UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Enhanced Phishing Webpage Detection

《DeepPhish-X: Multi-Modal Feature Engineering for Phishing Detection Using Hybrid Models of Computer Vision, Natural Language Processing, and Graph Neural Networks》

CANDIDATE : Cui Zhiwen
MATRIC NO : MCS241040

VENUE : BLOCK N28A
DATE : 2025
LECTURER : ASSOC. PROF. DR MOHD SHAHIZAN BIN OTHMAN

FACULTY OF COMPUTING,
UNIVERSITI TEKNOLOGI MALAYSIA

INTRODUCTION

BACKGROUND

- The rapid expansion of internet services has revolutionized how individuals and organizations communicate, conduct transactions, and access information . However, this growth has also led to an increase in cybersecurity threats, with phishing attacks becoming one of the most widespread and serious forms of online fraud. Phishing attacks involve creating fake websites that appear to be benign, in order to steal sensitive information such as passwords, credit card numbers, and personal identification details from users . As phishing techniques become increasingly sophisticated, detecting these fraudulent activities has become more challenging.
- Phishing websites often attempt to resemble the URLs of benign websites. For example, phishing websites may use domains or paths similar to those of benign websites to deceive users. In such cases, it is difficult to detect phishing solely by analyzing the URL. However, while phishing websites may imitate the URLs of benign websites, they rarely replicate the HTML structure completely .

PROBLEM STATEMENT

- Detecting phishing webpages is a critical task in the field of cybersecurity, with significant implications for online safety and data protection. Traditional methods have primarily relied on analyzing URL features, which can be limited in capturing the full context of phishing attacks. In this study, we propose an innovative approach that integrates HTML DOM graph modeling with URL feature analysis using advanced deep learning techniques. The proposed method leverages Graph Convolutional Networks (GCNs) to model the structure of HTML DOM graphs, combined with Convolutional Neural Networks (CNNs) and Transformer Networks to capture the character and word sequence features of URLs, respectively. These multi-modal features are then integrated using a Transformer network, which is adept at selectively capturing the interdependencies and complementary relationships between different feature sets. We evaluated our approach on a realworld dataset comprising URL and HTML DOM graph data collected from 2012 to 2024. This dataset includes over 80 million nodes and edges, providing a robust foundation for testing. Our method demonstrated a significant improvement in performance, achieving a 7.03 percentage point increase in classification accuracy compared to state-of-the-art techniques. Additionally, we conducted ablation tests to further validate the effectiveness of individual features in our model. The results validate the efficacy of integrating HTML DOM structure and URL features using deep learning. Our framework significantly enhances phishing detection capabilities, providing a more accurate and comprehensive solution to identifying malicious webpages.

LITERATURE REVIEW

LITERATURE REVIEW

Dhamija, R.; Tygar, J.D.; Hearst, M. Why phishing works. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 22–27 April 2006; pp. 581–590.

Yoon, J.-H.; Buu, S.-J.; Kim, H.-J. Phishing Webpage Detection via Multi-Modal Integration of HTML DOM Graphs and URL Features Based on Graph Convolutional and Transformer Networks. Electronics

Lee, J.; Wang, J.; de Guzman, M.C.; Gupta, M.; Rao, H.R. Can I Help Prevent Data Breaches in the Workplace? From Routine Activities to Extra-Role Security Behaviors. IEEE Trans. Technol. Soc. 2024. Early Access.

Alsharnouby, M.; Alaca, F.; Chiasson, S. Why phishing still works: User strategies for combating phishing attacks. Int. J. Hum.-Comput. Stud. 2015, 82, 69–82.

METHODOLOGY

The Challenge of Phishing Detection

Traditional URL-based methods often fall short in capturing the full context of sophisticated phishing attacks. Phishing websites can mimic legitimate URLs, making detection difficult.

Our research highlights that while URLs might appear similar, the underlying HTML DOM structures often reveal significant differences between benign and phishing sites.

