Innovative Journal of Computing, Security and Digital Technologies
Volume 1, Issue 1, Jan – Apr 2026
Page no : 23-28

Detecting Phishing Websites Using Transformer-Based Language Models and URL Features

Al-Ahmadi. S., AbuSharkh. M

Department of Computer Science and Engineering, Delhi Technological University, Delhi–110042, India.

Abstract

Phishing websites use deceptive links to direct users to fraudulent pages and obtain sensitive information. This paper proposes a hybrid phishing detection framework that combines transformer-based language models with engineered URL features. A transformer encoder processes URL tokens to learn contextual and sequential patterns, while a feature extraction module calculates lexical and structural properties such as URL length, symbol frequency, domain composition, and suspicious token patterns. The two representations are fused and passed to a classification layer that predicts whether a URL is legitimate or phishing-related. The proposed workflow includes data preparation, model training, evaluation, and a decision stage for classifying incoming URLs. The framework is designed to support early detection without requiring full webpage rendering. Its effectiveness must be established through experiments using separate training, validation, and test data, comparison with baseline models, and evaluation using precision, recall, F1-score, and false-positive rate.

Keywords: Phishing detection, transformer, URL analysis, language models, feature fusion, cybersecurity, classification.

Publication Date : 25.01.2026

1. INTRODUCTION

Phishing remains a cybersecurity threat because attackers can create URLs that imitate trusted services, conceal their destinations, or use misleading words and character patterns. URL-based detection can help identify suspicious links before a user visits the associated website. However, simple rules and manually selected features may not capture the contextual relationships among URL components.

Transformer-based language models provide a way to represent URL strings as sequences of tokens and learn relationships between their components. Previous research has examined transformer models for phishing URL detection, including URLTran and BERT-based approaches. Other work has explored combining transformer representations with engineered URL features to capture both contextual and structural signals.

This study proposes a hybrid framework that uses both forms of information. The transformer branch learns contextual representations from URL tokens, while the feature-engineering branch calculates explicit URL characteristics. The outputs are combined to produce a phishing or legitimate classification.

How to Cite
Detecting Phishing Websites Using Transformer-Based Language Models and URL Features. (2026). Innovative Journal of Computing, Security and Digital Technologies, 1(1), 23-28.
https://ijcdst.in/index.php/about/article/view/9

2. PROBLEM STATEMENT

A URL may contain useful indicators of phishing, including unusual length, misleading subdomains, excessive separators, suspicious words, or irregular path and parameter structures. These indicators are not always sufficient individually. A legitimate URL may contain suspicious-looking terms, while a phishing URL may be designed to appear ordinary.

The research problem is therefore to determine whether combining contextual representations learned by a transformer with engineered URL features can provide more reliable phishing detection than relying on either representation alone.

3. OBJECTIVES

1. Develop a URL preprocessing pipeline for consistent tokenization and feature extraction.
2. Use a transformer-based language model to learn contextual URL representations.
3. Extract lexical and structural URL features.
4. Fuse the transformer representation and engineered features for binary classification.
5. Evaluate the proposed method against suitable baseline models using a held-out test set.

4. PROPOSED METHODOLOGY

4.1 Dataset Preparation

The dataset should contain URLs labelled as phishing or legitimate. Duplicate entries should be removed, labels checked, and the data divided into training, validation, and test sets. The split should be performed carefully to reduce leakage, such as near-identical URLs appearing in both training and test data.

4.2 URL Preprocessing

Each URL is normalized using consistent rules while preserving meaningful components. The system separates the scheme, hostname, path, and query parameters where available. The transformer tokenizer converts the URL into tokens, while the feature extraction module calculates numerical attributes.

4.3 Transformer-Based Representation

The tokenized URL is passed to a pretrained transformer language model. The model produces a contextual representation of the URL sequence. This representation is intended to capture relationships among tokens, including suspicious combinations of domain, path, and query terms.

4.4 Engineered URL Features

The feature extraction branch calculates selected URL characteristics, such as:

- Total URL length and hostname length
- Number of dots, hyphens, slashes, and digits
- Presence of an IP address in place of a domain name

- Number and composition of subdomains
- Suspicious terms and unusual token patterns
- Path and query-parameter characteristics

Features should be defined consistently and checked for missing or invalid values before model training.

