Authors	Title	Journal Name	Publishing House	Year	Volume	Issue	Problem Solved	Method Used to Solve the Problem	Data & Its Availability
Diana T. Mosa; Mahmoud Y. Shams; Amr A. Abohany; El-Sayed M. El-kenawy; M. Thabet	Machine Learning Techniques for Detecting Phishing URL Attacks	Computers, Materials & Continua	Not specified	2023	75	1	Detect phishing URL attacks in a cyber-security context by identifying malicious websites that mislead users	Survey and experimental evaluation using machine learning models (Neural Networks, Naïve Bayes, and Adaboost) applied on website features extracted from a Kaggle dataset	Kaggle dataset comprising over 11,000 website URLs, available in text and CSV formats; each sample includes 30 features and a class label (1 or –1)

Authors	Title	Journal Name	Publishing House	Year	Volume	Issue	Problem Solved	Method used to solve the problem	Data & its availability

Abdul Karim, Samir Brahim Belhaouari, Mobeenshahrooz, Khabib Mustofa, Ands. Ramanakumarjoga	Phishing Detection System Through Hybrid Machine Learning Based on URL	IEEE Access (inferred from DOI)	IEEE	2023	11	Not specified	Detect phishing attacks via URL classification to protect users from cybercrimes and online fraud.	A hybrid machine learning approach that employs multiple algorithms (decision tree, linear regression, random forest, naive Bayes, gradient boosting classifier, K-neighbors classifier, support vector classifier) alongside a novel hybrid	A publicly available phishing URL dataset from Kaggle consisting of 11,054 records with 33 attributes extracted from over 11,000 websites (including both phishing and legitimate URLs).
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method Used to Solve the Problem**	**Data & Its Availability**

Authors	Title	Journal Name	Publishing House	Year	Volume	Issue	Problem Solved	Method used to solve the problem	Data & its availability
Kanishka Misra, Julia Taylor Rayz	LMs go Phishing: Adapting Pre-trained Language Models to Detect Phishing Emails	2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)	IEEE	2022	N/A	N/A	Addresses the persistent challenge of phishing in email communications by improving detection methods—especially handling the gap between in-domain and out-of-domain emails.	Adapts pre-trained GPT2 models to the email domain by (a) fine-tuning them on a large corpus of ~725k training emails with a classification objective and (b) employing an in-context priming approach to facilitate zero-shot-like	Aggregated from several well-known sources: legitimate emails from Enron, Avocado, and IWSPA-AP; phishing emails from Nazario, FRAUD, and UNTROUBLED. After rigorous pre-processing and deduplication, about

Kunle Oloyede, Chinenye Obunadike, Simo Yufenyuy Simo, Emmanuel Elom, Abdul-Waliyyu Bello, Somtobe Olisah, Callistus Obunadike, Oluwadamilola Ogunleye, Sulaimon Adeniji	Impact Of Web (URL) Phishing and Its Detection	International Journal of Scientific Research and Management (IJSRM)	IJSRM (via https://ijsrm.net)	2024	12	4	Mitigate the risks posed by web phishing attacks by identifying fraudulent URLs and phishing emails in order to protect sensitive information.	A hybrid approach that involves collecting phishing data (via web scraping and public datasets like the Phish Tank Database), data cleaning and feature extraction using Python, followed by applying machine learning models (such as	Data are drawn from publicly available phishing datasets and APIs; while several sources (e.g., Phish Tank Database) are mentioned, the article does not provide direct links to specific open datasets.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Samer Atawneh; Hamzah Aljehani	Phishing Email Detection Model Using Deep Learning	Electronics	MDPI, Basel, Switzerland	2023	12	N/A	Detect and prevent email phishing attacks	Deep learning models—including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, recurrent neural networks (RNNs), and bidirectional encoder representations from transformers (BERT) using natural language	A dataset consisting of phishing and benign emails was used; details on public availability are not specified
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Mohamed Hassan	Comparative Analysis of Deep Learning Algorithms for Phishing Email Detection (a.k.a. “Evaluation of Deep Learning Algorithms in Comparison for Phishing Email Identification”)	Applied Mathematics on Science and Engineering	Not specified (self-archived under CC BY)	2024	1	1	Detecting and accurately classifying phishing emails in an inherently imbalanced dataset	Preprocessing text (tokenization, padding, label-encoding) followed by employing deep learning architectures—specifically CNN, a hybrid CNN-RNN, and RCNN for binary classification	A dataset comprising 13,055 training samples and 5,595 testing samples of emails is used. The study highlights an imbalance in the “Email Type” labels, but does not specify public availability.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Authors	Title	Journal Name	Publishing House	Year	Volume	Issue	Problem Solved	Method used to solve the problem	Data & its availability
Parisa Mehdi Gholampour, Rakesh M. Verma	Adversarial Robustness of Phishing Email Detection Models	Proceedings of the 9th ACM International Workshop on Security and Privacy Analytics (IWSPA ’23)	ACM, New York, NY, USA	2023	N/A	N/A	Enhance the robustness of phishing detection models against adversarial attacks and imbalanced datasets	Generation of adversarial examples using TextAttack techniques (Textfooler, PWWS, DeepWordBug, BAE), fine-tuning GPT-2 for synthetic phishing email generation, and applying a K-Nearest Neighbor defense to correctly reclassify adversarial	Utilized the public IWSPA 2.0 phishing/legitimate email dataset (initially composed of 629 phishing and 5092 legitimate emails, later refined), with the augmented adversarial and synthetic datasets publicly available on GitHub (https://github.com/ReDASers

