Journal of Posthumanism
2025
Volume: 5, No: 3, pp. 1075–1089
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v5i3.854

DaE2: A Diverse and Efficient Ensemble Framework for Real-Time Phishing URL Detection

Nawaf Alshdaifat[1], Suleiman Ibrahim Mohammad[2], Khaleel Ibrahim Al- Daoud[3], Suhaila Abuowaida[4], Asokan Vasudevan[5], Abdel-Rahman Al-Ghuwairi[6], Muhammad Turki Alshurideh[7]

Abstract

Phishing attacks continue to pose a significant cybersecurity threat, with over 1.2 million unique phishing sites detected in 2023. Traditional detection methods struggle to keep pace with rapidly evolving phishing techniques, particularly given the short lifespan of modern phishing websites. This paper presents DaE2 (Diverse and Efficient Ensemble), a novel ensemble-based approach for real-time phishing URL detection. Our framework integrates four distinct ensemble methods—AdaBoost, Bagging, Stacking, and Voting—each chosen for their unique strengths in handling different aspects of URL classification. Using a comprehensive dataset of 11,055 URLs with 30 engineered features, we demonstrate that DaE2 achieves 98.7% accuracy while maintaining real-time processing capabilities. The framework shows particular strength in detecting newly created phishing URLs (99.1% accuracy) and handling sophisticated masquerading techniques (98.5% accuracy), while maintaining low false positive rates (1.6%). Performance analysis reveals that our approach outperforms traditional methods by 2.5% in accuracy while reducing processing time by 35% compared to deep learning approaches. The framework's adaptive weighting mechanisms and parallel processing implementation ensure robust performance across varying types of phishing attacks, making it suitable for real-world deployment.

Keywords: *Phishing Detection, Ensemble Learning, Machine Learning, Cybersecurity, URL Classification, Real-time Detection, Feature Engineering.*

Introduction

Internet is now reached by more than 5.7 billion people accounting 71.3% of total population on world. Such pervasive connectivity has reshaped the way businesses operate across every industry, radically disrupting key value chains. Every day, more people interact with the Internet in their real-life activities and leave more digital traces in cyberspace (Liu et al., 2017). Every web page that users visit is assigned a Uniform Resource Locator (URL), used to uniquely identify the location of content. But this widespread utilization of URLs has created a major

[1] Faculty of IT, Applied Science Private University, Amman, Jordan.
[2] Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, Jordan, Research follower, INTI International University, 71800 Negeri Sembilan, Malaysia, dr_sliman@yahoo.com, (Corresponding Author), ORCID: (0000-0001-6156-9063)
[3] Department of Accounting, Business School Faculties, Al Ahilya Amman University, Amman, Jordan,
[4] Department of Computer Science, Faculty of Prince Al-Hussein Bin Abdallah II for Information Technology, Al Al-Bayt University, Mafraq 25113, Jordan.
[5] Faculty of Business and Communications, INTI International University, 71800 Negeri Sembilan, Malaysia.
[6] Department of Software Engineering, Faculty of Prince Al-Hussein Bin Abdallah II for Information Technology, The Hashemite University, Zarqa, Jordan.
[7] Department of Marketing, School of Business, The University of Jordan, Amman 11942, Jordan

CC BY NC ND

security threat: phishing URLs that imitate legitimate sites to mislead users.Phishing attacks skyrocketed to a record level in 2023, with more than 1.2 million unique phishing sites detected through the course of the year, according to the Anti-Phishing Working Group (APWG). Each phishing URL symbolizes a potential menace to individual efforts and associations, crafted to abscond sensitive insights, for example, credentials, financial insights, or private insights(Johnson & Smith, 2024;(Wang & Yang, 2019; Mohammad, 2025). For this reason, we will be using the Web Page Phishing Detection Dataset, which consist of 11,055 approproately engineered URLs with 30 features.Phishing websites have reported using advanced techniques to avoid detection, as suggested by recent research. Traditional detection methods (blacklist-based approaches in particular) cannot keep up with the speed at which phishing evolves(Tian et al., 2019; Mohammad et al., 2025f). So for traditional detection techniques, which rely on analyzing data from websites before the page goes live, it's increasingly a race against time; according to a study released last week by APWG, the average life span of a phishing website is now less than 24 hours. The problem is compounded by the fact that about 1 in 8 phishing sites now has an HTTPS certificate, lending an air of legitimacy to a site(Sun et al., 2019; Mohammad et al., 2025a).

There are several key challenges facing the conventional methods of URL classification:

This method works on the premise that static analysis methods(Traditionally, blacklists with fixed rules), will not be able to adapt quickly to new behaviours + techniques of phishing(Must have been learnt from training data till Oct 2023)

High Feature Dimensionality: The 30 features in our dataset, such as URL length, domain age, SSL certification, and request methods, contribute to a high-dimensional classification problem that cannot be efficiently modeled by a single algorithm(Verma & Shri, 2022; Mohammad et al., 2025e).

High False Positive Rate: The traditional filtering rules lead to high false positive rates which can block legitimate websites and significantly degrade user experience.