4.5 Feature Fusion and Classification

The transformer representation is combined with the engineered feature vector. The fused representation is passed to a classification layer that outputs a phishing probability. A decision threshold converts the probability into a phishing or legitimate label. The threshold should be selected using validation data rather than the test set.

4.6 Evaluation

The model should be evaluated using precision, recall, F1-score, accuracy, and false-positive rate. Results should be compared with baselines such as an engineered-feature classifier and a transformer-only classifier. The evaluation should also examine errors and test the model on URLs not used during training.

5. PROPOSED FLOWCHARTS

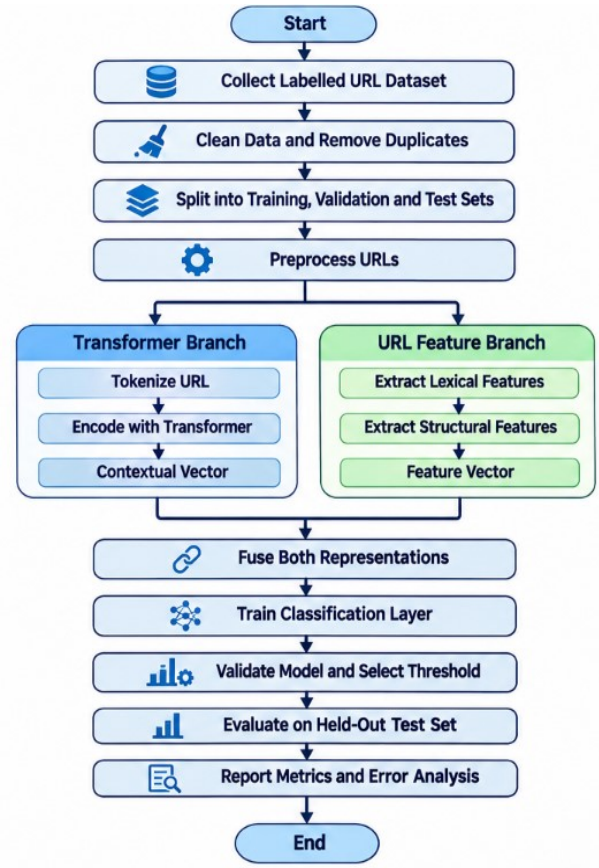


Figure 1. Proposed Training and Model Development Workflow

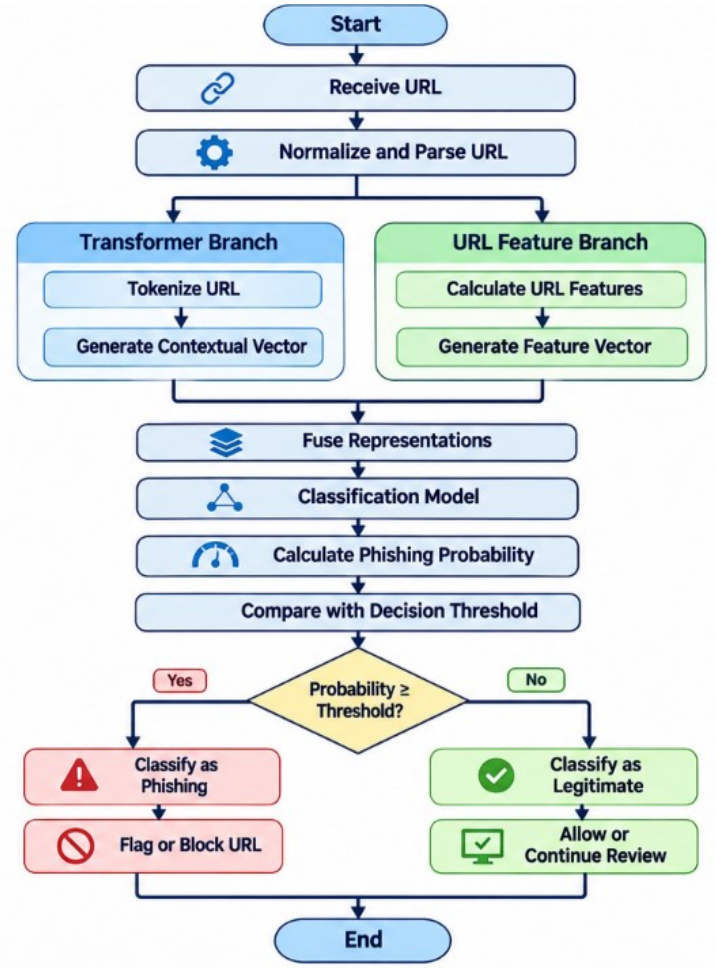


Figure 2. Proposed URL Detection and Classification Workflow

6. EXPECTED CONTRIBUTION

The proposed framework combines two complementary forms of URL information: contextual patterns learned by a transformer and explicit lexical or structural indicators. The expected contribution is a reproducible hybrid design that can be compared with transformer-only and feature-only baselines. Whether the combination improves detection performance, generalization, or false-positive rates remains an empirical question and should be reported only after testing.