Dam Minh Linh, Ha Duy Hung, Han Minh Chau, Quang Sy Vu, Thanh-Nam Tran	Real-time phishing detection using deep learning methods by extensions	International Journal of Electrical and Computer Engineering (IJECE)	Not specified	2024	14	3	Provides real-time detection and prevention of phishing attacks by detecting malicious URL links via a browser extension.	A browser extension is developed that integrates deep learning—primarily a convolutional neural network (CNN)—to classify URLs in real time. The paper also compares various ML models (e.g., LR, DT, RF, SVM, CNN, and CNN-LSTM) and employs a	Uses a large malicious URL dataset containing 651,191 samples compiled from five benchmark sources, with a detailed breakdown (benign, defacement, phishing, malware) provided in Table 1.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

P.C.R. Chinta; C.S. Moore; L.M. Karaka; M. Sakuru; V. Bodepudi; S.R. Maka	Building an Intelligent Phishing Email Detection System Using Machine Learning and Feature Engineering	European Journal of Applied Science, Engineering and Technology	EJASET (via www.ejaset.com)	2025	3	2	Identifying and mitigating phishing email threats in digital communications	Comprehensive pipeline of data preprocessing (tokenization, stop word removal, etc.), feature engineering, and training multiple ML models (CNN, XGBoost, RNN, SVM) with the best performance achieved by a BERT-LSTM hybrid	A large-scale phishing email dataset combining phishing and legitimate emails, curated from sources such as Spam Assassin and the UCI ML library; public availability is not explicitly stated
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Safaa Magdy, Yasmine Abouelseoud, Mervat Mikhail	Efficient spam and phishing emails filtering based on deep learning	Computer Networks	Elsevier B.V.	2022	206	N/A (article ID 108826 is provided instead)	Tackles the threat of spam and phishing emails that waste bandwidth, jeopardize security, and incur financial losses	A deep learning–based neural network classifier incorporating feature extraction from emails, feature selection techniques (Low Variance, PCA, Chi-squared), and grid search hyperparameter tuning to discriminate among ham, spam, and phishing	Uses three benchmark datasets: SpamBase (from the UCI repository), CSDMC2010 spam corpus, and a merged Phishing_ corpus (from SpamAssassin and Nazario) which are publicly available
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Chidimma Opara, Yingke Chen, Bo Wei	Look before you leap: Detecting phishing webpages by exploiting raw URL and HTML characteristics	Expert Systems With Applications	Elsevier Ltd	2024	236	N/A	To develop a reliable mechanism for detecting phishing webpages—helping to prevent email/internet fraud—by addressing the limitations of manual feature extraction.	**WebPhish:** An end-to-end deep neural network that automatically learns feature representations from raw URLs and HTML content. It employs an embedding layer for characters (URL) and words (HTML), concatenates these representations, and	Experiments were conducted on a real-world phishing dataset, and the authors have made the dataset available to promote verification and replicability.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Authors	Title	Journal Name	Publishing House	Year	Volume	Issue	Problem Solved	Method used to solve the problem	Data & its availability
Liqun Yang; Jiawei Zhang; Xiaozhe Wang; Zhi Li; Zhoujun Li; Yueying He	An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features	Expert Systems With Applications	Elsevier Ltd.	2021	165	N/A (Article ID: 113863)	The paper addresses the challenge of detecting phishing websites efficiently, aiming to overcome the drawbacks of slow training and suboptimal detection accuracy in existing methods.	A novel non-inverse matrix online sequence extreme learning machine (NIOSELM) is proposed. This approach avoids matrix inversion using the Sherman–Morrison Woodbury equation and integrates online sequential learning. In addition,	Normal websites are sourced from Alexa's top 2000 and 58,000 DMOZ websites covering various sectors; phishing websites (5000 examples) are retrieved from PhishTank. Data is collected via a Python-based web crawler,

