**Response to Reviewers' Comments**

Dear Editor and Dear Reviewers,

We are submitting the manuscript titled "**FD-SGIL: Fraudulent Website Detection Model based on Structure Growing Incremental Learning**" to IEEE Transactions on Information Forensics and Security. This submission is related to a previously rejected manuscript—originally titled "**FD-SGIL: Fraudulent Website Detection Model based on Structure Growing Incremental Learning**" (**Manuscript ID: IoT-48797-2025**)—which was previously submitted to IEEE Internet of Things Journal and subsequently rejected.

We highly value the constructive comments from the previous review process at IEEE Internet of Things Journal, as they have been critical in enhancing the quality of our manuscript. In this submission to IEEE Transactions on Information Forensics and Security, we have thoroughly addressed all feedback from the prior reviewers, and the following content provides a detailed, point-by-point response to those comments. To ensure clarity for your review, we have formatted the previous reviewers' comments in bold red with clear headings, and marked all revised content in the manuscript in black. Moreover, for each response, we list the corresponding references. The indexes of the references cited in the manuscript correspond to Arabic number. The key revisions to the manuscript are summarized as follows:

1. First, we have supplemented the latest statistical data from the FBI Internet Crime Report [7], added a threat model, and conducted a detailed analysis of the attack methods of fraudulent websites as well as the necessity of researching them. For specific modification details, please refer to the responses to **Comments 2** and **3** in the major comments from **Reviewer #1**.
2. Second, we have reorganized and analyzed the latest key studies in the fields of phone fraud detection, phishing website detection, and incremental learning (e.g., [15], [34], and [47]). This addresses the issue that the literature review insufficiently covers important recent works. Meanwhile, we have conducted in-depth discussions on the newly added literature in the introduction and related work sections to strengthen the demonstration of research background and innovation. For specific modification details, please refer to the response to **Comment 4** in the major comments from **Reviewer #1** and the response to **Comment 7** in "**Comments on Introduction (Section I)** and **Related Work (Section II)**" from **Reviewer #2**.
3. Third, we have supplemented key details of the detection network. Specifically, we added formula descriptions in the problem definition part of the research content. Meanwhile, we have redrawn the structural diagram of the entire detection network and supplemented its key detection details. This fills the gap in the description of the detection network—where key details and mathematical modeling were missing—and makes the implementation logic of the model clearer and more complete. For specific modification details, please refer to the response to **Comment 7** in the major comments from **Reviewer #1** and the response to **Comment 5** from **Reviewer #2**.
4. Fourth, we have supplemented detailed statistical information about the dataset. A new dataset statistics table has been added, which clearly presents the number of samples, class distribution, types of features used, and feature extraction methods. Additionally, considering that the dataset includes 7 categories of fraudulent websites, we explained the process of using the "structure-growing" incremental learning strategy (each incremental task contains data from 2 categories) to simulate training for unknown attack strategies, so as to prove the generalization ability of the model for emerging fraud types. For specific modification details, please refer to the response to **Comment 7** in the major comments from **Reviewer #1** and the response to **Comment 5** from **Reviewer #2**.
5. Fifth, we have enhanced the experimental evaluation. Specifically, we introduced two new evaluation metrics—mean accuracy (mACC) and backward transfer (BWT)—to better measure the effectiveness of incremental detection. Furthermore, we supplemented comparative experiments against baseline methods from both telephone fraud detection and phishing website detection to demonstrate the broader applicability of our method. In addition, we conducted systematic comparisons between FD-SGIL and representative incremental learning baselines under the mACC and BWT metrics. Finally, we carried out ablation studies on the Effectiveness of FD-SGIL Components and the Impact of Multi-source Fused Features, which further validate the contributions of our design. For specific modification details, please refer to the responses to **Comments 3**, **4**, and **8** in the comments from **Reviewer #2**.
6. Besides, we have checked and double-checked the spelling, grammar and punctuation of the whole manuscript in the revised version. Meanwhile, we have checked the experiments for ensuring the correctness of the results, and updated some data. Last but not least, let us thank deeply AE and Reviewers for their constructive comments.

References:

[7] “2024 internet crime report,” Internet Crime Complaint Center, Federal Bureau of Investigation, USA, pp. 1–47, 2024.

[15] A. Tong, B. Chen, Z. Wang, J. Gao, and C. K. Lam, “Gdfgat: Graph attention network based on feature difference weight assignment for telecom fraud detection,” PLoS One, vol. 20, no. 5, p. e0322004, 2025. 18720–18729.

[34] F. Rashid, N. Ranaweera, B. Doyle, and S. Seneviratne, “Llms are one shot url classifiers and explainers,” Computer Networks, vol. 258, p. 111004, 2025.

[47] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11 858–11 867.

**Response to Major Comments from Reviewer #1**

Reviewer’s major comment 1:

The paper is outside of the scope of the journal.

We sincerely thank the reviewer for this important comment. We respectfully clarify that our work is closely related to the IoT security domain. Telecom fraud is no longer confined to traditional telephony or computer networks; it increasingly occurs via mobile devices, browsers, and other IoT-enabled endpoints. Users accessing websites through smartphones, tablets, or other IoT devices are all potentially exposed to fraudulent websites, and these users are distributed across diverse IoT application scenarios, including smart homes, wearable devices, and mobile payment platforms.

Our proposed FD-SGIL framework directly addresses these challenges by enabling continuous and adaptive detection of emerging fraudulent websites. By incrementally learning new fraud types while retaining knowledge of previously seen ones, the framework can adapt to the rapidly evolving threat landscape typical in IoT environments, where static detection models quickly become outdated. Additionally, the multi-source feature fusion strategy—integrating URL, webpage content, and WHOIS features—is conceptually analogous to IoT security scenarios that require combining heterogeneous data streams, such as device logs, network traffic, and contextual metadata.