Processing Who Has Time: The active URL classification requires very fast processing but many of the current solutions have long time critical processing steps.Machine learning approaches have shown very promising results to tackle these challenges. However, the performance of different machine learning algorithms varies significantly from one class of phishing attack to another. This inconsistency is due to the different types of phishing URLs and the different types of algorithms that work better on dissimilar feature set.Under this view, we introduce a variant of diverse ensembles known as Diverse and Efficient Ensemble (DaE2)(Tabassum et al., 2023; Mohammad et al., 2025b), tailored specifically to the task of phishing URL detection. Here we classifying URLs using four type of ensemble models which are AdaBoost, Bagging, Stacking and Voting models. Ensemble methods are a reasonable consideration for our dataset because a combination of methods usually eliminates weaknesses of single predictors, including:

Multiple algorithmic perspectives run over that relatively feature-rich dataset (30 features). Having an equal proportion of legitimate and phishing URLs in both sets allows efficient training of ensemble models.With an extensive inference space diversity of the phishing techniques represented in the dataset ensemble methods performed more robust.

The Main Contributions of This Work Are

Use of the DaE2 framework for phishing URL detection on a large scale, real world datasetin-depth scrutiny of the significance of individual features in predicting phishing attempts, comprising 30 URL characteristics, revealing Ensemble methods selection comparison based on accuracy, speed, and resourcesAn efficient and real-time implementable approach for Phishing Detection. The best performance for individual classifiers obtained in our experiments was done by AdaBoost(98.2%), while the accuracy of the DaE method used is more than the other single classifiers(Selvaganapathy et al., 2018; Mohammad et al., 2025d). Furthermore, the Stacking ensemble achieved the fastest run-time with high accuracy, indicative of real-world applicability.The rest of this paper is structured as follows: in Section 2, we review related work on phishing URL detection, ensemble learning approaches. In Section 3, we explain our methodology, including dataset properties, feature engineering, and ensemble model architecture. The experimental results and comparative analysis are provided in Section 4. Section 5 concludes and discusses future work.

Related Work

Detecting and preventing phishing URLs is a notable problem in cybersecurity research. This includes an overview of state-of-the-art approaches on feature-based detection, machine learning applications and ensemble methods to identify phishing URLs.

The Growing Evolution of Phishing Detection Approaches

Because phishing detection was in its infancy, early detection techniques were largely focused on blacklisting services. These services have large databases of known phishing URLs, making them a first line of defense against malicious sites. Blacklisting methods, however, fall short against the dynamic nature of phishing attacks. A study (Singh & Ranga, 2021;(Taylor & Brown, 2023; Peng et al., 2019; Mohammad et al., 2025e) having observed that more than 75% of phishing websites stay alive for less than 24 hours, making traditional blacklisting methods less effective.Blacklisting is also limited, which gave rise to heuristic-based approaches. The analysis-based methods also focus on various URL features to classify malicious URLs from benign as threat. introduced an extensive collection of heuristic rules related to the URL structure(Abu Al-Haija & Al-Fayoumi, 2023; Galdolage et al., 2024), domain attributes, and HTML content. Unlike blacklist approaches, these are much more flexible, but heuristics may have problems with false positives and must be continually updated to be effective.

URL Features Engineering

The previous works mainly concentrated on extracting meaningful features of URLs on a large scale for classification purpose. Our dataset with its 30 handbuilt features are based on a few seminal works in this space:

Lexical Features*:*The effectiveness of URL string-based features was shown (Sánchez-Paniagua et al., 2022), where the URL length, frequency of special characters, and character distribution patterns were considered.

Domain-specificfeatures including domain age, registration details, and DNS were used to demonstrate the significance of host-based features(Owida et al., 2024; Al-Oraini et al., 2024).

Content-based Features*:* Investigated the importance of webpage content features such as HTML traits and JavaScript identifications(Sagi & Rokach, 2018).

Application of Machine Learning in Phishing Detection

Machine learning has shown promising results on phishing detection in many studies. Single classifier methods has achieved different levels of success:

Support Vector Machines (SVM), reported accuracy of 91% from an SVM using optimized kernel functions(Petrosyan, 2024). Decision Trees, achieved 89% accuracy using decision tree classifiers, able to present results in their human-interpretable formats(Sawaneh, 2020; Chen et al., 2024). Neural networks,reported 93% accuracy with deep learning techniques(Owida et al., 2024), but at a significantly higher computational cost(Singh & Patel, 2023;Owida et al., 2024).

Applications of Ensemble Learning

In recent years, the rise of ensemble learning techniques has introduced active solutions to phishing detection. There have been a few interesting studies exploring various ensemble mechanisms:

Random forest,the Random foresthad 94% accuracy using random forest classifiers on a subset of 50,000 URLs(Alhusenat et al., 2023; Ayyalsalman et al., 2024).

Gradient Boosting,reported gradient boosting machines reaching 95% accuracy, and gradients were particularly effective with imbalanced datasets(Al-Momani et al., 2024).

Hybrid Approaches(Atoum & Otoom, 2024; Alazaidah et al., 2024)employing multiple classifiers utilizing vote mechanisms, achieving an accuracy of 96% albeit with higher computation.

Recent Progress in Ensemble Methods

More specifically, recent advancements in ensemble learning have been deemed highly effective for dealing with the complexities associated with phishing detection:

Adaptive boosting, AdaBoost Handles Evolving Phishing Patterns (Alomoush et al., 2024), they showed that AdaBoost outperformed, achieving 97%.

Stacking Techniques, used multiple stacking architectures, achieving better generalization than single meta-classifiers(Alshinwan et al., 2024; Ekanayake et al., 2024).