Case	Ground Truth	URL
(a)	Benign	https://script.google.com
(b)	Phishing	https://script.google.com/macros/s/AKfycbyjt18r7uGwlzNe6SekQE0yCNgfYHE5mHe1ib-9SIKfuT0KKTu8oaCrdXoXhbg3ixjl/exec

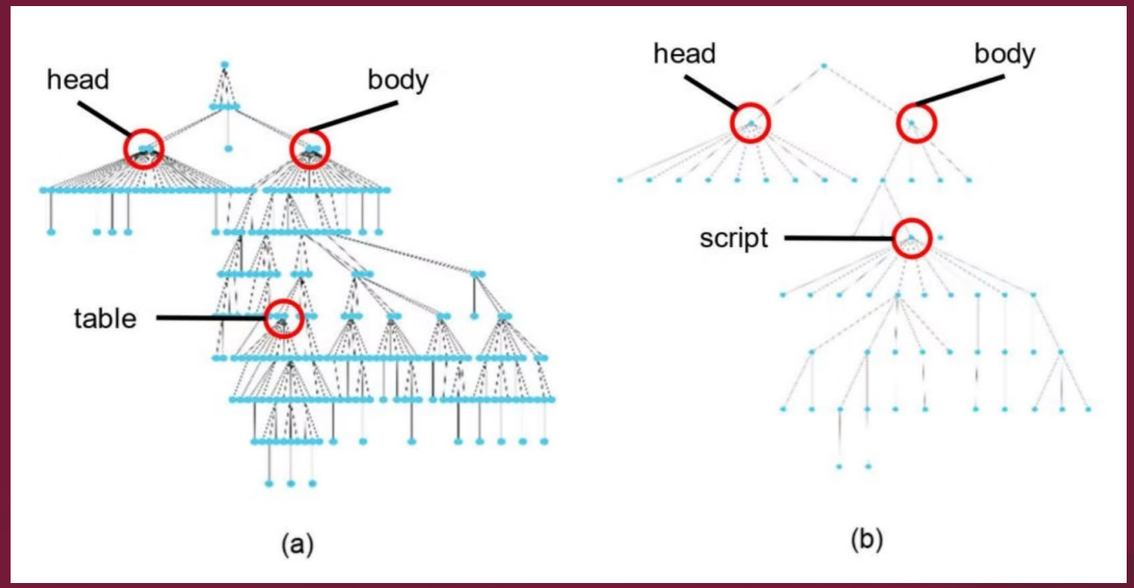
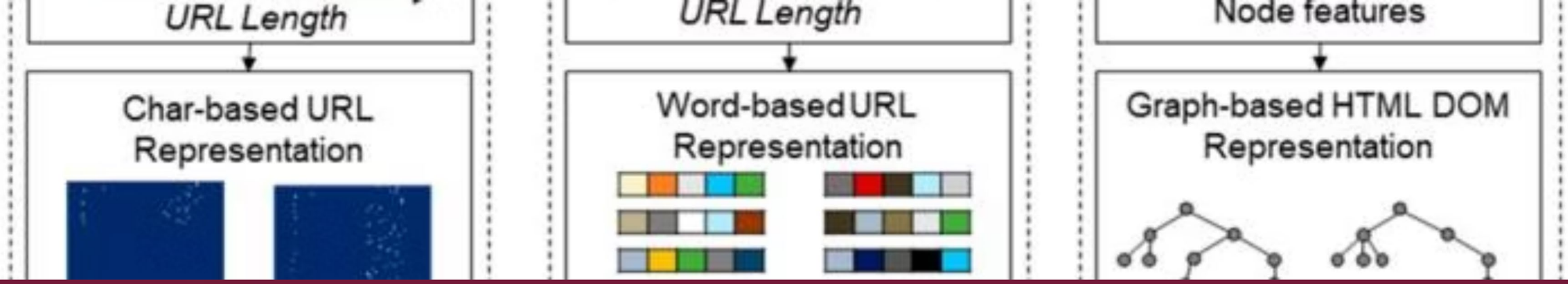


Figure 2. Comparison of HTML DOM structures: (a) benign vs. (b) phishing, showing structural differences despite similar URLs.



