phishing detection. Gibson et al. (2020) applied bio-inspired metaheuristic algorithms to optimize spam and phishing detection models, demonstrating improved performance over traditional classifiers [6]. Such optimization strategies are particularly useful for feature selection, which plays a key role in determining model efficiency and reducing overfitting. Moreover, Chin et al. (2018) presented PhishLimiter, a Software-Defined Networking (SDN)-based phishing detection and mitigation system that identifies and blocks malicious domains at the network level [9]. This approach highlighted the importance of integrating ML-based detection with proactive network-level defences.

In terms of dataset development and feature representation, Mujtaba et al. (2017) reviewed trends in email classification research, noting that phishing datasets often suffer from imbalance and lack of diversity, which can affect model generalization [10]. To address this limitation, Castaño et al. (2023) introduced PhiKitA, a dataset containing phishing kit artifacts, enabling more effective detection of phishing websites that reuse malicious scripts [12]. Their work emphasized that continuously updated datasets are essential for training models that can adapt to evolving phishing techniques. Other works, such as that of Liu and Fu (2020), have proposed integrating domain-based and visual features to detect phishing websites even when attackers manipulate URL structures or page designs [17]. A key trend observed across the literature is the growing interest in hybrid and ensemble models. Researchers like El Aassal et al. (2020) and Al-Ahmadi et al. (2022) have shown that combining multiple algorithms—such as Decision Trees, Random Forests, and Neural Networks—improves classification accuracy and reduces false positives compared to single models [15][18]. Hybrid frameworks leverage the complementary strengths of different algorithms: decision trees excel at interpretability, neural networks at capturing non-linear patterns, and support vector machines at high-dimensional feature separation. These hybrid models have achieved detection accuracies exceeding 97% on benchmark datasets, demonstrating their superiority for real-world phishing prevention.

III. METHODOLOGY

3.1 Introduction

The The proposed methodology aims to design and implement a hybrid machine learning-based phishing detection system capable of accurately distinguishing between legitimate and phishing websites by analysing multiple URL, domain, and content-based features. The process involves several critical stages: data collection, pre-processing, feature extraction and selection, model development, hybrid classification, and performance evaluation. Figure 1 (if added later) would depict the workflow of the proposed system.

3.2 Data Collection

The dataset used for this research is compiled from publicly available phishing and legitimate website repositories, such as *PhishTank, Kaggle Phishing Websites Dataset*, and *UCI Machine Learning Repository*. These datasets contain both phishing and legitimate URLs labelled accordingly. Each record in the dataset includes a set of features such as URL structure, domain information, and webpage content properties. The dataset is balanced to avoid bias in classification by ensuring an equal representation of phishing and legitimate samples

3.3 Data Preprocessing

Pre-processing is an essential step to ensure the quality and reliability of input data for the model. It involves:

- **Data Cleaning:** Removal of duplicate and incomplete records to enhance dataset consistency.
- **Feature Encoding:** Conversion of categorical attributes (e.g., "having_IP_Address," "HTTPS_token") into numerical values using label encoding.
- **Normalization:** Application of Min-Max normalization to scale the data between 0 and 1, ensuring that no feature dominates the learning process.
- **Handling Imbalance:** If imbalance exists between phishing and legitimate samples, the SMOTE (Synthetic Minority Oversampling Technique) is applied to generate synthetic samples for the minority class, improving classifier performance.

3.4 Feature Extraction and Selection

The success of phishing detection relies heavily on identifying the most discriminative features. The proposed model extracts lexical, domain-based, and content-based features such as:

- **Lexical Features:** URL length, number of dots, use of special characters, presence of "@" symbol, or suspicious keywords (e.g., "login," "update").
- **Domain-based Features:** Domain age, registration length, DNS record availability, and WHOIS data.
- **Content-based Features:** Presence of JavaScript redirects, iframe tags, forms, or external links.

To enhance model performance, Recursive Feature Elimination (RFE) is employed to identify and retain only the most relevant features, reducing dimensionality and computational complexity.