Jay Doshi, Kunal Parmar, Raj Sanghavi, Narendra Shekokar	A comprehensive dual-layer architecture for phishing and spam email detection	Computers & Security	Elsevier	2023	133	Unspecified	Tackles the dual challenge of spam and phishing email detection by addressing limitations in earlier studies that either focus on only one type of malicious email or use a single feature source (body or content).	Proposes a novel dual-layer architecture that employs deep learning methods—specifically using techniques such as Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN)—combined with	Uses real-world, publicly available datasets: phishing emails from Nazario's corpus and spam/ham emails from the Spam Assassin project, comprising a total of 8,218 raw emails.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Abdulla Al-Subaiey; Fatema Antora; Mohammed Al-Thani; Amith Khandakar; Naser Abdullah Alam; SM Ashfaq Uz Zaman; Kaniz	Novel interpretable and robust web-based AI platform for phishing email detection	Computers and Electrical Engineering	Elsevier Ltd.	2024	120	N/A	Tackles the persistent threat of phishing emails by overcoming prior limitations—namely, the dependence on proprietary datasets and the lack of real-world applicability in detection methods.	Merged multiple public phishing/spam email datasets followed by rigorous text preprocessing (cleaning, tokenization, and feature extraction via TF-IDF), training and evaluating machine learning models (SVM, Multinomial Naive Bayes,	Combined public dataset from sources including the Enron Phishing Email Dataset, CEAS 2008 Spam Challenge Corpus, Ling-Spam Corpus, Nazario Spam Dataset, Nigerian Fraud Dataset, and SpamAssassin Public Corpus
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

N. Swapna Goud, Dr. Anjali Mathur	Feature Engineering Framework to detect Phishing Websites using URL Analysis	International Journal of Advanced Computer Science and Applications (IJACSA)	IJACSA website (hosted on thesai.org) or not explicitly stated	2021	12	7	Detection of phishing websites by identifying and selecting critical URL-based features	A pipeline that applies recursive feature elimination (RFE) alongside ensemble machine learning techniques – including bagging (with a Decision Tree base), AdaBoost, XGBoost, Gradient Boosting, and Extra Tree Classifier – to automatically select	A dataset comprising 112 URL features (with a target attribute) that is pre-processed via standard scaling and split into 80% training and 20% testing sets
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Dong-Jie Liu; Guang-Gang Geng; Xiao-Bo Jin; Wei Wang	An efficient multistage phishing website detection model based on the CASE feature framework: Aiming at the real web environment	Computers & Security	Elsevier Ltd.	2021	110	N/A	Addresses the need for fast and accurate phishing website detection in real web environments by overcoming shortcomings of existing methods that do not fully analyze phishing features.	Proposes a comprehensive and interpretable CASE feature framework using quaternary features—Counterfeiting, Stealing, Affiliation, and Evaluation—and designs a multistage detection model that incorporates (1) whitelist filtering, (2) fast	A practical large-scale dataset is constructed by gathering websites from varied sources (different languages, content qualities, and obfuscation levels) to simulate the real web environment. Two types of experiments were performed:
uthors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Raniyah Wazirali, Rami Ahmad, Ashraf Abdel-Karim Abu-Ein	*Sustaining accurate detection of phishing URLs using SDN and feature selection approaches*	Computer Networks	Elsevier B.V.	2021	201	N/A (Article No. 108591)	Improving the accurate detection of phishing URLs while reducing the computational burden on end-user devices (addressing low accuracy, lengthy learning curves, and hardware limitations)	A novel integration that offloads URL classification to the SDN controller through the combined use of feature selection (Recursive Feature Elimination with SVM) and deep learning (Conventional Neural Network, forming the FS-CNN	A dataset comprising 51,200 URL samples: legitimate URLs collected from https://5000best.com/websites and phishing URLs obtained from https://www.phishtank.com
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Jibrilla Tanimu, Stavros Shiaeles, Mo Adda	A Comparative Analysis of Feature Eliminator Methods to Improve Machine Learning Phishing Detection	Journal of Data Science and Intelligent Systems	BONVIEW PUBLISHING PTE. LTD.	2024	2	2	Enhance ML-based phishing detection by selecting the most significant features – thereby reducing computational overhead and improving detection accuracy in real time.	Comparative evaluation of various feature elimination methods (e.g., Recursive Feature Elimination, Univariate Feature Selection, correlation-based selection) integrated with machine learning classifiers such as Random Forest, SVM, etc.	Data was collected via a crawler from the PhishTank repository, amassing over 50,000 phishing websites (with nonphishing data included) stored in a MySQL database.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Emre Kocyigit, Mehmet Korkmaz, Ozgur Koray Sahingoz, and Banu Diri	Enhanced Feature Selection Using Genetic Algorithm for Machine-Learning-Based Phishing URL Detection	Applied Sciences	MDPI (Basel, Switzerland)	2024	14	6081 (article number)	Addresses the challenges of phishing detection by mitigating overfitting, reducing computational cost, and improving model performance caused by an excessive number of features in URL-based phishing detection systems.	Proposes a Genetic Algorithm (GA)–based approach enhanced with a local optimization step. The method leverages uniform crossover, bit-flip mutation, and tournament selection to efficiently select an optimal subset of features for	The study uses URL-based features from phishing datasets. However, precise details on dataset size, source, or public availability are not explicitly provided.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Murathan OK, Ilker Kara, Ahmet Ozaday	Characteristics of Understanding URLs and Domain Names Features: The Detection of Phishing Websites With Machine Learning Methods	IEEE Access	IEEE	2022	10	Not specified	Detect phishing websites by analyzing URL and domain name features, thereby countering online phishing attacks.	Extraction of eleven predetermined features from URLs and domain names with six machine learning algorithms (including Logistic Regression, LDA, KNN, Decision Trees, SVM, and with Random Forest showing the highest performan	Dataset of 32,928 records (approximately 20,614 phishing and 12,314 legitimate websites) gathered from the TR-CERT open-source data. The paper also provides an access link for dataset requests.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Abdul Karim, Samir Brahim Belhaouari, Mobeenshahrooz, Khabib Mustofa, Ands Ramanakumarjoga	Phishing Detection System Through Hybrid Machine Learning Based on URL	IEEE Access	IEEE	2023 (publication date: 3 March 2023)	11	Not Specified (N/A)	Detecting phishing URLs/websites to enhance cybersecurity and safeguard user data against phishing attacks	A hybrid machine learning approach employing several algorithms including Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVC) combined as the proposed LSD (LR+SVC+DT) model with soft/hard voting. The study also	Utilizes a phishing URL dataset sourced from Kaggle consisting of 11,054 records and 33 attributes (data presented as vectors from over 11,000 websites)
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method Used to Solve the Problem**	**Data & Its Availability**

