

Deep Learning-Based Phishing Website Detection System

Using URL and Content Analysis

Student Name : MD Shibli Hossain

Student ID Number : 2517990

Programme : MSc Computing

Module : Research Methods and Professional Development

Module Code : ACFA7004

Lecturer : Dr. Suresh Shanmugasundaram

Internal Verifier : Dr. Nitheesh Kaliyamurthy

Assignment Title : Research Proposal

Word Count : 3,883 words

Submission Deadline : 14, August 2026

Table of Contents

1. Introduction.....	2
1.1 Context and Background	2
1.2 Research Problem	3
1.3 Research Aim and Objectives.....	4
1.4 Importance of Research	5
2. Literature Review	6
3. Research Design and Methodology	8
3.1 Research Design	8
3.2 Research Methods.....	9
3.3 Materials	10
4. Expected Outcome.....	11
5. Timeline.....	12
6. Conclusion	13
References.....	14

1. Introduction

1.1 Context and Background

The rapid pace of change in the sphere of computing research has accelerated the rate of change in the field of cyber-security in response to the digital attacks individuals, organisations and critical infrastructure are subjected to. Phishing is among the most prevalent forms of cyber-attacks and is yet one of the most damaging, not only monetarily, but in the human suffering it causes, as it is based on people's trust rather than on a technical vulnerability. A phishing site is a duplicate of a legitimate site designed to trick users into providing information such as personal identity and bank account numbers, login names and passwords.

The size of this threat is on an increase. In 2025, the Anti-Phishing Working Group identified 3.8 million phishing attacks globally, which is a modest increase from the previous year, but a continuation of a 5-year trend of increasing volumes and sophistication of attacks each year. Attackers are more interested in high-value credential targets making the most exploited industries those in financial services, as well as SaaS/ webmail services. Still later on, with the introduction of generative artificial intelligence, the threat has been increased even more, with attackers now able to create grammatically correct, personalized, and multilingual phishing messages, which do not trigger the traditional detection heuristics of spelling mistakes or generic templates.

Using detection methods has changed significantly in the last 20 years. The first systems were blacklisting based: They produced databases of malicious URLs and would check incoming requests against them. Though blacklists are less expensive to compute, they are reactive in nature and can't catch any new or unknown phishing domains that are registered, which are also known as zero-day phishing websites. It was done by heuristic and rule-based methods based on the presence of handcrafted indicators such as URL length, presence of IP addresses or HTTPS. These increased coverage still, were not effective to attackers who could construct URLs in a calculated manner to evade existing rules.

Machine learning was a tremendous success, as models could now adopt a discriminative approach on many more labelled examples, instead of relying on hand-crafted rules. Even recently, deep learning models like convolutional networks, recurrent networks and transformer-based language models have demonstrated the ability to automatically learn complex lexical, structural and

semantic features of both URLs and the content of web pages, with consistently higher detection accuracy over the traditional machine learning baselines. This dynamic evolution of blacklisting to a dynamic, learning-based one models the contemporary context of the current study.

1.2 Research Problem

Despite much progress, the current “deep learning” style phishing detection approaches only consider one data modality, either the URL string or webpage content, but do not explicitly “deep learn” the correlation between the two data modalities. It is these real restrictions of this limited scope which provide impetus to this present study.

First, URL-only models can be adversarial obfuscated. Attackers can use a similar technique to (roughly) create new domains that seem similar to a legitimate brand name, using character substitution or manipulation with subdomains, and give the delivery lexical classifiers a false sense of harmlessness, while the contents of the page being delivered are obviously fake. Attackers can bypass these models since they do not analyze the page rendered and, given knowledge of the classifier weights of particular features of URLs, can be evaded systematically.

Second, content only models, have a complementary weakness. They tend to take the entire page before they can fetch and render for classification, and introduce latency and in some cases where they are implemented, expose the user to the partial page load before a verdict is reached. Content Based Classifiers are vulnerable to attacks due to the following reason: Attackers can synthetic create attack HTML content similar enough to keep semantically similar which can be hardly detected by content based classifiers but require URL level evidence.