We acknowledge that in the original manuscript, the relationship between telecom fraud and IoT was not explicitly emphasized, which may have led to the perception that the work was outside the journal’s scope. We have since thoroughly revised the manuscript to clarify these connections, and we believe the current version clearly demonstrates the relevance of our contributions to IoT cybersecurity. Therefore, we intend to submit this improved version to IEEE Transactions on Information Forensics and Security and sincerely hope the reviewers support this decision.

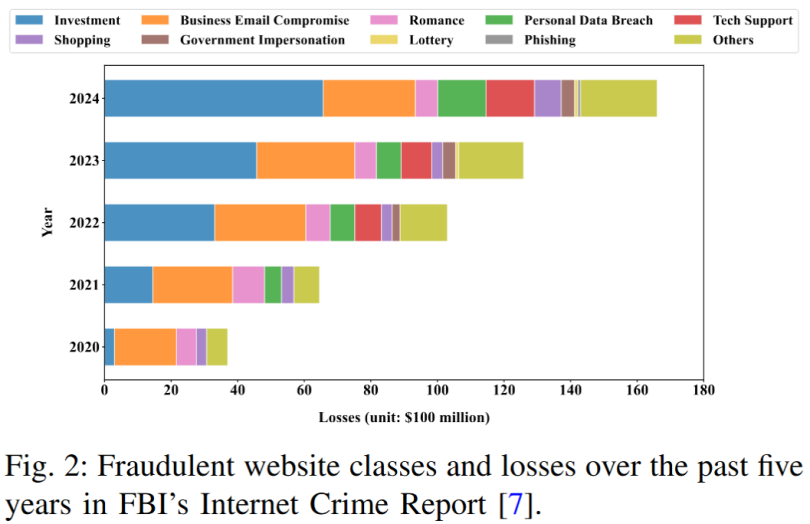
Reviewer’s major comment 2:

The introduction lacks proper definition of the research problem. A deeper exploration of the seriousness of the phishing URLs problem, along with some recent statistics could really help establish the need for this research.

We would like to sincerely thank the reviewer for the insightful suggestion. We have carefully revised the introduction to address both the lack of a clear research problem definition and the insufficient analysis of the seriousness of phishing URLs.

First, we have explicitly defined the research problem in the **section I. INTRODUCTION of Page 1**. Fraudulent websites are now clearly described as malicious web resources intentionally constructed to deceive users into disclosing sensitive information or transferring money. These fraudulent tactics extend beyond conventional phishing schemes, encompassing a broader spectrum of deceptive practices such as investment fraud and business email compromise. Furthermore, we emphasize that such fraudulent methodologies are not static but continuously emerging and evolving, which makes them particularly challenging to detect.

Second, we have reinforced the motivation by analyzing the seriousness of this threat in the **section I. INTRODUCTION of Pages 1~2**. We incorporated recent statistics from the FBI’s Internet Crime Report [7], which shows that the financial losses associated with fraudulent websites in the United States alone had exceeded 16 billion USD by 2024 (see Fig. 2 for illustration). To further illustrate the impact, we summarized the growing diversity of attack categories—including investment scams, phishing fraud, and shopping fraud—and highlighted their increasing economic and social harm. This evidence clearly demonstrates the urgent need for more effective detection techniques tailored to fraudulent websites.



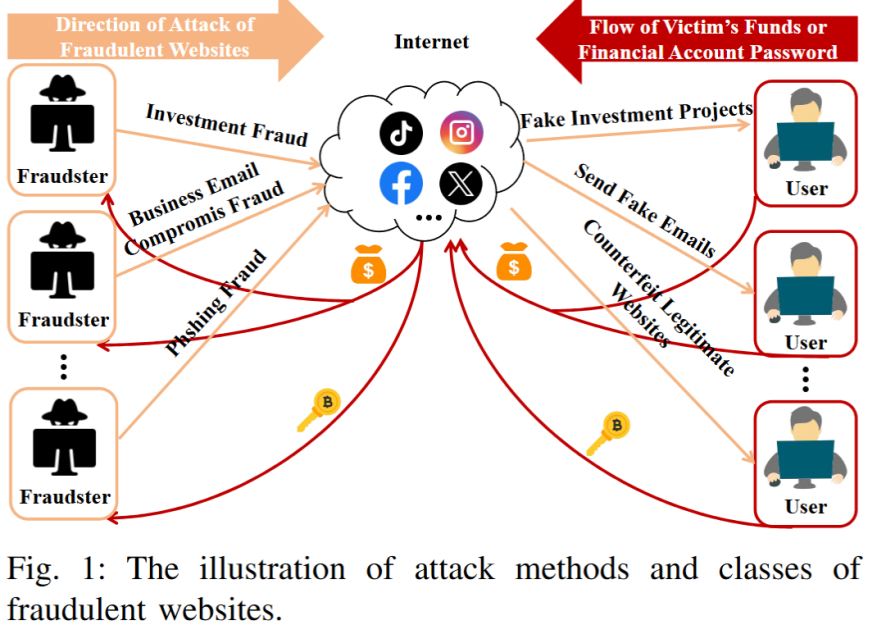
References:

[6] “Research report on telecommunication network fraud governance in 2021,” Tencent, Shenzhen, China, pp. 1–53, 2022.

Reviewer’s major comment 3:

The paper needs to include a proper threat model either in the introduction of in a separate section.

We thank the reviewer very much for providing the instructive comment and valuable suggestion. In the introduction, we have organized and summarized the attack methods and types of fraudulent websites, and presented their threat model.