Rizka Widyarini Purwanto, Arindam Pal, Alan Blair, Sanjay Jha	PHISHSIM: Aiding Phishing Website Detection With a Feature-Free Tool	IEEE Transactions on Information Forensics and Security	IEEE	2022	17	N/A	Effectively detect phishing websites without needing manual feature extraction, even when phishing sites are slight variations of known attacks.	Uses a feature-free approach by computing the Normalized Compression Distance (NCD) over website HTMLs. The method employs the Furthest Point First algorithm for extracting prototypes from clusters and uses	Evaluated on a large dataset comprising phishing websites (e.g., reports from PhishTank and manually verified brand-specific cases). The paper does not explicitly state if the dataset is publicly available.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Ali Aljofey, Qingshan Jiang, Abdur Rasool, Hui Chen, Wenyin Liu, Qiang Qu, Yang Wang	An effective detection approach for phishing websites using URL and HTML features	Scientific Reports	Nature Publishing Group	2022	12	8842 (article number)	Detect phishing websites that masquerade as legitimate and mitigate risks such as 0-hour attacks and false positives	Extracts a hybrid set of features including: • URL character sequence features (without relying on phishing prior knowledge) • Textual content features (via character-level TF-IDF from plaintext and noisy HTML) • Various hyperlink features<	A custom dataset was built comprising 60,252 webpages (27,280 phishing and 32,972 benign), with testing also performed on a benchmark dataset.
Authors	**Title**	**Journal Name**	**Publishing House**	**Year**	**Volume**	**Issue**	**Problem Solved**	**Method used to solve the problem**	**Data & its availability**

Dr. A. Usha Ruby; Dr. George Chellin Chandran J	Enhancing Phishing URL Detection Accuracy in Software-Defined Networks (SDNs) through Feature Selection and Machine Learning Techniques	Research Square (preprint/ Research Article)	Not specified	2024	N/A	N/A	Increase the detection accuracy of phishing URLs in SDNs, addressing the challenges posed by evolving phishing attacks	A combined approach leveraging feature selection via Recursive Elimination (FSRE), k-means clustering, binary encoding for feature extraction, normalization, and deep learning using a Convolutional Neural Network integrated within the SDN	Real-world phishing URL datasets were used in an SDN testbed; however, detailed source information is not provided