Our Innovative Multi-Modal Approach

We propose an innovative framework that integrates HTML DOM graph modeling with URL feature analysis using advanced deep learning techniques.

HTML DOM Graphs

Leveraging Graph Convolutional Networks (GCNs) to model the structural nuances of HTML DOM trees.

URL Features

Utilizing CNNs and Transformer Networks to capture character and word sequence features from URLs.

Multi-Modal Integration

A Transformer network selectively captures interdependencies and complementary relationships between feature sets.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Data Representation and Feature Extraction

Our method meticulously processes URL and HTML data to extract critical features for robust phishing detection.

Char-Based URL Features

URLs are converted into a 128x128 one-hot encoded matrix, processed by CNNs to capture local dependencies and spatial hierarchies.

$$\mathbf{H}_l = f(\phi(\mathbf{W}_l, \mathbf{X}) + \mathbf{b}_l)$$

Word-Based URL Features

URLs are segmented into tokens, filtered, and transformed into integer sequences for Transformer network processing, capturing contextual relevance.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

DOM Graph Features

HTML documents are parsed into DOM trees, represented as graphs where nodes are elements and edges are parent-child relationships. GCNs capture relational and structural information.

$$\mathbf{H}(l+1) = \text{ReLU}(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)})$$

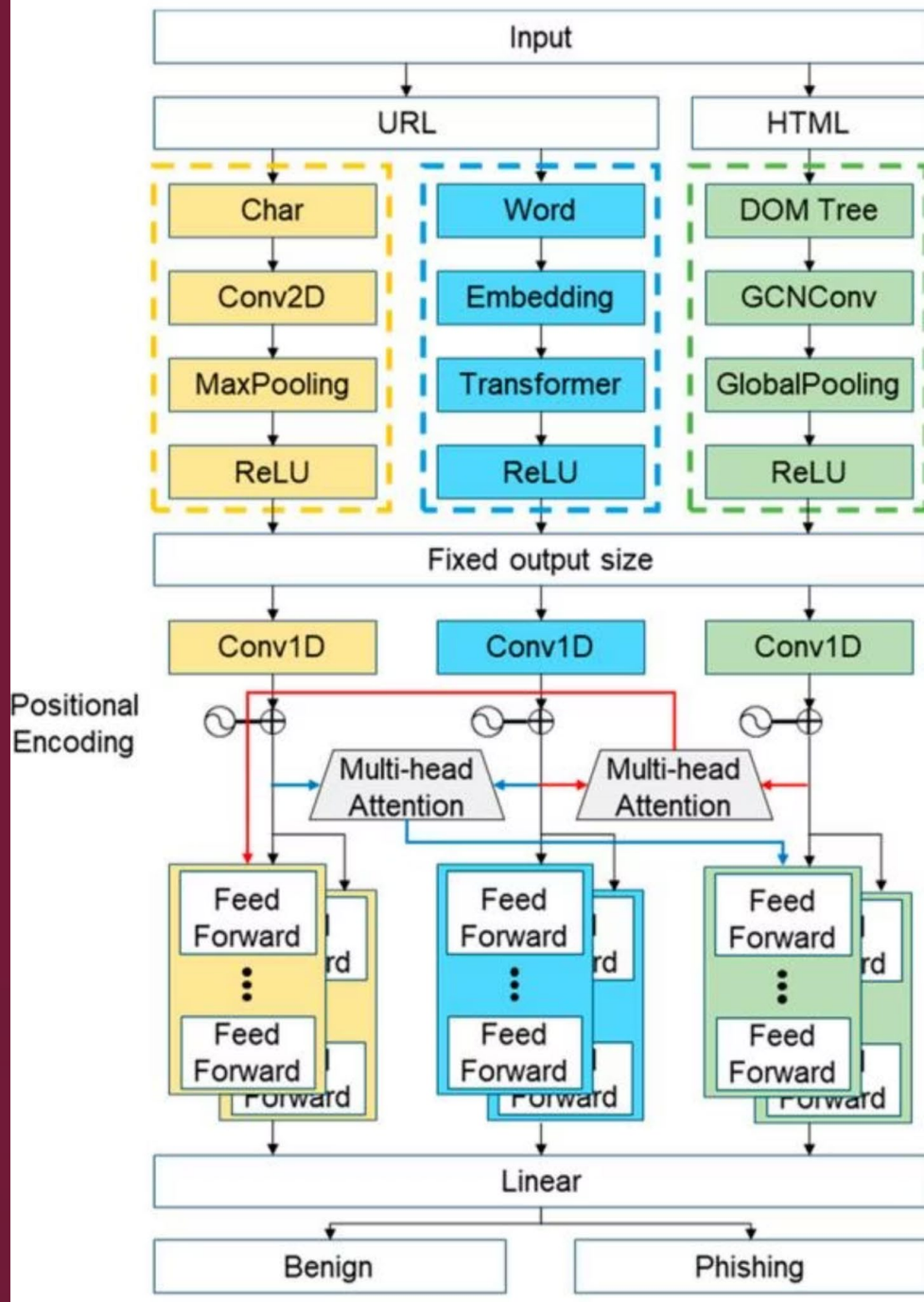
UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Ensemble Model Architecture