Most importantly, the third: literature often gets high performance on one benchmark data set, and gets a huge drop on another, new, unrelated, independently acquired data set. This gap in generalisation infers that the majority of published models are overgeneralised with artefacts specific to the data they are trained over and very little about the real nature of the phishing behaviour, greatly limiting their practical applicability.

Finally, the best raw-result deep learning models, in particular transformer based models, will likely be a black-box. The binary verdict often issued to security analysts and end users typically doesn't include any text to indicate which features led to the verdict, thus reducing trust, and making it more difficult to audit and test for false positives in operational systems.

The lack of interpretability, weak cross-dataset generalisation and adversarial vulnerability, along with modality isolation, are what constitute what is evident and unsolved gap in the literature of phishing detection on which this proposal is based.

1.3 Research Aim and Objectives

Goal: To design, test and evaluate a hybrid deep learning model synergistically analysing URL features and webpage features for phishing websites detection and explainable AI methods to generate explainable and transparent classification decisions.

In order to attain this objective, the study will be organized in the following objectives:

Objective 1: In the first four weeks of the project, to search the existing literature and select a systematic review published after 2019 on the use of deep learning to detect phishing to understand the current architectures, the used feature engineering techniques and the reported limitations in the use of the URL-only detection, URL content-only detection and hybrid detection forms.

Objective 2: Gather, merge and preprocess a labelled data set of phishing and legitimate web sites sampled by at least two independent public sources containing URL strings and the appropriate HTML documents, class balance and remove duplicates and corrupted records, all by week six.

Objective 3: By the end of week ten, develop and train a hybrid neural network design that includes a URL encoder (convolutional + bidirectional long short-term memory) and content encoder (customized transformer language model) and feature fusion layer, to achieve the final classification using both of the representations.

Objective 4: Within 14 days train and evaluate the proposed model based on accuracy, precision, recall, F1-score and AUC-ROC as evaluation metrics, and compare the proposed model with at least three baseline models (a deep learning model of URL only, content only deep learning model, and a traditional ML model) and analyze the generalisation using at least two independent datasets.

Objective 5: The trained model will feature explainability methods, where SHAP and LIME will give a feature-level explanation why the prediction was made, and the explainability will be evaluated for clarity and consistency, achieved at week sixteen.

All these are measurable with a defined product: summary of the review, dataset cleaned, trained architecture, evaluation report, explainability output; within the scope of the project resources; directly related to the above objective; and have clear milestones, meeting the SMART criteria.

1.4 Importance of Research.

This study may have significance in both academic and real-life activities in cybersecurity. In theory, it will assist in building the longstanding body of research on multimodal deep learning by showing how URL-level and content-level representations can be pooled together in a single architecture but also adds to the paucity of empirical research on cross-dataset generalisation, which still faces underrepresentation compared to single-dataset reports on accuracy.

In real practice, the investigation is a solution to an issue having a direct practical implication. It is better for end users because it offers more defensible protection against credential theft and financial fraud, even against the latest, most sophisticated phishing attacks that are evading traditional phishing filters and are driven by artificial intelligence. The organisations including financial institutions or SaaS providers are the most frequently attacked by the currently reported APWG and can gain from a detection layer that generalises or frames of attack patterns, instead of the ever-changing rules.

It is also very practical to have an explainability component in the process. Security operations teams require more than just detection, they need to make reasonable decisions, particularly when it comes to automated blocking, which has an impact on regular business communications. A faster analyst triage and lower operational costs in the event of a false positive can be an indicator of a process when a site was flagged, such as because of a suspicious domain structure, or credentialharvesting forms.

The possible implications of this study include browser extensions, integrating into email gateways, and enterprise API-driven scanning platforms, which provide an avenue of an academic prototype to practical security technology.

2. Literature Review

This review of literature seeks to summarise the current body of knowledge on deep learning (DL) models to detect phishing websites in the past seven years (2019-2026) that has seen the transition of traditional machine learning-based models to transformer-based and multi-modal models. The review is organized around four strands: URL-based deep learning models, content-, or HTMLbased, hybrid multimodal approaches and the new paradigm of explainable detection. Throughout, individual studies are placed in the context of the larger phishing/surveys literature [3, 7, 14].