Performance Metrics & Values	Future Works (if any)	Limitation (if any)	Critique (if any)
Neural Network: 90.23% accuracy; Naïve Bayes: 92.97% accuracy; Adaboost: 95.43% accuracy; additional metrics such as precision, sensitivity, specificity, and F1-score are also reported	Future research could focus on enhanced feature extraction techniques, incorporation of advanced ML/DL algorithms, and adaptation for real-time detection to tackle evolving phishing strategies	Reliance on a single Kaggle dataset, potential dependence on rule-based feature extraction, longer training times, and possible challenges in adapting to fast-changing phishing tactics	Although high accuracies are achieved, the study would benefit from broader dataset validation and deeper analysis of model robustness and real-world applicability

Performance metrics & values	Future works (if any)	Limitation (if any)	Critique (if any)

Evaluated using accuracy, precision, recall, F1-score, and specificity. The comparative analyses showed that the proposed method outperforms individual models – with one instance reporting an accuracy of up to 99.55% when using third-party	Not explicitly discussed in the provided content.	The use of third-party services increases detection time and may impact computational efficiency.	The paper does not offer an explicit critique; the proposed approach is presented as effective, although issues like processing speed and scalability might benefit from further exploration.
Performance Metrics & Values	**Future Works (if any)**	**Limitation (if any)**	**Critique (if any)**

Performance metrics & values	Future works (if any)	Limitation (if any)	Critique (if any)
The modified language models achieve near-perfect in-domain performance—with MCC scores close to 0.98–0.99—and improved robustness on out-of-domain data (MCC up to 0.84 with the priming approach). Perplexity values improve	Not explicitly detailed; the discussion hints at exploring strategies to further enhance robustness to domain shifts and generalization beyond the training distribution.	Fine-tuned models exhibit overfitting to the training domain, causing a reduction in performance when handling emails from unseen or different distributions.	Although the dual approach (fine-tuning and priming) shows promising performance, the heavy reliance on large pre-trained models and extensive data cleaning could limit replicability and practical deployment in real-world, noisy settings.

The paper reports metrics like a precision rate of up to 99.14%, an error rate reduction of 30% when using supplemental protocols (e.g., WHOIS integration), and employs standard performance measures (TP, TN, FP, FN, F1-score.	While not explicitly detailed, the discussion hints at future improvements such as further refinement of detection algorithms, optimizing classification thresholds, and adapting to evolving phishing techniques.	The system may not catch every phishing instance due to the continuously evolving nature of attackers; its performance is dependent on the quality and structure of the available data, with potential for false positives/ negatives.	Although the methodology is exhaustive and technical, the paper could be critiqued for not elaborating on issues of reproducibility (e.g., direct dataset accessibility) and scalability in real-world implementations, as well as for providing limited comment
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Performance metrics & values	Future works (if any)	Limitation (if any)	Critique (if any)
Best performance reached 99.61% accuracy (observed for the combination of BERT and LSTM)	Not explicitly mentioned	Not explicitly discussed	No explicit critique provided; additional discussion on dataset generalizability could be beneficial.

Reported evaluation metrics indicate an overall accuracy around 97% (with detailed metrics per model: e.g., CNN: Acc ≈ 97%, Precision ≈ 96–98%, Recall ≈ 96–97%, F1-Score ≈ 96–97%)	Future research is suggested to further address challenges such as the imbalance in data and to optimize computational efficiency for real-time phishing detection	The imbalanced nature of the dataset leads to challenges including a non-negligible proportion of false negatives and, in some architectures (e.g., CNN-RNN), increased training/testing time	The paper does not provide external validation with independent data and offers limited discussion on the practical impact of false negatives; in addition, the complexity of the proposed models may constrain scalability in real-world applicatio
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Baseline models achieved approximately 0.99 accuracy with F1 scores around 0.95–0.97; however, under adversarial attacks the F1 score dropped dramatically (in some cases as low as 0.10), while the defensive K-NN approach yielded 94%	Investigate more robust augmentation methods and scalable defense mechanisms—particularly to improve the classification of legitimate emails under adversarial conditions	Synthetic phishing email generation using GPT-2 did not consistently boost robustness; the defense methods were less effective for legitimate emails; some adversarial attack techniques (e.g., Textfooler) are computationally expensive	Although the study makes a valuable contribution by addressing adversarial vulnerabilities in detection models, it is limited by its focus on specific attack scenarios and models, and further work is needed to balance robustness with
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The CNN model achieved an accuracy of 98.4%. Additional metrics include high detection scores (with malicious URLs scoring around 0.999, surpassing the default threshold of 0.5) along with evaluations based on precision, recall, and	While the paper does not include an explicit “future works” section, it hints at the potential for enhancing the character encoding strategy and adapting the model to future phishing techniques.	Limitations are not explicitly detailed. However, potential drawbacks include reliance on a fixed threshold for classification (0.5) and the challenge of ensuring that the dataset remains representative as phishing techniques evolve.	The approach is robust and technically detailed, demonstrating high accuracy. Critically, one might note that further discussion on handling false positives in diverse real-world environments, scalability issues, and continuous model updates in an
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Performance metrics & values	Future works (if any)	Limitation (if any)	Critique (if any)
BERT-LSTM: Accuracy = 99.55%, Precision = 99.61%, Recall = 99.55%, F1-score = 99.24% (with comparative analysis against models such as Naïve Bayes, RNN, and SVM)	Extend the current work toward real-time detection, refine feature engineering, incorporate larger and more diverse datasets, and optimize computational efficiency	Relies on large, annotated datasets; high computational requirements; may need frequent updates to cope with evolving phishing tactics	Although the performance metrics are outstanding, there is a concern about overfitting and limited generalization when faced with real-world, noisy data. Additionally, the method’s high complexity might hinder deployment in resource-