Our ensemble classifier integrates CNNs, Transformers, and GCNs to leverage complementary feature representations for enhanced accuracy.

The outputs from each component (char-based URL, word-based URL, and HTML DOM graph features) are concatenated and fed into a Transformer layer for final classification.

$$\mathbf{H}_{\text{concat}} = \text{Concat}(\mathbf{H}_{\text{char}}\mathbf{H}_{word}\mathbf{H}_{graph})\mathbf{Z} =$$
$$\mathbf{W}_{fc}\,\text{Transformer}(\mathbf{H}_{concat}) + \mathbf{b}_{fc}\,\mathbf{P} = \text{softmax}(\mathbf{Z})$$



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Experimental Results and Performance

Our model was evaluated on a real-world dataset from 2012-2024, comprising over 80 million nodes and edges, demonstrating significant performance improvements.

Method	Accuracy	Precision	Recall	F1 Score
CNN	0.8952	0.8626	0.8952	0.8736
Transformer	0.8868	0.8859	0.8868	0.8856
GCN	0.8739	0.8684	0.8740	0.8605
PhishDet	0.9853	0.9522	0.9721	0.9620
Ours	0.9812	0.9658	0.9765	0.9709

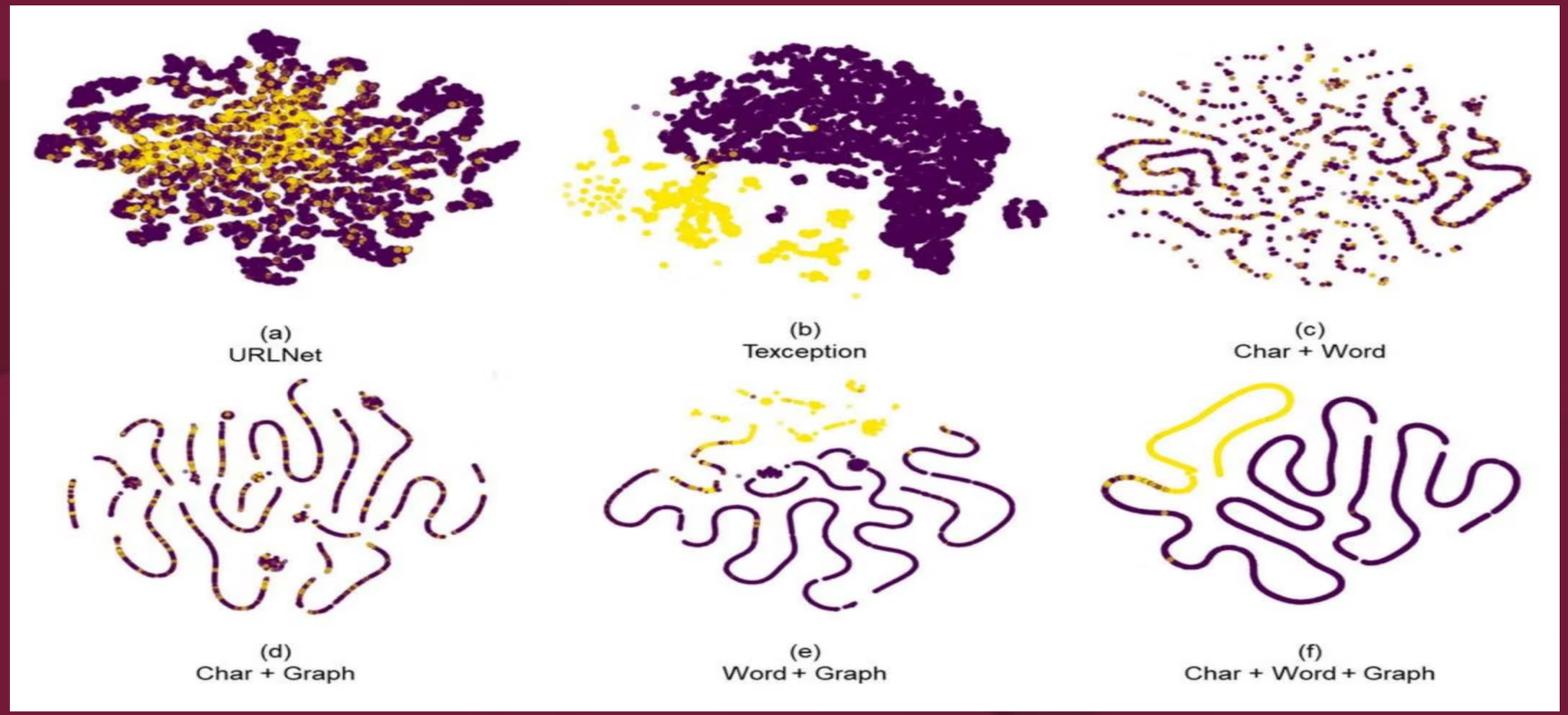
Our ensemble model achieved a 7.03 percentage point increase in classification accuracy compared to state-of-the-art techniques, with up to a 22% improvement in precision and 23% in recall.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Ablation Study and Feature Importance

Our ablation study confirms the critical importance of integrating all three feature types for optimal phishing detection performance.

Data	Char-Based URL	Word-Based URL	HTML DOM Graph	Accuracy	Precision	Recall	F1 Score
	✔			0.8952	0.8626	0.8952	0.8736
		✔		0.8868	0.8859	0.8868	0.8856
			✔	0.8739	0.8684	0.8740	0.8605
	✔	✔		0.9540	0.9167	0.8544	0.8844
	✔		✔	0.9817	0.9436	0.9647	0.9541
		✔	✔	0.9789	0.9245	0.9730	0.9481
✔	✔	✔	✔	0.9884	0.9677	0.9759	0.9718



Conclusion and Future Directions

Our study demonstrates the efficacy of integrating HTML DOM structure and URL features using deep learning, significantly enhancing phishing detection capabilities.

Key Contributions

First to combine URL features and HTML DOM graph features with a neural network that effectively merges these complementary characteristics.

Limitations

Reliance on static features, misclassification of complex benign URLs, and increased computational complexity in real-time applications.

Future Work

Incorporate additional HTML features, user behavior analysis, and adversarial learning to improve robustness and adaptability against evolving phishing tactics.

THANK YOU

univteknologimalaysia utm.my utmofficial