URL-Based Deep Learning Models

The original deep learning phishing models were based on the URL itself, leveraging the fact that most malicious URLs are composed of distinct lexical and structural patterns, which can be found prior to a phishing page being displayed. Jalil et al. showed that the high detection accuracy can be achieved by carefully designed lexical features and machine learning classifiers on benchmark datasets [5] and Samad et al. further proved that this accuracy can be enhanced by fine tuning the hyper parameter of the classifiers [6]. Ozcan et al. proposed a hybrid DNN-LSTM design and demonstrated that a hybrid network performed better than either a deep network or an LSTM network, in capturing the static and sequential features of the URL, respectively [17]. Likewise, Javeed et al. demonstrated that multi-learning algorithm led to better accuracy of URL classification [16], and Ujah-Ogbuagu et al. demonstrated a real-time able hybrid deep learning network that was trained to classify URLs of spoofed websites [8]. Recent work has benchmarked these architectures against more modern benchmark datasets such as PhiUSIIL [24, 25] and has demonstrated that still deep sequential models outperform traditional classifiers on more recent datasets. Convolutional and recurrent models do better than their counterparts in all cases in this strand, although the vast majority of existing studies assume the URL string serves as the feature space.

Content and HTML-Based Approaches

Another set of work has looked at the real contents of the page and proposed that phishing pages are an indication of their bad intent through structural or semantic information that is not reflected in the URL itself, with Çolhak et al. suggesting a multi-model methodology to examine the HTML content directly, and achieved better scores in concluding that the page was being disguised by a

harmless URL but was actually malicious in nature [10]. Majgave and Gavankar applied a deep belief network on a transformer-based approach on webpage content which produced deeper semantic features than bag-of-words/ TF-IDF approaches previously was used [11]. The studies tend to do well on pages that have been cloned or made to look like a legitimate page, which is exactly where URL-only phishing detection models struggle, but they generally take a long time to make a decision as they need the entire page to be retrieved and processed before they can make a decision.

Hybrid Multimodal Architectures

It has also been found in a small body of research which has begun to incorporate the URL and content signals into the same model. This proposal proposes the fusion based design which has been directly referred to or more precisely replicated on the work of Opara et al. that demonstrated that fusion of raw URL and HTML characteristics give more detection power [13]. As an example, with simplicity of structure in HTML DOM graph and URL features, graph convolutional and transformer networks combined with these concepts yielded good performance, but with extremely complicated architecture [12]. Besides the above, Mousavi and Bahaghighat also indicated that more reasonable feature selection could lead to better deep learning compared to high numbers of features across modalities [9]. Besides the phishing domain, other sequential-pattern securityrelated applications, including IoT threat detection [18] and IoHT network intrusion detection [19], have been found to be well-performed by the hybrid CNN-BiLSTM and CNN-BiLSTM-DNN, whereasGRU-BiLSTM has been shown to be well-performed with other intrusion detection domains [20]. It is important to note that these results across domains are in favor of the architectural of the design of the hybrid encoder which is proposed in the present paper, but the studies per se are not phishing specific.

Explainability Trend

Lastly, and, perhaps, more conspicuously, there is a call of plausible phishing detection. Fatima et al. applied SHAP and LIME to the field of phishing classification in the IoT and Robotic Communications to show that feature-based attributions are possible with leading to minimal impact on phishing detection accuracy. Chavakala et al. also achieved the same results in the case of financial fraud and credit scoring classification models and demonstrated that the methods can be applied to other classification issues related to security [22]. Lim et al. introduced EXPLICATE

in the case of phishing detection, combining SHAP, LIME, and a large language model to translate technical explanations into natural language as a result, significantly, promoting interpretability and performance that is notably similar to black-box deep learning models [23]. Combined, this strand implies that explainability does not need to give up a significant amount of accuracy but more specifically explainability in combination with high capacity multimodal architectures is a relatively under-explored area.