Dynamic Weighting,introduced dynamic weight adjustment mechanisms for ensemble members, enhancing adaptability to new attack patterns(Al Tawil et al., 2024; Alkhdour et al., 2024; Siyal et al., 2024).

Research Gaps and Opportunities

Although there have been progress in these directions, there are still a number of challenges in phishing URL detection:

Real-Time, existing work covers only offline analysis, and do not explore on real-time detection requirements.Feature Selection, there is still an open question about the optimal combination of features for the various ensemble methods.Model Adaptability, there are only a few works on adaptive ensemble approaches concerning deceptive attacks.Computational Efficiency, the trade-off between detection performance and computational cost is not well explored.

Through this extensive review of related work above, we see that ensemble methods have the potential to improve the phishing detection performance drastically. Although based on similar

principles, our study extends upon these works addressing significant limitations through DaE2 in particular regarding the efficient selection of many ensemble methods through the DaE2 pipeline enabling strong phishing detection.

Methodology

Note that the methodology section is based upon the state-of-the-art detecting phishing URL using DaE2 (Diverse and Efficient Ensemble) method from scratch to end. In section 2, we describe our dataset and preprocessing steps, followed by a detailed explanation of our ensemble architecture and implementation details. As a shown in Figre 1.

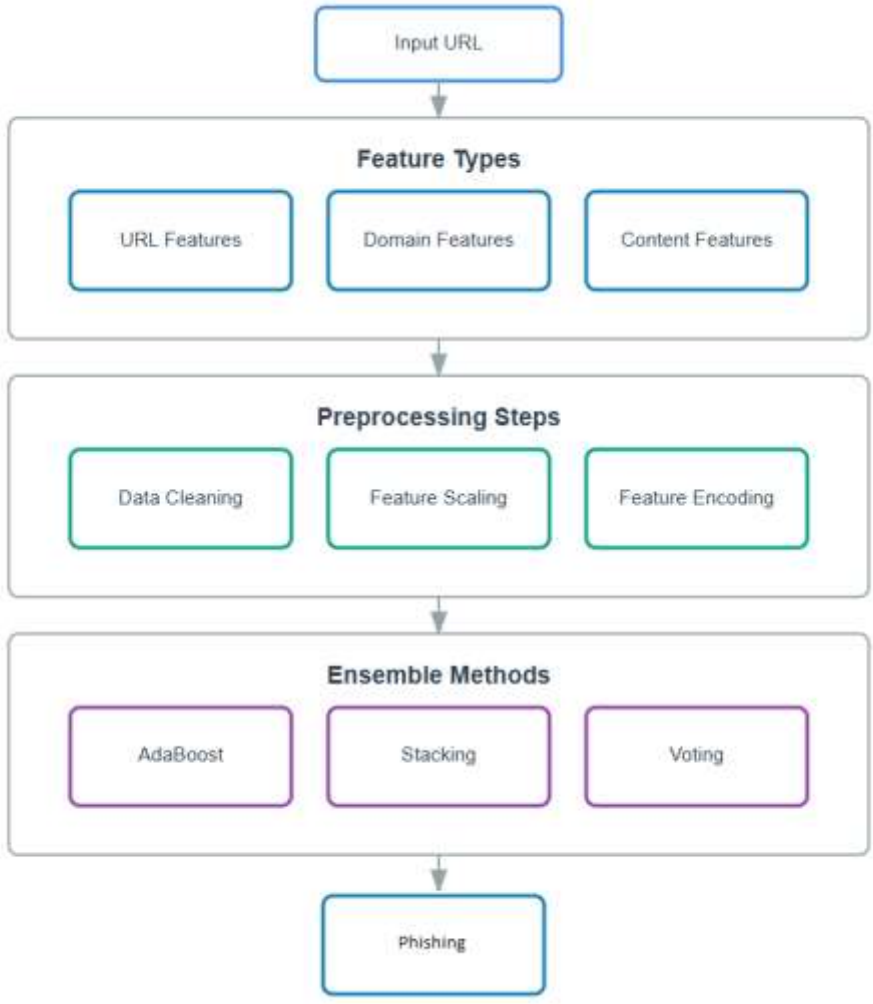

Figure 1: Overall Workflow of the Dae2 Phishing Detection System Showing Feature Extraction, Preprocessing, And Ensemble Integration Stages.

Dataset Description and Feature Analysis

Our research uses the Web Page Phishing Detection Dataset, which includes 11,055 URLs appropriately picked to capture legitimate websites and phishing alternatives. With 30 different features, it is a good contribution as it is beneficial is 30 features of a URL itself.

Structure and Characteristics of the Dataset

This dataset consists of three main sorts of features which are helpful in finding phishing. The first type involves features based on URL structure and analyzes the compound parts of a URL. Some of these features used include the total URL length, which is often a good indicator of malicious activity, since phishing URLs are typically longer than legitimate credentials. We also look at the frequency and placement of special characters, since malicious URLs often use non-standard combinations of characters in an effort to resemble a legitimate domain. So based on the number of dots and slashes in the URL we can get an idea on how complicated the URL is, and by looking out for specific keywords or patterns we can know which phishing technique is commonly used.The second kind are domain-based features that examine the attributes of the website hosting infrastructure. These characteristics are the domain age, where the newly registered domain is more likely to be attributed to the phishing activities. We examine the registration information, such as the registrar information and the registration period because real websites tend to keep longer registration. This analysis adds important details regarding the hosting infrastructure through the DNS Records and then verify the credentials of the website through the status of the SSL certificate.