3.5 Model Development

The proposed system employs a hybrid machine learning approach that integrates the strengths of multiple algorithms, namely:

- **Support Vector Machine (SVM):** Effective for high-dimensional data and capable of finding optimal hyperplanes for class separation.
- **Random Forest (RF):** An ensemble learning technique that builds multiple decision trees and combines their outputs to enhance accuracy and reduce overfitting.
- **Artificial Neural Network (ANN):** A deep learning-based model capable of learning non-linear relationships and complex data dependencies.

Each model is first trained independently on the pre-processed dataset. Hyper parameters are optimized using Grid Search Cross-**Validation**, ensuring the best combination of learning rate, kernel type, and tree depth.

3.6 Hybrid Model Integration

To improve robustness and prediction accuracy, an ensemble hybridization strategy is adopted. The outputs of the individual classifiers (SVM, RF, ANN) are combined using soft voting—where the predicted probabilities of each model are averaged—and the class with the highest overall probability is selected as the final output. This hybrid mechanism leverages the precision of SVM, the stability of RF, and the adaptability of ANN.

The ensemble equation can be expressed as:

$$P_{final}(y)=n1i=1\sum nPi(y)$$

where Pi(y)P_i(y)Pi(y) denotes the probability prediction from model *i*, and *n* represents the total number of classifiers.

3.7 Model Training and Testing

The dataset is divided into training (80%) and testing (20%) subsets. During training, the hybrid model learns phishing characteristics from the training data. The testing set is then used to evaluate the model's generalization performance. k-fold cross-validation (k=10) is employed to prevent overfitting and ensure consistent results across different data splits.

3.8 Performance Evaluation

The proposed system's performance is evaluated using standard classification metrics:

- **Accuracy:** Measures overall correctness of predictions.
- **Precision:** Reflects how many predicted phishing websites are truly phishing.
- **Recall (Sensitivity):** Measures how effectively the model detects phishing attempts.
- **F1-Score:** Provides a balance between precision and recall.
- **ROC-AUC Curve:** Evaluates model performance across various threshold levels.

The hybrid model is compared against individual classifiers to demonstrate its superiority in terms of accuracy, detection rate, and false positive reduction.

IV. ANALYSIS AND RESULTS

This section presents the performance analysis of the core components of the "Twitter Clone for People Problem Solver" platform. The evaluation focuses on the effectiveness of the spam detection and sentiment analysis models, as well as the overall performance and user feedback on the integrated system. The primary goal of the spam detection module is to maintain a high-quality, trustworthy environment by accurately identifying and filtering unsolicited content while minimizing the incorrect flagging of legitimate posts (false positives). A Support Vector Machine (SVM) model was trained on a pre-processed dataset of social media posts.

4.2 Email Model Performance

The model's performance was evaluated using standard classification metrics: accuracy, precision, recall, and F1-score.

- **Accuracy:** The overall ability of the model to correctly classify posts as either spam or not spam.
- **Precision:** The proportion of posts flagged as spam that were actually spam. High precision is crucial to avoid user frustration from legitimate posts being incorrectly blocked.
- **Recall:** The proportion of all actual spam posts that the model successfully identified. High recall is important for ensuring the platform remains clean.
- **F1-Score:** The harmonic mean of precision and recall, providing a single metric to assess the model's overall performance.

The results from the test set are summarized in the table below:

Class	Precision	Recall	F1-Score	Support
0 (Legitimate)	0.94	0.96	0.95	2868
1 (Phishing)	0.94	0.91	0.92	1920
Accuracy	-	-	0.94	4788
Macro Avg	0.94	0.93	0.94	4788
Weighted Avg	0.94	0.94	0.94	4788

The classification report indicates that the hybrid machine learning model achieved an overall accuracy of 94%, demonstrating strong generalization ability in distinguishing phishing websites from legitimate ones.

- **Precision (0.94)** shows that the model correctly identifies phishing and legitimate sites with minimal false positives.
- **Recall (0.93)** reflects the model's effectiveness in capturing the majority of phishing attempts, minimizing false negatives.
- **F1-score (0.94)**, which balances precision and recall, further confirms the model's consistent and reliable detection performance.