Fraudulent websites: Fraud is perpetrated by building fraudulent websites on the Internet and luring users into visiting them. As shown in Fig. 1, the fraudulent tactics employed by these websites not only encompass traditional phishing fraud but also include schemes such as investment fraud and romance fraud, with new fraud tactics continuously evolving.

The above modified content is placed in the section **I.INTRODUCTION of Page 1** in the revised version.

Reviewer’s major comment 4:

Literature review is missing a few important recent works that combine the WHOIS features with lexical features of the URL with higher accuracy and other performance metrics. This needs to be expanded.

We thank the reviewer’s helpful comment and valuable suggestion. We have reorganized the latest review studies relevant to this research, including those focusing on telephone fraud detection, phishing website detection, and incremental learning methods. Furthermore, we have conducted a detailed analysis of the shortcomings of existing review studies in addressing fraudulent website detection. Meanwhile, we have supplemented the Motivation to clarify the rationale behind this research.

III. MOTIVATION

Traditional research on telecom fraud detection has primarily focused on telephone fraud, with a lack of studies dedicated to fraudulent website detection. The data structure of fraudulent websites lacks inherent graph topological characteristics, making traditional GNN-based telephone fraud detection methods [4], [8], [12] difficult to apply directly. Existing phishing website detection methods [16], [48], [49] mainly rely on URL structures, assisted by HTML and WHOIS records for binary classification. The diverse classes and attack patterns of fraudulent websites make it difficult for traditional phishing website detection methods to be directly applicable. Most existing incremental learning methods [39], [41], [46], [47] are designed for balanced datasets, while the highly imbalanced characteristics of fraudulent website data exacerbate the model's catastrophic forgetting problem [50].

To this end, this paper proposes the FD-SGIL method. First, a multi-source feature fusion framework for URLs, webpage content, and WHOIS records is constructed based on BERT, which improves the accuracy of multi-classification detection of fraudulent websites. Second, an incremental learning strategy with dynamic structure adjustment is adopted to dynamically expand detection branches for unseen classes of fraudulent websites. Finally, a transfer loss function is designed to dynamically adjust the loss weights of misclassified samples according to sample distribution, effectively alleviating data imbalance and forgetting problems.

The above modified content is placed in section **Ⅱ. RELATED WORK of Pages 2~4 and** section **Ⅲ.MOTIVATION of Page 4** in the revised version.

References:

1. X. Hu, H. Chen, S. Liu, H. Jiang, G. Chu, and R. Li, “Btg: A bridge to graph machine learning in telecommunications fraud detection,” Future Generation Computer Systems, vol. 137, pp. 274–287, 2022.

[8] X. Hu, H. Chen, J. Zhang, H. Chen, S. Liu, X. Li, Y. Wang, and X. Xue, “Gat-cobo: Cost-sensitive graph neural network for telecom fraud detection,” IEEE Transactions on Big Data, vol. 10, no. 4, pp. 528–542, 2024.

[12] X. Hu, H. Chen, H. Chen, S. Zhang, S. Liu, and X. Li, “Telecom fraud detection via imbalanced graph learning,” in 2022 IEEE 22nd International Conference on Communication Technology (ICCT). IEEE, 2022, pp. 1312–1317.

[16] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, and M. Wo´zniak, “Accurate and fast url phishing detector: a convolutional neural network approach,” Computer Networks, vol. 178, p. 107275, 2020.

[39] Z. Li and D. Hoiem, “Learning without forgetting,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 12, pp. 2935–2947, 2017.

[41] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 2001–2010.

[46] S. Yan, J. Xie, and X. He, “Der: Dynamically expandable representation for class incremental learning,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 3014–3023.

[47] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11 858–11 867.

[48] B. Sabir, M. A. Babar, R. Gaire, and A. Abuadbba, “Reliability and robustness analysis of machine learning based phishing url detectors,” IEEE Transactions on Dependable and Secure Computing, 2022.

[49] A. Aljofey, Q. Jiang, A. Rasool, H. Chen, W. Liu, Q. Qu, and Y. Wang, “An effective detection approach for phishing websites using url and html features,” Scientific Reports, vol. 12, no. 1, p. 8842, 2022.

[50] Y. Yang, D. Ren, C. Peng, J. Huo, W. Li, and Y. Gao, “Dynamic replay training for class-incremental learning,” in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024, pp. 5915–5919.

Reviewer’s major comment 5:

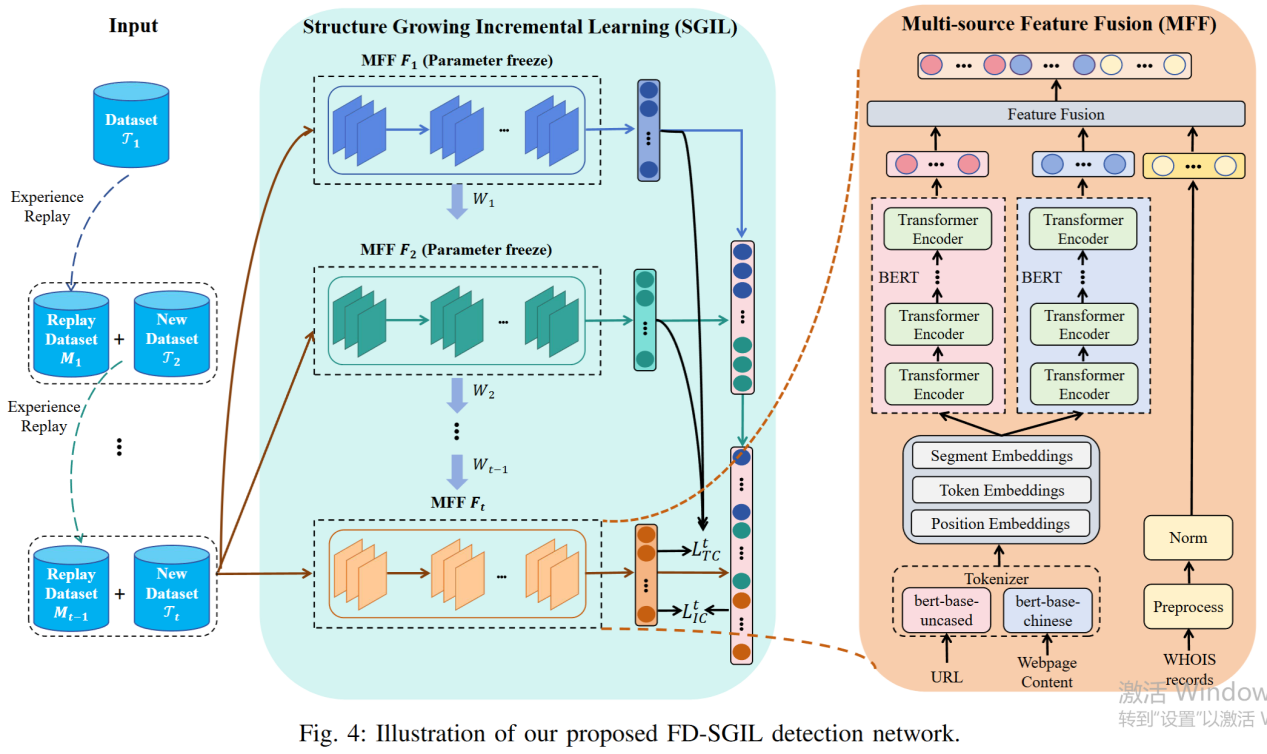
Mentioning specific countries with "weak legal oversight" without proper support of these claims is not scientifically correct. Why mention countries? The point could've been delivered without mentioning country names.

We would like to thank the reviewer very much for providing this helpful comment. We have realized that it is highly inappropriate to mention specific countries as having "weak legal oversight," so we have removed all content referring to specific countries labeled as such.

Reviewer’s major comment 6:

The description of the detection network is missing some important details. The description and mathematical modeling needs to be expanded.

We would like to thank the reviewer very much for providing this helpful comment. We have supplemented the key details of the detection network. Specifically, we have added formula descriptions in the problem definition. Meanwhile, Meanwhile, we have redrawn the structural diagram of the entire detection network and supplemented its key detection details.

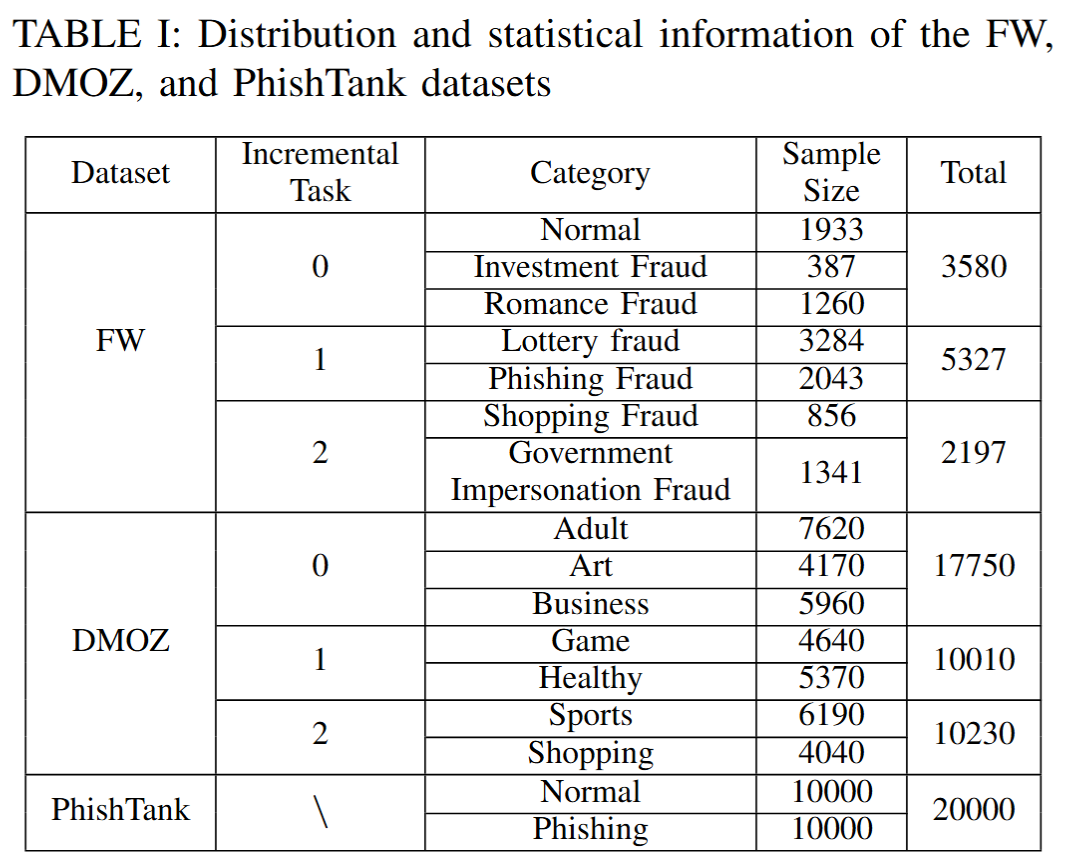


The above modified content is placed in the section **Ⅳ.METHOD of Pages 4~7** in the revised version.