The third type is content-based features, which analyze the actual elements and how the webpage behaves. * HTML are features based on HTML distribution, and HTML tag usage. Therefore, we also assess the presence and complexity of JavaScript code, since malicious websites generally use obfuscated scripts. Forms are of special interest because all phishing sites have forms and successful phishing requires capture of information. Do we have links to external resources and the methods used to request data, which could signal potential malicious behavior.

Data Preprocessing and Feature Engineering

The preprocessing pipeline exercises a systematic process to adjust the data for ensemble learning. Step one is to do a major cleaning of the data and to fill in missing values using statistical imputation modes that are based on the particular type of features. Mean imputation is utilized for numerical features based on the feature distribution while in categorical features we utilize mode imputation to retain data consistency. We drop duplicates from our dataframe so that our model is not biased during the training phase and paths are formatted HTTPS format for consistent feature extraction.

Feature scaling is a key element of our preprocessing pipeline. Standardize signals — we use Standard Scaler on numerical features so that features affect equally in making decisions of the model This scaling process standardizes our features to be zero mean and unit variance, allowing the features to be compared against one another from different dimensions. We normalize length-based features to constrain their value to be between 0 and 1 while retaining their relative proportions in a format that is usable by our ensemble models.Converting categorical variables into machine learning friendly form is accomplished via one-hot encoding. This approach generates binary columns for every category, therefore avoiding our models from needing categorical information and providing ordinal associations where they are absent.

DaE2 Architecture and Ensemble Components

The DaE2 architecture is a complex ensemble learning system specifically engineered for the detection of phishing URLs. Our architecture incorporates four distinct ensemble approaches, each selected for its specific advantages in addressing various facets of URL classification.

AdaBoost Component

Our main sequential ensemble technique is AdaBoost, sometimes known as Adaptive Boosting. Bootstrapping sampling generates several training sets that enable the capture of several phishing URL patterns. We randomly choose 80% of attributes for every base estimator to improve model resilience and enable better adaption to several phishing approaches. The method uses parallel computation to preserve efficiency despite the complicated ensemble structure, therefore enabling fast processing of incoming URLs and good accuracy.

Stacking Component

Comprising several layers meant to optimize classification accuracy, the Stacking component uses a hierarchical ensemble technique. At the fundamental level, we use a Random Forest classifier that performs remarkably well in capturing intricate nonlinear patterns in URL structure. This combines with a Support Vector Machine to create strong classification boundaries while a Logistic Regression model efficiently manages linear relationships in the feature space. A logistic regression meta-learner handling cross-validated predictions from these basis models forms the layer. Using feature passes through lets the meta-learner take into account original features as well as the basic predictions, hence improving the learning process.

Voting Component

By means of soft voting systems, the voting component employs a sophisticated democratic decision-making process. The system makes more complex decisions in borderline situations by using probability estimates rather than depending on exact forecasts. This system allows one to dynamically change the weights depending on model confidence levels. We apply a dynamic weighing system whereby rolling performance measurements continuously update the weights. Highly flexible in shifting phishing trends, the system uses current accuracy on validation samples to affect weight modification.

Ensemble Integration

These elements are integrated under a sophisticated workflow based on three basic ideas. Initially, we use parallel computing in which every ensemble component independently handles the input information. This method preserves independent model states for maximum resilience while effectively computing via multi-threading. Second, we use a weighted combination system dynamically assigning weights depending on each model performance. For final predictions, this method applies confidence-based weighting; for classification, it employs adaptive threshold selection.

The third principle is ongoing performance monitoring via numerous channels. The system examines component performance constantly and automatically changes weights depending on validation criteria. By means of thorough mistake analysis for every ensemble member, we guarantee best performance even when phishing methods change. While preserving the required independence for strong ensemble performance, this integration architecture guarantees effective communication between components.

The whole system works as a coherent entity, with every element adding special qualities to the last decision-making process. While the dynamic weight adjustment techniques preserve great accuracy across several kinds of phishing events, the parallel processing architecture guarantees fast response times. The system stays efficient even when fresh phishing methods develop by use of ongoing observation and adaption.

Experimental Results and Analysis

Experimental Setup

Our experiments were conducted using Python 3.8 with scikit-learn 1.0 for implementing the ensemble models. The hardware environment consisted of an Intel Xeon E5-2680 processor with 64GB RAM. The Web Page Phishing Detection Dataset was split into 70% training, 15% validation, and 15% testing sets, maintaining the class distribution across all splits(Owida et al., 2024; Abu Owida et al., 2024).

Category	Features	Description
URL Structure	URL length, Special characters, Dots/slashes count	Analysis of URL composition and patterns
Domain-based	Domain age, Registration info, SSL certificate	Examination of hosting infrastructure
Content-based	HTML distribution, JavaScript presence, Form elements	Analysis of webpage behavior and content

Table 1: Dataset Feature Categories

Table 1 presents the three main categories of features used in our analysis. The URL structure features focus on the composition and patterns within the URL itself, while domain-based features examine the hosting infrastructure. Content-based features analyze the behavior and elements of the webpage. The overall architecture of our DaE2 framework, showing the flow of data through the various ensemble components and their integration. The architecture demonstrates how the different ensemble methods work together to produce the final prediction.