The study reports competitive validation accuracy and fast testing performance (with a maximum test time of 0.07856 seconds), claiming to outperform state-of-the-art methods though detailed numeric accuracy values are not explicitly	Future work is mentioned in Section 6; it likely involves further enhancements in network architecture and refining feature selection methods, though details are not fully elaborated in the excerpt	Limitations are not explicitly stated; however, the discussion hints that model performance may be sensitive to the chosen feature selection method and that some techniques (e.g., metaheuristic optimization) can lead to longer model-	The study provides a detailed and well-structured approach along with a comparative analysis to related work. A critique might be that more explicit reporting of numerical performance metrics and real-world limitations could enhance clarity and
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Accuracy: 98.1%	Not explicitly mentioned	The article does not explicitly list limitations of the proposed model; however, as with many DNN approaches, issues like heavy data requirements and computational burden may be inherent.	No explicit critique is provided; the paper positions WebPhish as outperforming existing baseline methods.

Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The experimental results indicate improved detection accuracy and faster training speed compared to other methods, though no specific numerical values are provided in the excerpt.	The article does not explicitly state future work directions.	The paper notes that domain and topological features may not be fully captured for all phishing websites due to temporary unavailability. Moreover, some steps (e.g., reliance on synthetic data via ADASYN) might introduce dependency on	While the method shows promise in improving training speed and detection accuracy, the reliance on multiple preprocessing steps (ADASYN and SDAE) and advanced matrix operations may add complexity when scaling or deploying in real-world environme
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Accuracy: 99.51%; Recall: 99.68%; Precision: 99.5%; F1-score: 99.52%	The paper’s concluding section (Section 9) outlines future scopes to further refine feature engineering approaches and improve scalability and application robustness in real-world scenarios.	The dual-layer deep learning approach may introduce high computational complexity and resource demands, and while the work mitigates data imbalance, the underlying dataset size and balance may still pose challenges in broader deployme	Although the performance metrics are outstanding, the study could benefit from extended evaluation on more varied and larger datasets as well as further discussion on model interpretability and real-world scalability.
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

f1 score of 0.99 (as reported in the abstract).	Not explicitly mentioned within the provided content.	The article does not explicitly discuss limitations; however, aspects such as scalability or generalizability beyond the selected public datasets might require further evaluation.	The approach is comprehensive and robust—integrating data merging, preprocessing, explainable AI, and web deployment. Nonetheless, it could benefit from a deeper discussion on real-world deployment challenges, ablation
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Multiple accuracy reports are provided: XGBoost achieved 93.0% accuracy; with the optimal subset of 29 features the accuracy reached 94% – while Logistic Regression (89.8%), AdaBoost (90.0%) and Gradient Boost (92.8%) were also	Not explicitly mentioned	The study focuses solely on URL-based features and does not discuss potential limitations in capturing other phishing behaviors	Although the methodology is detailed and the comparative study is thorough, the paper could benefit from a deeper discussion on limitations (e.g. reliance on URL features alone) and future directions for incorporating broader data (such
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The method achieves better detection results with high efficiency, including high recall rates and very low false alarm rates, while significantly shortening execution time. However, the article does not provide detailed numerical values for	The article does not explicitly outline future works, though it implies that further research into advanced feature extraction and model generalization could be explored.	Not explicitly discussed; however, as with many phishing detection approaches, challenges may arise from rapidly evolving phishing techniques and the difficulties inherent in handling extremely imbalanced (real-world) datasets.	Although the framework is comprehensive and practical, the work would benefit from more detailed performance metrics and validation on larger, more diverse datasets. The reliance on hand-crafted feature extraction may also limit adaptabili
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Achieved 99.5% phishing detection accuracy; additional simulation-based performance evaluation using metrics like True Positive rate, among others	Future work is mentioned (in Section 5) but not detailed in the provided excerpt	Limitations are not explicitly detailed; possible concerns include evaluation limited to a simulated environment and potential scalability issues of using a centralized controller	The approach is innovative in offloading heavy detection tasks to an SDN controller while integrating advanced feature selection and deep learning. However, real-world validation and further discussion on computational overhead
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The RF classifier demonstrated superior performance compared to others. Exact numerical values are not fully delineated in the text (with literature comparisons showing accuracies in the range of ~97%–99%).	Section 7 outlines future directions, including further refinement of feature selection and exploration of advanced neural network algorithms to further improve phishing detection.	Some feature elimination methods (chi-squared test, stepwise regression, forward feature selection, etc.) were dropped due to poor performance and multicollinearity issues; results depend on dataset quality.	The study does not include a detailed quantitative breakdown of its own experimental performance and could be extended by broader evaluation across different phishing attack types.

Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The GA approach uses “Recall” as the fitness metric to evaluate model performance. Although improvements in recall are stressed, specific numerical values or benchmark comparisons are not detailed in the extract.	The conclusion mentions that further work can explore additional optimizations and possibly extend the approach to different domains, though no detailed roadmap is provided.	The method may incur high computational costs due to repeatedly evaluating the fitness function for each chromosome and managing the local optimization process.	While the approach is thorough in integrating GA with local optimization for feature selection, the paper could benefit by providing more detailed quantitative benchmarks, clearer data availability information, and discussion on
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

Performance metrics & values	Future works (if any)	Limitation (if any)	Critique (if any)
Reported accuracy up to 98.90%, with test results showing 98% detection on phishing pages and 97% on legitimate websites (overall ~98% correct prediction rate)	Integration into real-time email link blocking and secure connection mechanisms. Also, further dynamic updates to track new phishing tactics and continuously update the database are planned.	Reliance on high-quality, balanced data. The performance may deteriorate over time as attackers further evolve their techniques; the dataset is mainly national, which might limit diversity.	The study could be critiqued for potential issues in generalizability due to the use of a national dataset and the possibility that the feature extraction process may not fully account for sophisticated or evolving adversarial tactics.

Evaluated using accuracy, precision, recall, F1-score, and specificity. Comparative analysis indicates that the proposed approach outperforms conventional models with reported accuracy figures reaching as high as approximately 99.55% in	Not explicitly specified; however, the discussion hints at potential enhancements such as incorporating more advanced NLP-based feature extraction and extended real-world validations.	Not explicitly discussed; potential limitations include dependence on dataset quality and increased detection time when integrating third-party features.	While the study presents a comprehensive and innovative hybrid approach, it lacks detailed quantitative comparisons for each algorithm and an in-depth discussion on the generalizability and potential evolving nature of phishing attacks.
Performance Metrics & Values	**Future Works (if any)**	**Limitation (if any)**	**Critique (if any)**

AUC of 98.68%, TPR of about 90%, FPR of 0.58%, with an average processing time around 0.3 seconds.	Future work may extend the framework to cover phishing websites with completely novel HTML structures and incorporate additional web content features to enhance robustness.	The methodology is inherently limited to detecting variations of phishing websites that have previously occurred, potentially missing novel phishing templates with entirely different HTML structures.	While the feature-free, NCD-based approach is elegant and achieves strong performance, it might encounter scalability issues with very large datasets and could benefit from complementing HTML similarity with other features (e.g., dynamic
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

On the custom dataset: 96.76% accuracy with a 1.39% false positive rate; on the benchmark dataset: 98.48% accuracy with a 2.09% false positive rate	The paper’s conclusion mentions future work, though detailed plans were not provided in the extracted text.	• The approach relies on HTML/text-based features which are language-dependent • It requires access to the webpage’s HTML source code	While robust and outperforming several baseline methods, the method’s reliance on HTML content and plain-text extraction may limit its effectiveness if webpage structures change, if content is embedded as images, or in cases of
Performance metrics & values	**Future works (if any)**	**Limitation (if any)**	**Critique (if any)**

The paper reports improvements in detection accuracy and a reduction in false positives compared to conventional methods, but no explicit numerical values are given	Directions include further exploration of scalability, enhanced adaptation to emerging phishing tactics, and potentially mitigating computational/hardware resource demands	Potential high computational costs and hardware demands; lack of extensive quantitative performance data	The integration of multiple techniques is innovative yet the paper could benefit from more detailed performance metrics and a deeper discussion regarding scalability and energy consumption