Reviewer’s major comment 7:

In section IV.A there is absolutely no explanation of the features included in these datasets. No discussion of class balance or imbalance. This significantly undermines the validity of the obtained results. Did all of these datasets have only URLs? or complete featuresets? Did they contain the same features? How were these features extracted?

We would like to thank the reviewer for this constructive comment. In the revised manuscript, we have provided a detailed introduction of the three datasets, including the types of features contained in each dataset as well as the statistical distribution of the samples, as summarized in Table 1. We have also explained the feature extraction process in detail. To address class imbalance, we designed a gradient-based distillation loss, which effectively alleviates this problem.



Specifically, the PhishTank and DMOZ datasets only contain URL strings. Since most of the PhishTank URLs are inactive and the DMOZ dataset consists entirely of legitimate websites, webpage content and WHOIS records cannot be utilized. By contrast, the FW dataset contains active URLs, allowing us to extract not only URLs but also webpage content and WHOIS records, thus providing richer multi-source features. While all three datasets contain URL features, only the FW dataset includes additional webpage content and WHOIS features. Therefore, the proposed multi-source feature fusion method (MFF) is mainly evaluated on the FW dataset, whereas PhishTank and DMOZ serve as benchmarks for URL-based detection tasks.

The above modified content is placed in subsection **C. Multi-source Feature Fusion Method of Pages 5~6** and subsection **A. Experimental Setup of Page 9** in the revised version.

Reviewer’s major comment 8:

The results obtained are not validated. I'd suggest aproper validation method such as 10-fold cross-validation.

We would like to thank the reviewer very much for providing this helpful comment. In this study, we did not employ 10-fold cross-validation. Instead, we adopted a train/test split strategy, where the datasets were divided into separate subsets to ensure fair evaluation of model performance. To evaluate the model comprehensively, we report mean accuracy (mACC), which measures the average performance across all previously learned tasks, and backward transfer (BWT), which assesses the impact of learning new tasks on the performance of previously learned tasks in incremental learning. This evaluation protocol is consistent with prior studies in the incremental learning literature.

**Response to Minor Comments from Reviewer #1**

Reviewer’s minor comment 1:

The language of the paper needs intensive review to improve readability and eliminate grammatical errors.

We would like to thank the reviewer very much for providing this helpful comment. We have carefully revised the manuscript to improve readability and eliminate grammatical errors throughout the paper.

Reviewer’s minor comment 2:

The first time the term is mentioned, it needs to be in expanded form. For example, the first work in the introduction is IoT without the long form. The same applies to many acronyms mentioned in the paper across the sections.

We would like to thank the reviewer very much for providing this helpful comment. We have ensured that all acronyms are expanded at their first occurrence across the manuscript, and consistently applied this practice in all sections.

Reviewer’s minor comment 3:

Figure 2 is too small to be legible. This is the core of the presented work. It needs to be given adequate space.

We would like to thank the reviewer very much for providing this helpful comment. We have enlarged Figure 2 in the revised manuscript and adjusted its layout to occupy adequate space, making all details clearly legible. Last but not least, let us thank deeply **Reviewer #1** for his/her constructive reviews.

**Response to Comments from Reviewer #2**

Reviewer’s comment 1:

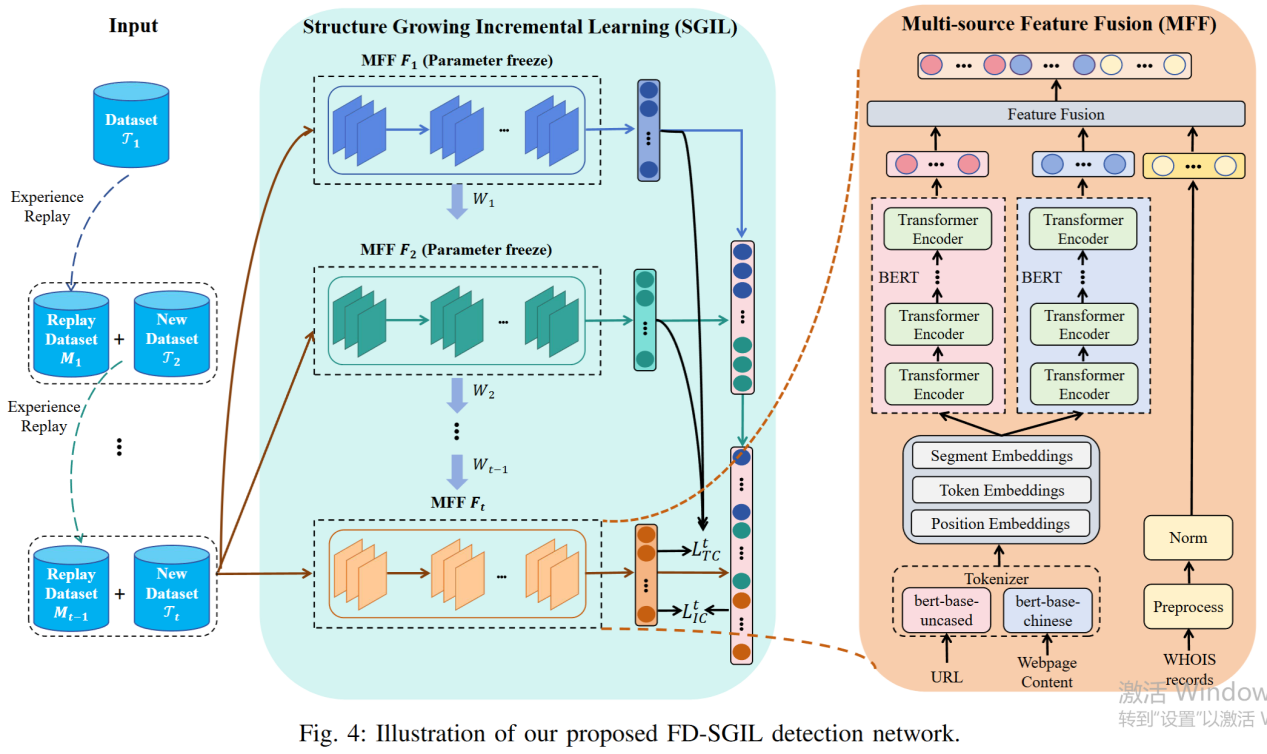
Dependence on BERT for Feature Extraction BERT, while powerful, is computationally expensive and may not be optimal for real-time detection on resource-constrained IoT devices. This could limit deployment scalability, especially for edge devices with limited processing power.