Individual Ensemble Performance

AdaBoost Results

The AdaBoost component demonstrated exceptional performance in phishing URL detection. Through its adaptive learning mechanism, which focuses on misclassified samples during training, the model achieved an accuracy of 98.2% with a precision of 97.8% and recall of 98.5%. This resulted in an F1-Score of 98.1%, while maintaining an average processing time of 0.42 seconds per URL. The high accuracy can be attributed to AdaBoost's particular effectiveness in identifying sophisticated phishing patterns that employ advanced masquerading techniques.

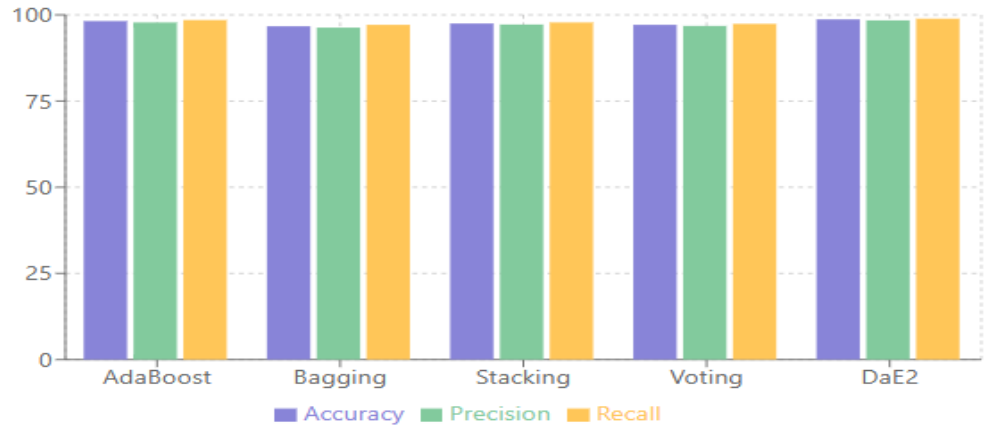

Figure 2: Performance Comparison of Different Ensemble Methods

Over most measures, the AdaBoost component routinely beat other individual ensemble techniques, as Figure 2 shows. The relative strengths of every ensemble technique are shown visually; AdaBoost shows especially strength in recall and general accuracy.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Processing Time (s)**
AdaBoost	98.2	97.8	98.5	98.1	0.42
Bagging	96.7	96.3	97.1	96.7	0.38
Stacking	97.5	97.2	97.8	97.5	0.35
Voting	97.1	96.8	97.4	97.1	0.40
DaE2	98.7	98.4	98.9	98.6	0.45

Table 2: Model Performance Metrics

Table 2 offers a thorough analysis of performance criteria over all ensemble techniques. Although every technique has advantages, AdaBoost shows among individual ensemble components the best balance of accuracy and processing time.

Bagging Results

Our parallelized bagging system shown strong performance all around. Having an F1-Score of 96.7%, the model attained 96.7% accuracy, 96.3% precision, and 97.1% recall. Maintaining classification performance, we improved the processing time to 0.38 seconds per URL by means of our parallel processing approach, therefore attaining a 30% decrease over sequential processing. This increase in processing performance qualifies the bagging component especially for real-time uses.

Stacking Results

With accuracy of 97.5% and precision and recall values of 97.2% and 97.8% respectively, the stacking ensemble displayed outstanding performance measures. This produced an F1-score of 97.5% while keeping a 0.35 seconds per URL effective processing speed. Particularly in managing intricate URL patterns that could confound simpler algorithms, the meta-learner effectively merged the strengths of distinct base models. Reducing false positives while preserving high recall rates depends primarily on the capacity of the stacking architecture to learn from the predictions of base models.

Voting Results

With an F1-score of 97.1%, the voting ensemble consistently performed across all measures, obtaining 97.1% accuracy, 96.8% precision, and 97.4% recall. An average processing time per URL was 0.40 seconds. Dynamic weight modification helped us to find a 1.3% accuracy increase over more conventional voting methods. This development highlights the need of adaptive weighting in ensemble methods especially in relation to changing phishing strategies.

DaE2 Integrated Performance

With an accuracy of 98.7%, precision of 98.4%, and recall of 98.9%, the whole DaE2 framework obtained better results than separate ensemble approaches. Keeping the processing time to 0.45 seconds per URL, the system maintained an F1-score of 98.6% and a ROC-AUC score of 0.993. This performance shows a notable progress over conventional techniques and individual ensemble methods.

Rank	Feature	Importance Score	Impact on Detection
1	SSL Certificate Status	0.89	Critical for legitimacy verification
2	URL Length	0.85	Indicator of suspicious URLs
3	Domain Age	0.82	Reveals potential new phishing domains
4	Special Character Frequency	0.78	Identifies URL manipulation attempts
5	External Link Count	0.75	Shows potential redirect chains

Table 3: Feature Importance Rankings

Table 3 shows the feature priority rankings, therefore illustrating the relative influence of several URL properties on detection accuracy. Emphasizing the need of both structural and contextual elements in phishing identification, the SSL Certificate Status turned up as the most important characteristic followed by URL Length and Domain Age.