Gaps and Positioning of the Present Study

The literature reviewed has three gaps that still exist. To begin with, despite the fact that, multimodal fusion models have been demonstrated to be effective [12, 13], the majority of the models that have been published have been tested on one data set which implies that cross-dataset generalisation, one of the most important tests in determining the practical efficacy of a model, has not been effectively checked. Second, Content-based and graph-based multimodal structures [11] and [12] are much more accuracy-oriented than deployment-latency (speed) ones, and are not particularly engaged in making things both real-time-enabled and real-time viable. Third, the techniques of explainability have been experimented on in isolation [21] or applied to relatively simple architectures [23] but not to a real model of hybrid URL-content fusion such as that presented here.

This study fills in three gaps. The proposed study further builds on the multimodal shift of other studies, such as Opara et al. and Yoon et al. [13, 12] as it integrates a URL encoder with a content encoder using a single fusion architecture, evaluates explicitly across at least two independent datasets, and embeds SHAP and LIME directly into the trained hybrid model instead of a simplified proxy, addressing the generalisation and interpretability issues that are common across all reviewed works.

3. Research Design and Methodology.

3.1 Research Design

The present work focuses on a Design Science Research (DSR) method, which fits well considering that the first deliverable is a new technical artefact, a hybrid neural network, as opposed to a theoretical one. DSR is quite appropriate to the style of research in the field of

computing that seeks to construct and test rigorously a resolution to a well-articulated practical difficulty, here the modality-isolation and interpretability gulf found in the literature review [13], [12].

The study adheres to the conventional DSR cycle: problem identification (Chapter 1), literaturebased design (Chapter 2), artefact construction (the hybrid URL-content fusion model), demonstration (training and inference on labelled datasets), and evaluation (comparing the model to baseline models and measurements in a number of metrics and datasets). Such a cyclic structure enables the design to progressively evolve along with empirical evaluation findings as opposed to being fixed on a specific architecture early in the design.

DSR is better than a completely experimental or a survey-only design, given that the research question is inherently constructive: how, not whether, hybrid multimodal fusion enhances phishing detection, how this architecture shall be structured, trained and interpretable. A DSR framing is so that artefact and its evaluation remain closely intertwined, with design choices, like which fusion mechanism to use or what layer to explain it with, being directly accountable to the problem statement and goals written in Section 1.3, instead of being discussed as a forlornly distinct matter.

3.2 Research Methods

The research is quantitative and experimental in nature, which has a controlled comparison between the proposed hybrid model and the existing baseline approaches. The strategy is suitable as the research questions are essentially measurable: all the parameters of detection accuracy, generalisation across datasets, and quality of explanations can be quantified and statistically compared, instead of the interpretative analysis inherent to qualitative research methods.

The proposed system architecture is made up of three components running consecutively. A URL encoder is a convolutional and bidirectional LSTM-based encoder that uses the raw URL string to encode the character-level and multi-character lexical patterns. Meanwhile, a content encoder, also grounded on a fine-tuned transformer language model, encodes the extracted HTML and visible text, to extract the semantic coheres within context. They are then fused in a feature fusion layer (implemented with concatenation and an attention mechanism) and then submitted to a final classification layer producing a phishing-versus-legitimate probability score. This end-to-end pipeline with the explainability module that will be added to the final output is depicted in the following diagram:

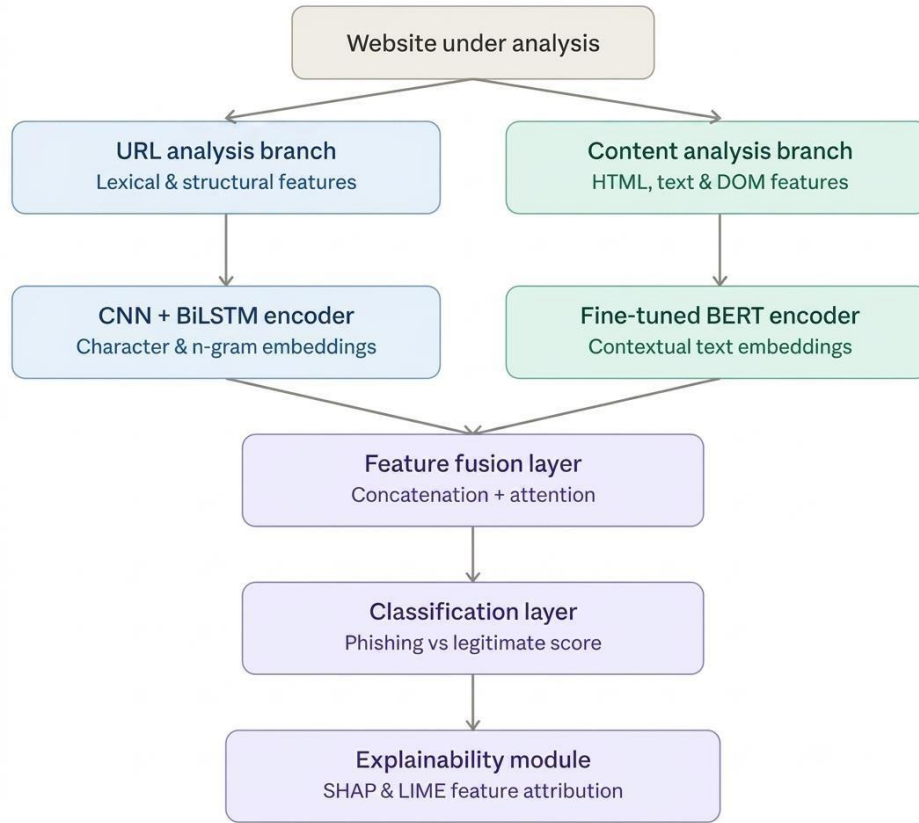


Figure 1: Architecture Diagram

This hybrid model is compared against three baselines: a URL-only deep learning model, contentonly deep learning model, and a traditional machine learning classifier, as encouraged in comparative designs of previous hybrid studies [17], [9]. All models are trained on an 80/10/10 train-validation-test split, and assessment is repeated on at least two independent datasets to directly test cross-dataset generalisation, the gap found in Section 2. Accuracy, precision, recall, F1-score, and AUC-ROC serve to measure model performance to give a balanced consideration of the overall accuracy and performance in class imbalanced domains.

3.3 Materials

Various public repositories will be used to supply data necessary to support the cross-dataset evaluation design: PhishTank and OpenPhish to get continuously updated, community-approved phishing URLs; the UCI Phishing Websites to get structure-based feature-labelled legitimate and phishing samples; and phishing datasets that are supported by Mendeley, like PhiUSIIL [24] and

PhiKitA [26] to make sure that the model is tested against up-to-date phishing methods as opposed to old samples.

The software environment will consist of Python, as the models will be implemented and trained with either TensorFlow or PyTorch, baseline classifiers and evaluation metrics will use Scikitlearn, and explainability module will use SHAP along with LIME. The training and development will follow the use of a GPU-accelerated cloud computing environment like Google Colab Pro or a similar resource procured through the university based on the computational power requirements of fine-tuning a transformer-based content encoder.

The datasets used are all publicly available, anonymised sets of URLs and page text with no personally identifying information added, and thus no primary data collection using human participants is necessary. However, the study will record the provenance of data sources and the license of every author of the data, and will not hold or distribute any raw scraped data other than that allowed by the terms of use of the underlying repository.

4. Expected Outcome

It is expected that the proposed hybrid model will do better than single-modality baselines, and it is likely to detect cases with better accuracy than URL-only or content-only architecture, especially when it comes to disguised URL paired with cloned page content, which is the Achilles heel of the proposed hybrid model in the first place. The cross-dataset assessment should result in a smaller amount of performance degradation than the amount reported in the single-dataset studies, reflecting better generalisation.

This explainability module should help generate stable feature-level explanations of individual predictions, enabling a security analyst to understand which features of URLs or content content elicited a particular classification, instead of receiving an unexplainable binary verdict. This must show that it is possible to obtain interpretability without inevitably compromising detection accuracy.