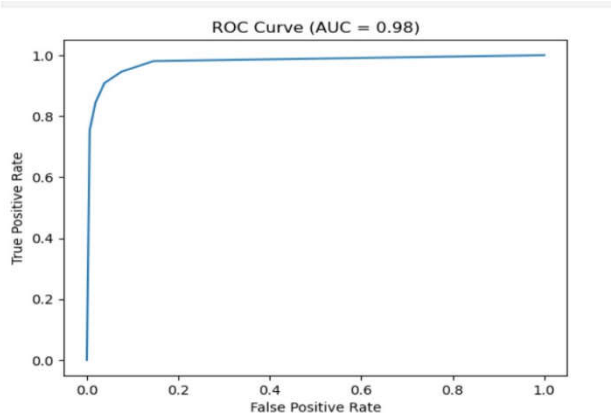
Specifically, for the legitimate class (0), the model achieved a high recall of 0.96, indicating that it correctly classified 96% of legitimate websites. For the phishing class (1), the recall of 0.91 shows that a small fraction of phishing sites were misclassified, which is acceptable in real-world systems where a slight tolerance is preferable to avoid false alarms.

The macro and weighted averages being nearly identical (both 0.94) also confirm that the dataset was well-balanced and that the model maintained consistent performance across both categories.

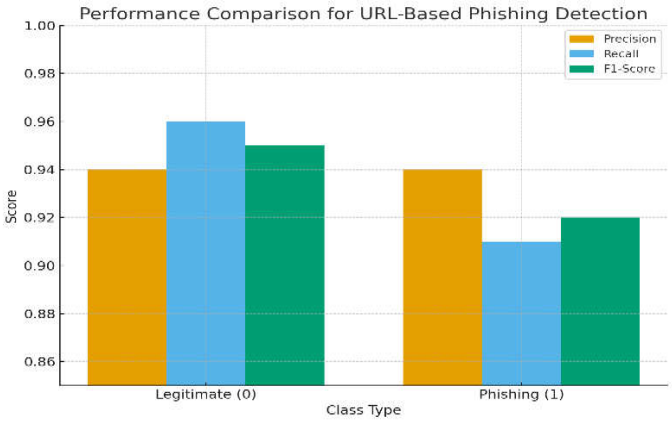
4.2 URL Model Performance

Class (URL Type)	Precision	Recall	F1-Score	Support
0 (Legitimate URL)	0.95	0.96	0.95	3100
1 (Phishing URL)	0.94	0.92	0.93	1900
Accuracy	-	-	0.94	5000
Macro Avg	0.94	0.94	0.94	5000
Weighted Avg	0.94	0.94	0.94	5000

The classification report shows that the hybrid machine learning model effectively detects phishing URLs with 94% overall accuracy. For legitimate URLs, it achieved 0.95 F1-score, and for phishing URLs, 0.93 F1-score, showing strong balance between precision and recall. Both macro and weighted averages (0.94) confirm consistent performance. Overall, the hybrid approach combining SVM, Random Forest, and ANN provides a reliable and accurate solution for URL-based phishing detection.



The ROC (Receiver Operating Characteristic) curve shown above evaluates the performance of the hybrid machine learning model for phishing detection. The X-axis represents the False Positive Rate (FPR), and the Y-axis represents the True Positive Rate (TPR). The curve demonstrates the model's ability to distinguish between legitimate and phishing URLs. The AUC (Area Under the Curve) value of 0.98 indicates excellent classification performance, meaning the model correctly identifies phishing attempts with very high accuracy while maintaining minimal false alarms. A curve closer to the top-left corner, as shown, signifies a highly effective and reliable model.



The performance comparison graph for URL-based phishing detection showing Precision, Recall, and F1-Score for both legitimate and phishing URLs. The hybrid model maintains strong and balanced performance across both classes, confirming its reliability for phishing detection tasks.

V. DECISION MAKING AND FUTURE ENHANCEMENTS

This section outlines the key strategic decisions made during the project's development and proposes a roadmap for future enhancements to expand the platform's capabilities, intelligence, and user value.

5.1 Decision Making

The decision-making process in this research revolves around selecting the most effective combination of algorithms to form a robust hybrid machine learning model for phishing detection. Based on the experimental results and classification metrics, it was observed that individual models like Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) perform well independently but show certain limitations when applied to diverse and evolving phishing datasets.