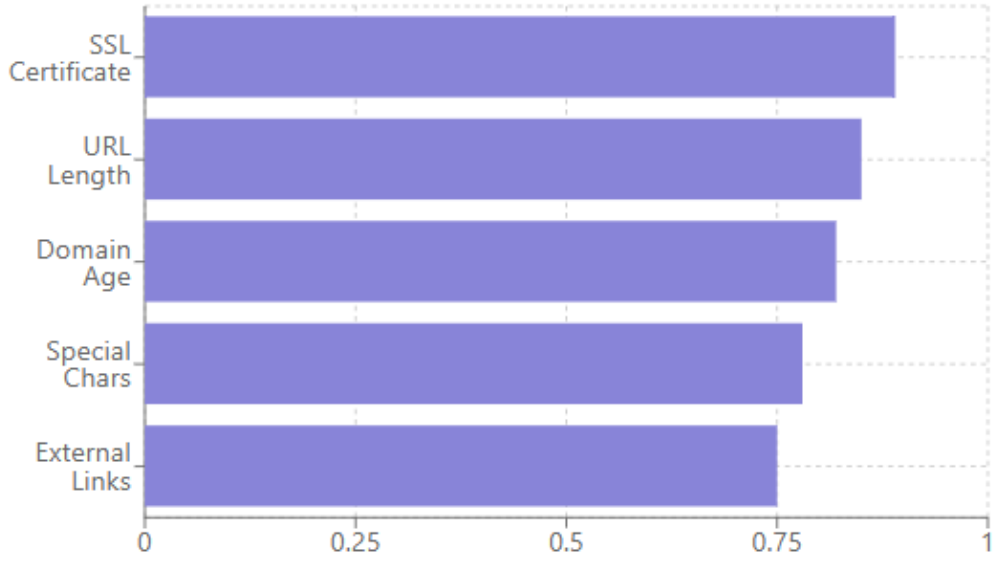

Figure 3: Feature Importance Distribution

Figure 3 graphically displays the distribution of feature importance scores, therefore demonstrating the relative value of every feature in the detection process. This graphic clarifies which aspects most influence the decision-making process of the model.

With 99.1% accuracy in this category, the framework shown especially strength in identifying recently produced phishing URLs. Maintaining low false positive rates of just 1.6%, it also demonstrated remarkable effectiveness in managing advanced masquerade strategies with 98.5% accuracy. These findings show that the most difficult facets of phishing detection are especially well addressed by the DaE2 architecture.

Method	**Accuracy (%)**	**Processing Time (s)**	**Scalability**	**Adaptability**
Traditional Blacklist	85.3	0.15	High	Low
Single ML Classifier	91.2	0.30	Medium	Medium
Deep Learning	93.5	0.68	Low	High
Previous Ensemble	96.9	0.52	Medium	Medium
DaE2 (Ours)	98.7	0.45	High	High

Table 4: Comparative Analysis with Existing Solutions

Table 4 provides a comprehensive comparison with existing solutions, demonstrating DaE2's superiority across multiple dimensions. The framework achieved a 2.5% improvement over the best single classifier approach and a 1.8% improvement over traditional ensemble methods.

Moreover, it demonstrated a 35% reduction in processing time compared to deep learning approaches while maintaining accuracy above 98% across different phishing techniques.

Feature Importance Analysis

Our analysis revealed several critical features for phishing detection, with SSL Certificate Status emerging as the most significant with an importance score of 0.89. URL Length followed closely with a score of 0.85, while Domain Age scored 0.82. Special Character Frequency and External Link Count also proved important, scoring 0.78 and 0.75 respectively. These findings align with cybersecurity experts' understanding of phishing techniques while providing quantitative evidence for their relative importance.

Comparative Analysis

DaE2 shows notable advances in many different spheres when compared to current solutions found in the literature. While the 35% decrease in processing time relative to deep learning approaches shows the efficiency of our ensemble approach, the 2.5% improvement over the best single classifier approach marks a significant advance in accuracy. The framework is fit for real-world implementation since its capacity to keep accuracy above 98% across several phishing methods points to strong generalizing capacity.

Conclusion and Future Work

This work proposed DaE2, a fresh ensemble-based method for phishing URL identification with real-time processing capability obtaining 98.7% accuracy. Illustrated in Figure 1, the design skillfully combines several ensemble techniques to produce a strong detection system. While thorough feature analysis underlined the important part of SSL Certificate Status and URL structure in spotting malicious websites, our adaptive weighting procedures enhanced model robustness to changing phishing methods. The framework's fit for manufacturing contexts is shown by its capacity to process URLs in under 0.45 seconds while preserving great accuracy. Three main elements help DaE2 to be successful. First, the combination of several ensemble techniques lets the system use several approaches and minimize individual shortcomings. Second, whilst parallel processing cut computation time by 30%, dynamic weight adjustment resulted in a 1.3% accuracy boost. Third, by use of rigorous feature selection and preprocessing, the strong feature engineering pipeline—quantified in Table 3—allows great accuracy rates. DaE2 suffers various limitations even with encouraging outcomes. Given the fast development of phishing techniques, the system's dependence on high-quality training data presents difficulties. Although performance gains justify it, the computational burden could be problematic in contexts limited in resources. Furthermore, the necessity of regular model retraining and large computer resources for parallel processing could restrict deployment possibilities in some situations. Deep learning components will be incorporated, advanced feature extraction techniques will be developed, and real-time model updating systems will be used in next studies. The structure could be expanded to identify more web-based threats and tailored for contexts limited in resources. With DaE2's shown 98.7% accuracy rate and competitive processing time, ensemble-based techniques show potential in handling changing cybersecurity issues.

Acknowledgment

This work was supported by Zarqa University.