A number of restrictions are expected. Phishing datasets caused to be publicly available might be selectively biased toward certain brands, geographies or attack types that are best reported to platforms like PhishTank, thereby making it difficult to generalise to phishing patterns that are

poorly represented. Also, phishing can dynamically develop, so the trained model can either need retraining periodically to stay operational against new obfuscated or other phishing attacks generated by AI that are not in the training samples.

5. Timeline

The project will last eight weeks and will be organized in six stages which will be sequential and overlapping to some extent. Overlapping of two consecutive steps such as literature review and the last stage of proposal finalisation are an indication of the actual situation on the ground where dataset scoping and preliminary planning of the architecture can be done before the literature review process is over. The hardest part, the model development, has the most significant amount of time dedicated to it due to the complexity of the implementation and integration of three different components related to the fusion layer to be described in Section 3.2 and URL and content encoders.

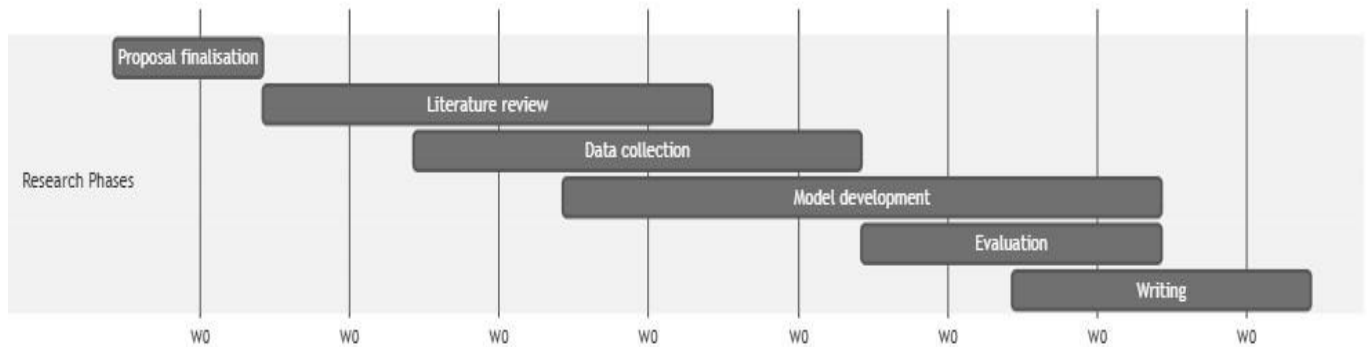


Figure 2: Gantt chart

This schedule is deliberately on the slower side during the model development phase, given the risk that there is more work needed in debugging and hyperparameter optimization of a transformer-based content encoder combined with a sequential URL encoder. In case of delays, the assessment stage can be condensed by focusing on the two most informative comparisons of the baseline instead of the entire three baselines as noted in Section 3.2, without jeopardising the essence of the research.

Table 1: Project timeline and milestones

Week	Milestone	Key Deliverable
1	Proposal finalisation	Confirmed research aim, objectives, and ethics form submitted
2–3	Literature review	Thematic review summary covering URL-based, content-based, hybrid, and explainability literature
3–4	Data collection and pre-processing	Cleaned, labelled dataset combining PhishTank, OpenPhish, UCI, and Mendeley sources
4–6	Model development	Implemented URL encoder, content encoder, and fusion-based classification layer
6–7	Evaluation	Benchmark results against baseline models across accuracy, precision, recall, F1-score, and AUC-ROC, tested on at least two datasets
7–8	Writing	Final analysis, explainability integration write-up, and complete project report

6. Conclusion

The current proposal has stated a clear research problem: the current systems of phishing detection by using deep learning are susceptible to adversarial URL obfuscation, poor generalisation across datasets, and not interpretable as they rely on a single data modality. To address this, this paper is a hybrid deep learning model, consisting of URL-level and content-level representations, a special encoder pair and attention-based fusion layer, trained on a diverse range of independent datasets and presented with SHAP and explainability by LIME. It is proposed that the given work is based on the Design Science Research approach and is backed by the eight-week plan, which will make it both methodologically justified and practically feasible within the means of the project. In the event of success, the resultant model would show measurable improvements in detection capability and disclosure, and guide future applications in browser-based defense, enterprise security

machinery and additional research would bring about expanded multimodal, elucidate phishing acknowledgment.