We would like to thank the reviewer very much for providing this helpful comment. We acknowledge that BERT is relatively resource-intensive and may not be optimal for latency-sensitive IoT deployments. However, in this study our primary objective is to validate the effectiveness of the proposed incremental learning framework for fraudulent detection. To this end, we adopt BERT as the feature extraction backbone because its strong representation ability provides more accurate and semantically rich features, thereby allowing us to better evaluate the contribution of the framework itself.

Reviewer’s comment 2:

Structure-Growing Incremental Learning Complexity The "structure-growing" mechanism may introduce additional hyperparameters and architectural complexity. This could make the model harder to tune and deploy in practice, requiring more expertise to maintain.

We would like to thank the reviewer very much for providing this helpful comment. We acknowledge that the structure-growing incremental learning mechanism introduces additional hyperparameters and architectural components. However, given the dynamic and evolving nature of fraudulent websites, it is essential for a detection model to continuously adapt to new attack types while maintaining performance on previously seen ones. The SGIL strategy is specifically designed to achieve this balance: by allowing the model to expand its structure when encountering new fraud patterns, it better preserves knowledge of earlier tasks and mitigates catastrophic forgetting. To facilitate deployment, we have redrawn the overall framework diagram of the FD-SGIL detection method (see Fig. 4), and we have also documented all hyperparameter settings and training configurations in the revised manuscript to ensure reproducibility and simplify future deployment.

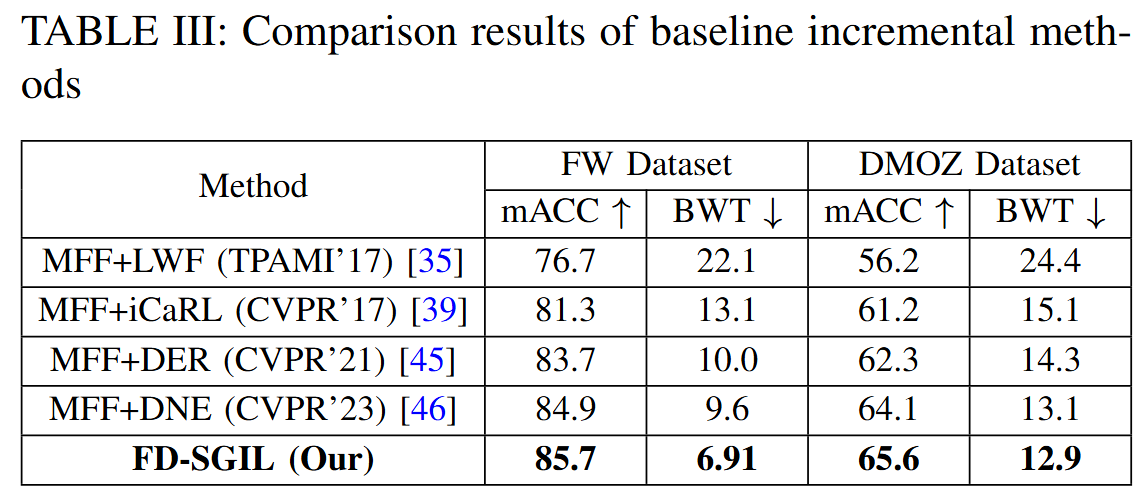


The above modified content is placed in subsection **C. Multi-source Feature Fusion Method of Pages 5~6**，subsection **D. Structure Growing Incremental Learning Method of Pages 6~7**，and **subsection A. Experimental Setup of Page 10** in the revised version.

Reviewer’s comment 3:

Catastrophic Forgetting Mitigation Claims While the paper mentions a "transfer loss mechanism," incremental learning methods often struggle with severe forgetting when new classes are drastically different from old ones. If the new fraudulent website types are highly dissimilar to previous ones, performance degradation may still occur.

We would like to thank the reviewer very much for providing this helpful comment. We acknowledge that catastrophic forgetting remains challenging when new fraudulent types are highly dissimilar to prior ones. Our proposed framework combines a dynamically weighted transfer loss with experience replay, which together aim to preserve past knowledge while learning new classes. Empirical results further support this design. As shown in Table Ⅲ, FD-SGIL consistently achieves the highest overall performance, obtaining optimal final mACC and BWT metrics while maintaining stable mACC values across all incremental tasks.