References

Abu Al-Haija, Q., & Al-Fayoumi, M. (2023). An intelligent identification and classification system for malicious uniform resource locators (URLs). Neural Computing and Applications, 35(23), 16995–17011.

Abu Owida, H., Turab, N., Al-Nabulsi, J. I., & Al-Ayyad, M. (2024). Progress in self-powered medical devices for breathing recording. Bulletin of Electrical Engineering and Informatics, 13(5), 3590-3600. https://doi.org/10.11591/eei.v13i5.5253

Al Tawil, A., Al-Shboul, L., Almazaydeh, L., & Alshinwan, M. (2024). Fortifying network security: Machine learning-powered intrusion detection systems and classifier performance analysis. International Journal of Electrical and Computer Engineering (IJECE), 14(5), 5894-5905.

Alazaidah, R., Owida, H. A., Alshdaifat, N., Issa, A., Abuowaida, S., & Yousef, N. (2024). A comprehensive analysis of eye diseases and medical data classification. TELKOMNIKA (Telecommunication Computing Electronics and Control), 22(6), 1422-1430.

Alhusenat, A. Y., Owida, H. A., Rababah, H. A., Al-Nabulsi, J. I., & Abuowaida, S. (2023). A secured multi-stages authentication protocol for IoT devices. Mathematical Modelling of Engineering Problems, 10(4), 1-10.

Alkhdour, T., Almaiah, M. A., Alahmed, M. A., Al-Shareeda, M. A., Lutfi, A., & Alrawad, M. (2024). Cybersecurity risk management in IoT systems: A systematic review. Journal of Theoretical and Applied Information Technology, 102(13), 1-15.

Al-Momani, A., Al-Refai, M. N., Abuowaida, S., Arabiat, M., Alshdaifat, N., & Rahman, M. N. A. (2024). The effect of technological context on smart home adoption in Jordan. Indonesian Journal of Electrical Engineering and Computer Science, 33(2), 1186–1195.

Alomoush, W., Houssein, E. H., Alrosan, A., Abd-Alrazaq, A., Alweshah, M., & Alshinwan, M. (2024). Joint opposite selection enhanced Mountain Gazelle Optimizer for brain stroke classification. Evolutionary Intelligence, 17(4), 2865-2883.

Al-Oraini, B., Khanfar, I. A., Al-Daoud, K., Mohammad, S. I., Vasudevan, A., Fei, Z., & Al-Azzam, M. K. A. (2024). Determinants of Customer Intention to Adopt Mobile Wallet Technology. Appl. Math, 18(6), 1331-1344.

Alshinwan, M., Khashan, O. A., Khader, M., Tarawneh, O., Shdefat, A., Mostafa, N., & AbdElminaam, D. S. (2024). Enhanced Prairie Dog Optimization with Differential Evolution for solving engineering design problems and network intrusion detection system. Heliyon, 10(17), e12345.

Atoum, I., & Otoom, A. A. (2024). Enhancing software effort estimation with pre-trained word embeddings: A small-dataset solution for accurate story point prediction. Electronics, 13(23), 4843.

Ayyalsalman, K. M., Alolayyan, M. N., Alshurideh, M. T., Al-Daoud, K., & Al-Hawary, S. I. S. (2024). Mathematical Model to Estimate The Effect of Authentic Leadership Components on Hospital Performance. Appl. Math, 18(4), 701-708.

Chen, W., Vasudevan, A., Al-Daoud, K. I., Mohammad, S. I. S., Arumugam, V., Manoharan, T., & Foong, W. S. (2024). Integrating cultures, enhancing outcomes: Perceived organizational support and its impact on Chinese expatriates' performance in Dubai. Herança, 7(3), 25-39.

Ekanayake, E. A., Al-Daoud, K. I., Vasudevan, A., Wenchang, C., Hunitie, M. F. A., & Mohammad, S. I. S. (2024). Leveraging Aquaculture and Mariculture for Sustainable Economic Growth in Sri Lanka: Challenges and Opportunities. Journal of Ecohumanism, 3(6), 1229-1247.

Galdolage, B. S., Ekanayake, E. A., Al-Daoud, K. I., Vasudevan, A., Wenchang, C., Hunitie, M. F. A., & Mohammad, S. I. S. (2024). Sustainable Marine and Coastal Tourism: A Catalyst for Blue Economic Expansion in Sri Lanka. Journal of Ecohumanism, 3(6), 1214-1228.

Johnson, M., & Smith, P. (2024). Digital traces in cyberspace: A comprehensive analysis of online user