References

- [1] E. Putra, Ubaidi Ubaidi, Achmad Zulfikri, Goffal Arifin, and Revi Mario Ilhamsyah, “Analysis of Phishing Attack Trends, Impacts and Prevention Methods: Literature Study,” *Brilliance Research of Artificial Intelligence*, vol. 4, no. 1, pp. 413–421, Aug. 2024, doi: 10.47709/brilliance.v4i1.4357.
- [2] E. Morrow, “Scamming Higher Ed: An Analysis of Phishing Content and Trends,” *Computers in human behavior*, vol. 158, no. 158, pp. 108274–108274, Apr. 2024, doi: 10.1016/j.chb.2024.108274.
- [3] S. Priya, D. Gutema, and S. Singh, “A Comprehensive Survey of Recent Phishing Attacks Detection Techniques,” *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*, pp. 1–6, Mar. 2024, doi: 10.1109/icitit61487.2024.10580446.
- [4] U. Imtiaz, “INVESTIGATING THE IMPACT OF PHISHING ATTACKS ON ORGANIZATIONAL CYBERSECURITY POSTURE,” *Spectrum of Engineering Sciences*, pp. 1758–1780, 2025, Accessed: July 08, 2026. [Online]. Available: <https://www.thesesjournal.com/index.php/1/article/view/2590>
- [5] S. Jalil, M. Usman, and A. Fong, “Highly accurate phishing URL detection based on machine learning,” *Journal of Ambient Intelligence and Humanized Computing*, Oct. 2022, doi: 10.1007/s12652-022-04426-3.
- [6] A. Samad *et al.*, “Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection,” vol. 12, no. 7, pp. 1642–1642, Mar. 2023, doi: 10.3390/electronics12071642.
- [7] A. Safi and S. Singh, “A systematic literature review on phishing website detection techniques,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, Jan. 2023, doi: 10.1016/j.jksuci.2023.01.004.
- [8] B. C. Ujah-Ogbuagu, Oluwatobi Noah Akande, and Emeka Ogbuju, “A hybrid deep learning technique for spoofing website URL detection in real-time applications,” *Journal*

- of Electrical Systems and Information Technology*, vol. 11, no. 1, Jan. 2024, doi: 10.1186/s43067-023-00128-8.
- [9] S. Mousavi and M. Bahaghighat, "Phishing Website Detection: An In-Depth Investigation of Feature Selection and Deep Learning," *Expert Systems*, vol. 42, no. 3, Jan. 2025, doi: 10.1111/exsy.13824.
 - [10] Furkan Çolhak, Mert İlhan Ecevit, Bilal Emir Uçar, Reiner Creutzburg, and Hasan Dağ, "Phishing Website Detection Through Multi-model Analysis of HTML Content," *Algorithms for intelligent systems*, pp. 171–184, Jan. 2025, doi: 10.1007/978-981-97-69575_15.
 - [11] A. B. Majgave and N. L. Gavankar, "Automatic phishing website detection and prevention model using transformer deep belief network," *Computers & Security*, vol. 147, p. 104071, Dec. 2024, doi: 10.1016/j.cose.2024.104071.
 - [12] J.-H. Yoon, S.-J. Buu, and H.-J. Kim, "Phishing Webpage Detection via Multi-Modal Integration of HTML DOM Graphs and URL Features Based on Graph Convolutional and Transformer Networks," *Electronics*, vol. 13, no. 16, p. 3344, Aug. 2024, doi: 10.3390/electronics13163344.
 - [13] C. Opara, Y. Chen, and B. Wei, "Look before you leap: Detecting phishing web pages by exploiting raw URL and HTML characteristics," *Expert Systems with Applications*, vol. 236, p. 121183, Feb. 2024, doi: 10.1016/j.eswa.2023.121183.
 - [14] S. Asiri, Y. Xiao, S. Alzahrani, S. Li, and T. Li, "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks," *IEEE Access*, pp. 1–1, 2023, doi: 10.1109/access.2023.3237798.
 - [15] "Phishing Detection System Through Hybrid Machine Learning Based on URL," *ieeexplore.ieee.org*. <https://ieeexplore.ieee.org/abstract/document/10058201> (accessed July 08, 2026).
 - [16] Muhammad Usman Javeed, S. M. Aslam, Hafiza Ayesha Sadiqa, A. Raza, M. M. Iqbal, and M. Akram, "Phishing Website URL Detection Using a Hybrid Machine Learning Approach," *Journal of Computing & Biomedical Informatics*, vol. 9, no. 01, 2023, Accessed: July 08, 2026. [Online]. Available: <https://jcbi.org/index.php/Main/article/view/989>