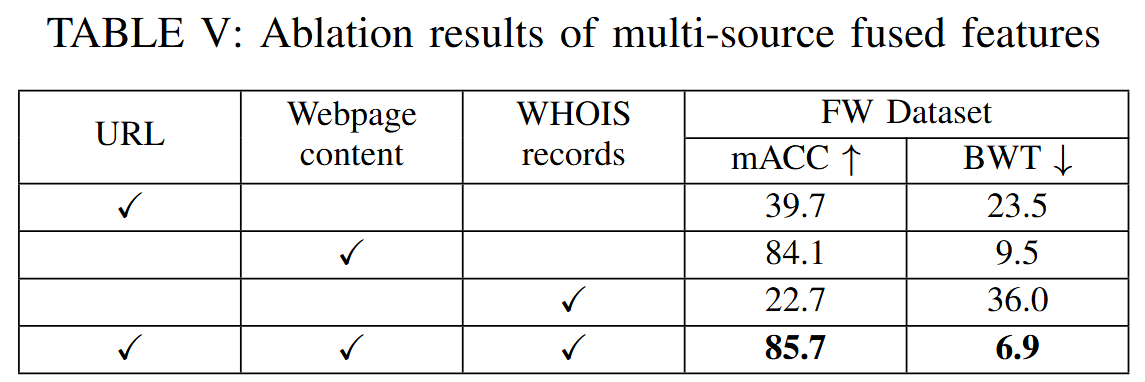


The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reviewer’s comment 4:​

Multi-Source Feature Fusion Overhead Combining URL, webpage content, and WHOIS records requires significant data preprocessing and alignment. This may slow down real-time detection, particularly if WHOIS data retrieval is delayed or unavailable.

We would like to thank the reviewer very much for providing this helpful comment. In our MFF method, URL and webpage content are represented using BERT, while WHOIS records are transformed into feature vectors through a simple feature mapping. These three representations are then concatenated for fusion. This process introduces only minor computational overhead compared to baseline methods and does not significantly slow down real-time detection.



Furthermore, we have added an ablation study on the impact of multi-source fused features (see Table Ⅴ). The results show that even when WHOIS data is unavailable, our detection method still achieves competitive and robust performance. This demonstrates that the framework remains effective under practical scenarios where WHOIS retrieval may be delayed or incomplete.

The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

Reviewer’s comment 5:​

Dataset Generalizability The experiments are conducted on only three datasets, which may not cover all emerging fraudulent website types. The model might not generalize well to unseen or rapidly evolving attack strategies.

We would like to thank the reviewer very much for providing this helpful comment. We acknowledge that the three datasets used in our experiments cannot fully cover all emerging fraudulent website types. However, the primary contribution of this work is the proposal of the FD-SGIL framework, which is specifically designed to incrementally detect new types of fraudulent websites while maintaining performance on previously learned ones.

To this end, we have additionally collected and constructed the FW dataset, which contains seven of the most harmful categories of fraudulent websites. We conducted extensive experiments on this dataset, and the results demonstrate the effectiveness of FD-SGIL in handling diverse and evolving fraud types. For newly emerging categories that were not covered in the present study, our framework can readily incorporate them through incremental training in future work.

Reviewer’s comment 6:​

Label Scarcity in Incremental Learning Fraudulent websites evolve quickly, but obtaining labeled data for new classes in real-world scenarios is challenging. The incremental learning performance may degrade if few labeled examples are available for new attack types.

We would like to thank the reviewer very much for providing this helpful comment. In our work, we have considered the problem of sample imbalance and label scarcity among different fraudulent website types. To address this challenge, we design a dynamic adjusted transfer loss that selects more suitable soft labels for each training sample during the distillation process, thereby alleviating catastrophic forgetting. At the same time, the loss weight of misclassified samples is increased while that of correctly classified samples is decreased, which helps mitigate the negative impact of imbalanced training data. This strategy enables the FD-SGIL to maintain balanced learning and stable detection performance even under label scarcity conditions.

Reviewer’s comment 7:​

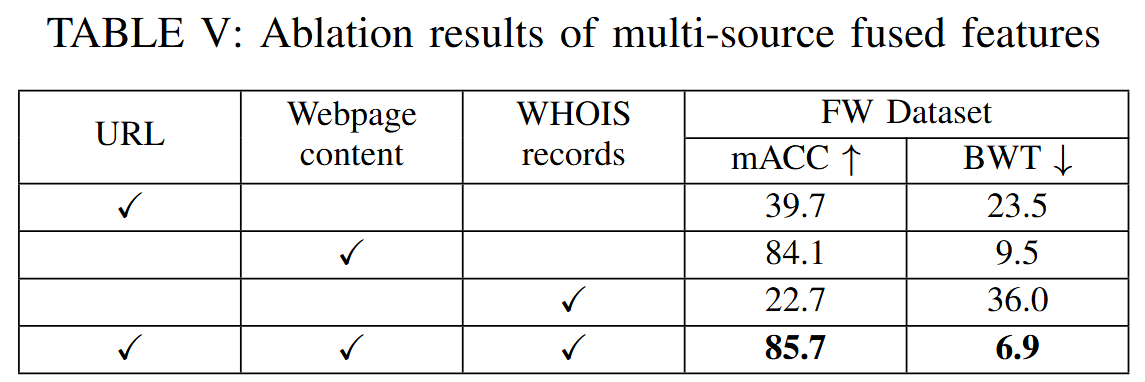
Real-World Deployment Feasibility The abstract does not discuss computational latency or model size, which are critical for IoT deployments. If the model is too large or slow, it may not be practical for real-time use in IoT gateways.

We would like to thank the reviewer very much for providing this helpful comment. Our main contribution lies in demonstrating the methodological effectiveness of FD-SGIL, rather than optimizing latency or model size. The purpose of FD-SGIL is to provide an incremental detection framework that can continuously learn and adapt to newly emerging fraudulent website types while maintaining robust detection ability for previously seen ones. This is particularly important for real-world deployment, where fraudulent websites evolve rapidly and static detection models quickly become outdated.

Reviewer’s comment 8:​

Adversarial Robustness Fraudulent websites may use adversarial techniques (e.g., URL obfuscation, dynamic content) to evade detection. The paper does not mention robustness testing against adversarial evasion attacks.