7. CONCLUSION

This paper presents a proposed hybrid framework for detecting phishing websites using transformer-based language models and engineered URL features. The method processes URLs through parallel representation branches, fuses the resulting features, and classifies each URL as phishing or legitimate. The two flowcharts describe the model-development process and its intended detection workflow. Future evaluation should use a clearly documented dataset, leakage-aware data splitting, baseline comparisons, and performance metrics that account for both missed phishing URLs and false alarms.

REFERENCES

1. Maneriker, P., Stokes, J. W., Garcia Lazo, E., Carutasu, D., Tajaddodianfar, F., & Gururajan, A. (2021). URLTran: Improving phishing URL detection using transformers. MILCOM 2021—2021 IEEE Military Communications Conference, 197–204. https://doi.org/10.1109/MILCOM52596.2021.9653028

2. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345–357. https://doi.org/10.1016/j.eswa.2018.09.029

3. Do, N. Q., Selamat, A., Fujita, H., & Krejcar, O. (2024). An integrated model based on deep learning classifiers and pre-trained transformer for phishing URL detection. Future Generation Computer Systems, 161, 269–285. https://doi.org/10.1016/j.future.2024.06.031

4. Wang, C., et al. (2022). TCURL: Exploring hybrid transformer and convolutional neural network on phishing URL detection. Knowledge-Based Systems, 258, Article 109955. https://doi.org/10.1016/j.knosys.2022.109955

5. Rashid, F., Doyle, B., Han, S. C., & Seneviratne, S. (2024). Phishing URL detection generalisation using unsupervised domain adaptation. Computer Networks, 245, Article 110398. https://doi.org/10.1016/j.comnet.2024.110398

6. Haynes, K., et al. (2021). Lightweight URL-based phishing detection using natural language processing transformers for mobile devices. Procedia Computer Science, 191, 127–134. https://doi.org/10.1016/j.procs.2021.07.040

7. Aljofey, A., Jiang, Q., Qu, Q., Huang, M., & Niyigena, J. P. (2022). An effective phishing detection model based on character-level convolutional neural network from URL. Electronics, 11(9), 1514. https://doi.org/10.3390/electronics11091514

8. Le, H., Pham, Q., S time? (2018). URLNet: Learning a URL representation with deep learning for malicious URL detection. Proceedings of the 2018 IEEE Security and Privacy Workshops. https://doi.org/10.1109/SPW.2018.00026

9. Chiew, K. L., Tan, C. L., Wong, K. S., Yong, K. S. C., & Tiong, W. K. (2019). A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Information Sciences, 484, 153–166. https://doi.org/10.1016/j.ins.2019.01.064

10. Li, Y., et al. (2019). A stacking model using URL and HTML features for phishing webpage detection. Future Generation Computer Systems, 94, 27–39. https://doi.org/10.1016/j.future.2018.11.004

11. Bu, S., et al. (2021). Deep character-level anomaly detection based on a convolutional autoencoder for zero-day phishing URL detection. Electronics, 10(12), 1492. https://doi.org/10.3390/electronics10121492

12. Rasheed, A., et al. (2021). Adversarial attacks on featureless deep learning malicious URLs detection. Computers, Materials & Continua, 68(1), 921–939.

13. Al-Ahmadi, S., & AbuSharkh, M. (2022). PDGAN: Phishing detection with generative adversarial networks. IEEE Access, 10, 42459–42472. https://doi.org/10.1109/ACCESS.2022.3168235

14. Shirazi, H., et al. (2021). Towards performance of NLP transformers on URL-based phishing detection for mobile devices. International Journal of Ubiquitous Systems and Pervasive Networks.

15. Detecting phishing URLs using the BERT transformer model. (2023). 2023 IEEE International Conference on Big Data, 2483–2492. https://doi.org/10.1109/BigData59044.2023.10386782