- [17] A. Ozcan, C. Catal, E. Donmez, and B. Senturk, "A hybrid DNN–LSTM model for detecting phishing URLs," *Neural Computing and Applications*, Aug. 2021, doi: 10.1007/s00521-021-06401-z.
- [18] B. A. Agbor, B. U.-A. Stephen, P. Asuquo, U. O. Luke, and V. Anaga, "Hybrid CNN–BiLSTM–DNN Approach for Detecting Cybersecurity Threats in IoT Networks," *Computers*, vol. 14, no. 2, p. 58, Feb. 2025, doi: 10.3390/computers14020058.
- [19] "Design and Development of A Novel CNN-BiLSTM Hybrid Deep Learning Model for Cyber Attacks Detection in an IoHT Environment," *Ieee.org*, 2026, doi: 10.1109/IATMSI64286.2025.10985456.
- [20] A. A. Ghani and S. A. Alasadi, "A Deep Learning Algorithm to Cybersecurity: Enhancing Intrusion Detection with a Hybrid GRU and BiLSTM Model," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23605–23612, June 2025, doi: 10.48084/etasr.10666.
- [21] Z. Fatima, M. H. Tanveer, R. C. Voicu, S. Chakravarty, and H. Desa, "Explainable AI for IOT Devices and Robotic Communication Phishing Detection: A Machine Learning Approach Using LIME and SHAP," *Engineering, Technology and Applied Science Research*, vol. 15, no. 5, pp. 26478–26486, Oct. 2025, doi: 10.48084/etasr.11595.
- [22] S. J. Chavakula, C. A. J. Albert, E. Ebenezer, M. H. Bhagat, and C. V. Mahamuni, "Explainable AI (XAI) Using SHAP and LIME for Financial Fraud Detection and Credit Scoring," *2025 International Conference on Advanced Computing Technologies (ICoACT)*, pp. 1–8, Mar. 2025, doi: 10.1109/icoact63339.2025.11005238.
- [23] B. Lim, R. Huerta, A. Sotelo, A. Quintela, and P. Kumar, "EXPLICATE: Enhancing Phishing Detection through Explainable AI and LLM-Powered Interpretability," *arXiv.org*, 2025. <https://arxiv.org/abs/2503.20796> (accessed July 08, 2026).
- [24] A. Rawla, S. Singh, M. Daniyal, and P. Dubey, "Detection of Phishing Attacks in PhiUSIL Dataset using Deep Learning," *Procedia Computer Science*, vol. 259, pp. 543–552, 2025, doi: 10.1016/j.procs.2025.04.003.
- [25] T. Choudhary, Siddhesh Mhapankar, Rohit Bhddha, Ashish Kharuk, and R. Patil, "Machine Learning Approach for Phishing Attack Detection," May 2023, doi: 10.37965/jait.2023.0197.

- [26] “PhiKitA: Phishing Kit Attacks Dataset for Phishing Websites Identification,” *Ieee.org*, 2026, doi: 10.1109/ACCESS.2023.3268027.