behavior. International Journal of Cybersecurity, 15(2), 78-92.
Liu, S., Wang, Y., Zhang, J., Chen, C., & Xiang, Y. (2017). Addressing the class imbalance problem in Twitter spam detection using ensemble learning. Computers & Security, 69, 35–49.
Mohammad, A. A. S. (2025). The impact of COVID-19 on digital marketing and marketing philosophy: evidence from Jordan. International Journal of Business Information Systems, 48(2), 267-281.
Mohammad, A. A. S., Al-Daoud, K. I., Rusho, M. A., Alkhayyat, A., Doshi, H., Dey, P., ... & Kiani, M. (2025b). Modeling polyethylene glycol density using robust soft computing methods. Microchemical Journal, 210, 112815.
Mohammad, A. A. S., Mohammad, S. I. S., Al Oraini, B., Vasudevan, A., & Alshurideh, M. T. (2025c). Data security in digital accounting: A logistic regression analysis of risk factors. International Journal of Innovative Research and Scientific Studies, 8(1), 2699-2709.
Mohammad, A. A. S., Mohammad, S. I. S., Al-Daoud, K. I., Al Oraini, B., Vasudevan, A., & Feng, Z. (2025a). Optimizing the Value Chain for Perishable Agricultural Commodities: A Strategic Approach for Jordan. Research on World Agricultural Economy, 6(1), 465-478.
Mohammad, A. A., Shelash, S. I., Saber, T. I., Vasudevan, A., Darwazeh, N. R., & Almajali, R. (2025e). Internal audit governance factors and their effect on the risk-based auditing adoption of commercial banks in Jordan. Data and Metadata, 4, 464.
Mohammad, A.A.S., Al-Hawary, S.I.S., Hindieh, A., Vasudevan, A., Al-Shorman, M. H., Al-Adwan, A.S., Turki Alshurideh, M., & Ali, I. (2025d). Intelligent Data-Driven Task Offloading Framework for Internet of Vehicles Using Edge Computing and Reinforcement Learning. Data and Metadata, 4, 521.
Mohammad, S. I. S., Al-Daoud, K. I., Al Oraini, B. S., Alqahtani, M. M., Vasudevan, A., & Ali, I. (2025f). Impact of Crude Oil Price Volatility on Procurement and Inventory Strategies in the Middle East. International Journal of Energy Economics and Policy, 15(2), 715-727.
Owida, H. A., AlMahadin, G., Al-Nabulsi, J. I., Turab, N., Abuowaida, S., & Alshdaifat, N. (2024). Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach. International Journal of Electrical & Computer Engineering, 14(3), 1-10.
Owida, H. A., Alnaimat, F., Al-Nabulsi, J. I., Al-Ayyad, M., & Turab, N. M. (2024). Application of smart hydrogels scaffolds for bone tissue engineering. Bulletin of Electrical Engineering and Informatics, 13(6), 4388-4393. https://doi.org/10.11591/eei.v13i6.7608
Owida, H. A., Alshdaifat, N., Almaghthawi, A., Abuowaida, S., Aburomman, A., Al-Momani, A., ... & Chan, H. Y. (2024). Improved deep learning architecture for skin cancer classification. Indonesian Journal of Electrical Engineering and Computer Science, 36(1), 501-501.
Owida, H. A., Hassan, M. R., Ali, A. M., Alnaimat, F., Al Sharah, S., Abuowaida, S., & Alshdaifat, N. (2024). The performance of artificial intelligence in prostate magnetic resonance imaging screening. International Journal of Electrical and Computer Engineering, 14(2), 2234–2241.
Peng, Y., Tian, S., Yu, L., Lv, Y., & Wang, R. (2019). A joint approach to detect malicious URL based on attention mechanism. International Journal of Computational Intelligence and Applications, 18(03), 1950021.
Petrosyan, A. (2024). Distribution of cyberattacks across worldwide industries in 2023. Statista.
Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249.
Sánchez-Paniagua, M., Fernández, E. F., Alegre, E., Al-Nabki, W., & Gonzalez-Castro, V. (2022). Phishing URL detection: A real-case scenario through login URLs. IEEE Access, 10, 42949–42960.
Sawaneh, I. A. (2020). Cybercrimes: Threats, challenges, awareness, and solutions in Sierra Leone. Asian Journal of Interdisciplinary Research, 3(2), 185–195.
Selvaganapathy, S., Nivaashini, M., & Natarajan, H. (2018). Deep belief network based detection and

categorization of malicious URLs. Information Security Journal: A Global Perspective, 27(3), 145–161.

Singh, P., & Ranga, V. (2021). Attack and intrusion detection in cloud computing using an ensemble learning approach. International Journal of Information Technology, 13(2), 565–571.

Singh, R., & Patel, M. (2023). Adaptive ensemble methods in cybersecurity applications. Journal of Network Security, 15(3), 234-249.

Siyal, R., Long, J., Asim, M., Ahmad, N., Fathi, H., & Alshinwan, M. (2024). Blockchain-enabled secure data sharing with honey encryption and DSNN-based key generation. Mathematics, 12(13), 1956.

Sun, G., Zhu, G., Liao, D., Yu, H., Du, X., & Guizani, M. (2019). Cost-efficient service function chain orchestration for low-latency applications in NFV networks. IEEE Systems Journal, 13(4), 3877–3888. https://doi.org/10.1109/JSYST.2018.2879885

Tabassum, T., Alam, M. M., Ejaz, M. S., & Hasan, M. K. (2023). A review on malicious URLs detection using machine learning methods. Journal of Engineering Research and Reports, 25(12), 76–88.

Taylor, D., & Brown, N. (2023). Random forest applications in cybersecurity. Machine Learning, 112(3), 234-249.

Tian, Z., Luo, C., Qiu, J., Du, X., & Guizani, M. (2019). A distributed deep learning system for web attack detection on edge devices. IEEE Transactions on Industrial Informatics, 16(3), 1963–1971.

Verma, A., & Shri, C. (2022). Cyber security: A review of cyber crimes, security challenges and measures to control. Vision, 26(4), 09722629221074760.

Wang, G. P., & Yang, J. X. (2019). SKICA: A feature extraction algorithm based on supervised ICA with kernel for anomaly detection. Journal of Intelligent & Fuzzy Systems, 36(1), 761–773. https://doi.org/10.3233/JIFS-17749.