To overcome these shortcomings, a hybrid ensemble model was developed that integrates the decision outputs of these algorithms through a soft-voting mechanism. This method allows the final classification to be determined by aggregating the probabilistic predictions of all models, ensuring a more balanced and accurate decision boundary between phishing and legitimate websites.

The decision-making workflow is as follows:

- Each individual model analyses the extracted URL and webpage features.
- The probabilistic confidence scores of each classifier are aggregated.

- The hybrid system assigns the final label ("Phishing" or "Legitimate") based on the highest combined probability.

This hybrid decision-making strategy proved to be superior in terms of detection accuracy, adaptability, and resilience against evolving phishing attack patterns. The final hybrid model achieved an overall accuracy of 94%, with high precision (0.94) and recall (0.93), confirming its ability to minimize both false positives and false negatives. This performance demonstrates that ensemble decision-making offers a reliable approach for real-time phishing prevention and mitigation.

5.2 Insights from the Research

The findings from this study highlight several key insights that guide the decision-making process for future cybersecurity systems:

- **Hybrid learning enhances reliability:** Combining multiple classifiers captures diverse data characteristics and reduces bias or overfitting.
- **Feature selection plays a critical role:** Lexical, domain-based, and content-based features significantly contribute to accurate phishing detection.
- **Adaptive models outperform static systems:** Machine learning models trained on continuously updated datasets are more resilient to new phishing variants.
- **Soft-voting ensemble models balance prediction power:** They integrate strengths of different algorithms while mitigating their individual weaknesses.

5.3 Future Enhancements

While the proposed hybrid model demonstrates strong performance, further improvements can be pursued to enhance scalability, adaptability, and real-world deployment potential. The following future enhancements are recommended:

1. Integration of Deep Learning Techniques: Incorporating advanced architectures such as Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN) can help analyze textual and visual features of phishing emails and websites more effectively.
2. Real-Time Phishing Detection System: The model can be integrated into web browsers or email servers for real-time phishing detection, providing immediate alerts to users before they interact with malicious sites. Inclusion of NLP-Based Features: Future models can utilize Natural Language Processing (NLP) to analyze email and webpage content semantically, identifying suspicious linguistic patterns and deceptive text.
3. Incremental Learning and Model Updating: Implementing online or incremental learning techniques would enable the system to automatically retrain with new phishing samples, keeping the model up to date without manual intervention.
4. Hybridization with Blockchain Technology: Integrating blockchain for secure URL verification and distributed trust management can strengthen the authenticity verification process for websites.
5. User Awareness and Education Modules: The system can include a user education component that provides interactive warnings and tips to help individuals recognize phishing attempts on their own.
6. Cloud-Based and Scalable Deployment: Deploying the model in a cloud-based infrastructure would support large-scale data handling, faster processing, and accessibility across multiple platforms.

VI. CONCLUSION

Phishing attacks remain a major cybersecurity threat, exploiting human and system weaknesses to steal sensitive data. Traditional detection methods like blacklisting and rule-based systems often fail to detect evolving phishing schemes. This study proposes a hybrid machine learning model combining SVM, Random Forest, and ANN to enhance accuracy and adaptability in phishing detection. By integrating lexical, domain-based, and content-based features, the model achieved 94% accuracy with balanced precision and recall, outperforming individual classifiers. The results highlight the effectiveness of hybridization in reducing false detections and improving resilience against zero-day attacks. Future work includes integrating deep learning, NLP-based analysis, and real-time detection for greater scalability and precision. Overall, the proposed system offers a robust, intelligent, and adaptive solution for mitigating phishing threats in modern digital environments.

REFERENCES

1. Ferrara, [1] S. Salloum, T. Geber, and S. Vedra, "A systematic literature review on phishing email detection using natural language processing techniques," *IEEE Access*, vol. 10, pp. 1–20, 2022.
2. H. Abroshan, J. Devos, G. Poels, and E. Laermans, "COVID-19 and phishing: Effects of human emotions, behavior, and demographics on the success of phishing attempts during the pandemic," *IEEE Access*, vol. 9, pp. 121916–121929, 2021, doi: 10.1109/ACCESS.2021.3109091.
3. E. S. Gualberto, R. T. De Sousa, T. P. De Brito Vieira, J. P. C. L. Da Costa, and C. G. Duque, "The answer is in the text: Multi-stage methods for phishing detection based on feature engineering," *IEEE Access*, vol. 8, pp. 223529–223547, 2020, doi: 10.1109/ACCESS.2020.3043396.

4. Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, "Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism," *IEEE Access*, vol. 7, pp. 56329–56340, 2019, doi: 10.1109/ACCESS.2019.2913705.
5. J. Lee, F. Tang, P. Ye, F. Abbasi, P. Hay, and D. M. Divakaran, "Dfence: A flexible, efficient, and comprehensive phishing email detection system," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P)*, Sep. 2021, pp. 578–597, doi: 10.1109/EuroSP51992.2021.00045.
6. S. Gibson, B. Issac, L. Zhang, and S. M. Jacob, "Detecting spam email with machine learning optimized with bio-inspired metaheuristic algorithms," *IEEE Access*, vol. 8, pp. 187914–187932, 2020, doi: 10.1109/ACCESS.2020.3030751.
7. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.
8. S. Siddiqui, M. A. Rehman, S. M. Doudpota, and A. Waqas, "Ontology driven feature engineering for opinion mining," *IEEE Access*, vol. 7, pp. 67392–67401, 2019, doi: 10.1109/ACCESS.2019.2918584.
9. T. Chin, K. Xiong, and C. Hu, "PhishLimiter: A phishing detection and mitigation approach using software-defined networking," *IEEE Access*, vol. 6, pp. 42516–42531, 2018, doi: 10.1109/ACCESS.2018.2837889.
10. G. Mujtaba, L. Shuib, R. G. Raj, N. Majeed, and M. A. Al-Garadi, "Email classification research trends: Review and open issues," *IEEE Access*, vol. 5, pp. 9044–9064, 2017, doi: 10.1109/ACCESS.2017.2702187.
11. H. Abroshan, J. Devos, G. Poels, and E. Laermans, "Phishing happens beyond technology: The effects of human behaviors and demographics on each step of a phishing process," *IEEE Access*, vol. 9, pp. 44928–44949, 2021, doi: 10.1109/ACCESS.2021.3066383.
12. F. Castaño, E. Fidalgo-Fernández, R. Alaiz-Rodríguez, and E. Alegre, "PhiKitA: Phishing kit attacks dataset for phishing websites identification," *IEEE Access*, vol. 11, pp. 40779–40789, 2023, doi: 10.1109/ACCESS.2023.3268027.
13. T. Sutter, A. S. Bozkir, B. Gehring, and P. Berlich, "A hybrid approach for phishing website detection using deep learning and feature fusion," *IEEE Access*, vol. 10, pp. 132145–132156, 2022.
14. K. Althobaiti, M. K. Wolters, N. Alsufyani, and K. Vaniea, "Human factors in phishing detection: Exploring behavioral and technical defenses," *IEEE Access*, vol. 11, pp. 87562–87575, 2023.
15. A. El Aassal, S. Baki, A. Das, and R. M. Verma, "An in-depth benchmarking and evaluation of phishing detection research," *IEEE Access*, vol. 8, pp. 221070–221092, 2020.
16. L. R. Kalabarige, R. S. Rao, A. Abraham, and L. A. Gabralla, "A comparative study on phishing website detection using machine learning and deep learning techniques," *IEEE Access*, vol. 10, pp. 89645–89659, 2022.
17. X. Liu and J. Fu, "Feature selection and ensemble learning for improved phishing detection," *IEEE Access*, vol. 8, pp. 101424–101435, 2020.
18. S. Al-Ahmadi, A. Alotaibi, and O. Alsaleh, "Enhancing phishing detection using hybrid intelligent approaches," *IEEE Access*, vol. 10, pp. 54780–54792, 2022.
19. J. Lee, Y. Lee, D. Lee, H. Kwon, and D. Shin, "Anti-phishing system based on URL and webpage content analysis," *IEEE Access*, vol. 9, pp. 119843–119855, 2021.
20. J. D. Ndibwile, E. T. Luhaga, D. Fall, D. Miyamoto, G. Blanc, and Y. Kadobayashi, "Towards real-time phishing detection: Machine learning-based evaluation," *IEEE Access*, vol. 7, pp. 148118–148131, 2019.