The problem statement is: "Despite the availability of extensive data on academic publications, there is a lack of co

The problem is to improve the detection of phishing URL attacks by developing more robust machine learning mode

The problem statement for the thesis is: "Despite the extensive research documented in various journals, there is a

How can a hybrid machine learning model, utilizing a combination of decision tree, linear regression, random forest

The problem statement is: "Despite the extensive data available on academic publications, there is a lack of compre

How can pre-trained language models be effectively adapted to detect phishing emails, addressing the challenges c

The problem statement is: "Despite the availability of extensive data on published research articles, there is a lack c

How can a hybrid machine learning approach be effectively developed and implemented to enhance the detection

The problem statement for the thesis is: "Despite the extensive research documented in various journals, there is a

The problem addressed in this thesis is the need for an effective and accurate model to detect and prevent email ph

The problem statement is: "Despite the extensive data available on publication metrics such as authorship, journal

The problem addressed in this thesis is the challenge of accurately detecting and classifying phishing emails within

The problem statement is: "Despite the extensive data available on publication metrics such as authorship, journal

How can phishing email detection models be enhanced to maintain high accuracy and robustness against adversar

The problem statement for the thesis is: "Despite the extensive research documented in various journals, there is a

The problem addressed in this thesis is the need for an effective real-time phishing detection system that can accur

The problem statement is: "Despite the extensive research documented in various journals, there is a lack of compr

The problem is to develop a real-time, computationally efficient phishing email detection system that maintains hig

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] Empty balance. Go to Add-ins > GPT for Excel™ Word™ > Billing

[GPT ERROR] cancelled

mprehensive analysis on the correlation between the methods used to solve problems and the performance metrics

ls that can adapt to evolving phishing strategies, enhance feature extraction techniques, and validate findings acros

lack of comprehensive analysis on the effectiveness of different methodologies used to solve specific problems, as

, naive Bayes, gradient boosting classifier, K-neighbors classifier, support vector classifier, and a novel hybrid LSD m

ehensive analysis on the correlation between the methods used to solve problems and the performance metrics ach of domain shift and generalization, while minimizing overfitting and ensuring practical deployment in real-world setti

of comprehensive analysis on the correlation between the methods used to solve problems and the performance me of web phishing attacks, considering the challenges of data quality, evolving phishing techniques, and the need for re

lack of comprehensive analysis on the effectiveness of different methodologies used to solve specific problems, as

ıishing attacks, leveraging deep learning techniques such as CNNs, LSTMs, RNNs, and BERT for feature extraction, w

impact, and methodological approaches, there is a lack of comprehensive analysis on how these factors collectivel an imbalanced dataset using deep learning algorithms, with a focus on optimizing model performance and comput

impact, and methodological approaches, there is a lack of comprehensive analysis on how these factors collectivel

lack of comprehensive analysis on the effectiveness of different methodologies used to solve specific problems, as

rately classify malicious URLs using deep learning methods, specifically through a browser extension, while overcom

rehensive analysis on the effectiveness of different methodologies used to solve specific problems, as well as the av h accuracy and generalization across diverse and evolving datasets, while addressing overfitting and deployment ch

well as the availability and impact of data on performance metrics. This thesis aims to evaluate the correlation betw

nodel, be optimized to improve the accuracy and efficiency of phishing URL detection while addressing challenges re

nieved, particularly in relation to the limitations and critiques identified in the literature. This thesis aims to address t etrics achieved, which hinders the identification of optimal research practices and the advancement of future works

well as the availability and impact of data on performance metrics. This thesis aims to evaluate the correlation betw

y influence the effectiveness and efficiency of problem-solving in academic research, as measured by performance y influence the effectiveness and limitations of problem-solving methods in academic research, as evidenced by pe

well as the availability and impact of data on performance metrics. This thesis aims to evaluate the correlation betw

ning challenges such as evolving phishing techniques, false positives, and scalability in diverse real-world environm

ailability and impact of data on performance metrics. This thesis aims to address these gaps by systematically evalu

een the methods employed and their performance outcomes, while also addressing the limitations and critiques ide his gap by systematically examining the relationship between problem-solving methods and their effectiveness, as r

veen the methods employed and their performance outcomes, while also addressing the limitations and critiques ide

een the methods employed and their performance outcomes, while also addressing the limitations and critiques ide uating the methods, data availability, and performance outcomes reported in the literature, while also identifying lim

itations and future research directions to enhance the understanding and application of these solutions."