We would like to thank the reviewer very much for providing this helpful comment. As stated in the Introduction, our main contribution lies in demonstrating the methodological effectiveness of FD-SGIL in achieving accurate detection of fraudulent websites and sustaining detection performance on newly emerging types, rather than explicitly optimizing adversarial robustness. We acknowledge that adversarial techniques such as URL obfuscation or dynamic webpage content can be used by attackers to evade detection. While defense against such adversarial strategies is not the primary focus of this study, our ablation experiments on the impact of multi-source fused features (see Table Ⅴ) show that FD-SGIL can still maintain strong detection performance even when certain feature sources (e.g., WHOIS) are unavailable. This suggests that the proposed framework inherently provides a degree of adaptability to adversarial variations. We consider adversarial robustness an important and complementary direction for future work.



The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

**Response to Comments from Reviewer #2 on the Introduction (Section I) and Related Work (Section II)**

Reviewer’s comment 1:

Overemphasis on Incremental Learning Without Addressing Core Challenges

While the paper highlights incremental learning (IL) as a solution for real-time detection, it does not sufficiently address: Class Imbalance: New fraudulent website types may have scarce labeled data, making IL unstable. Concept Drift: Fraudulent websites evolve dynamically (e.g., adversarial URL obfuscation), but the proposed "structure-growing" mechanism may not adapt quickly enough. Impact: The model might fail in real-world scenarios where new attack types emerge rapidly with limited labeled examples.

We would like to thank the reviewer very much for providing this helpful comment. In our work, we do not merely apply incremental learning, but specifically address its key challenges in fraudulent website detection. First, we introduce a dynamic structure-adjusting strategy, where new detection branches are expanded for unseen fraudulent website categories, while the parameters of previously learned branches are frozen to preserve their detection capability. Second, we combine experience replay with a transfer loss function to guide the newly expanded branches and reduce the risk of concept drift. Finally, by dynamically adjusting the loss weights of misclassified versus correctly classified samples according to the data distribution, our method effectively mitigates the issues of data imbalance and catastrophic forgetting. These designs ensure that FD-SGIL not only supports incremental detection but also overcomes the core obstacles to practical deployment.

Reviewer’s comment 2:

Heavy Reliance on BERT for Multi-Source Feature Fusion BERT is used for URL, webpage content, and WHOIS record fusion, but: Computational Cost: BERT is resource-intensive, conflicting with the goal of lightweight real-time detection on IoT devices. Feature Alignment: No clear method is described for aligning heterogeneous features (e.g., textual content vs. structured WHOIS data), which could lead to suboptimal fusion. Impact: High latency and energy consumption, making deployment on edge devices impractical.

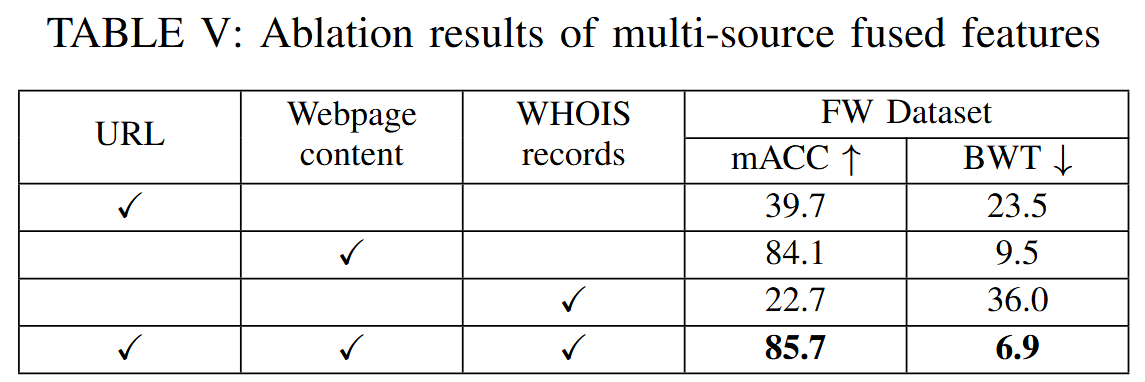
We would like to thank the reviewer very much for providing this helpful comment. The main contribution of this work is to demonstrate the methodological effectiveness of FD-SGIL in enabling continuous detection of newly emerging fraudulent website types while retaining robust detection ability for previously seen ones, rather than focusing on optimizing latency or model size. In our multi-source feature fusion design, BERT is employed to represent URL and webpage content, while WHOIS records are encoded through a lightweight feature mapping, and the three representations are concatenated for fusion. This design provides strong feature representations with only minor computational overhead compared to baseline methods and does not significantly slow down real-time detection.

Reviewer’s comment 3:

Lack of Adversarial Robustness Considerations Fraudulent websites actively evade detection via:

Dynamic Content: AIGC-generated text or adversarial perturbations in URLs/webpages. WHOIS Spoofing: Fake registration data to bypass WHOIS-based checks. Impact: The paper does not evaluate robustness against adversarial evasion techniques, risking high false negatives in practice.

We would like to thank the reviewer very much for this helpful comment. As stated in the Introduction, our main contribution lies in demonstrating the methodological effectiveness of FD-SGIL in achieving accurate detection of fraudulent websites and sustaining detection performance on newly emerging types, rather than explicitly optimizing adversarial robustness. We acknowledge that adversarial techniques such as URL obfuscation or dynamic webpage content can be leveraged by attackers to evade detection. While explicit defenses against such adversarial strategies are not the primary focus of this work, our ablation experiments on the impact of multi-source fused features (see Table Ⅴ) show that FD-SGIL can still maintain strong detection performance even when certain feature sources (e.g., WHOIS) are unavailable. This indicates that the framework inherently provides a degree of adaptability to adversarial variations. We consider adversarial robustness an important and complementary research direction for future work.

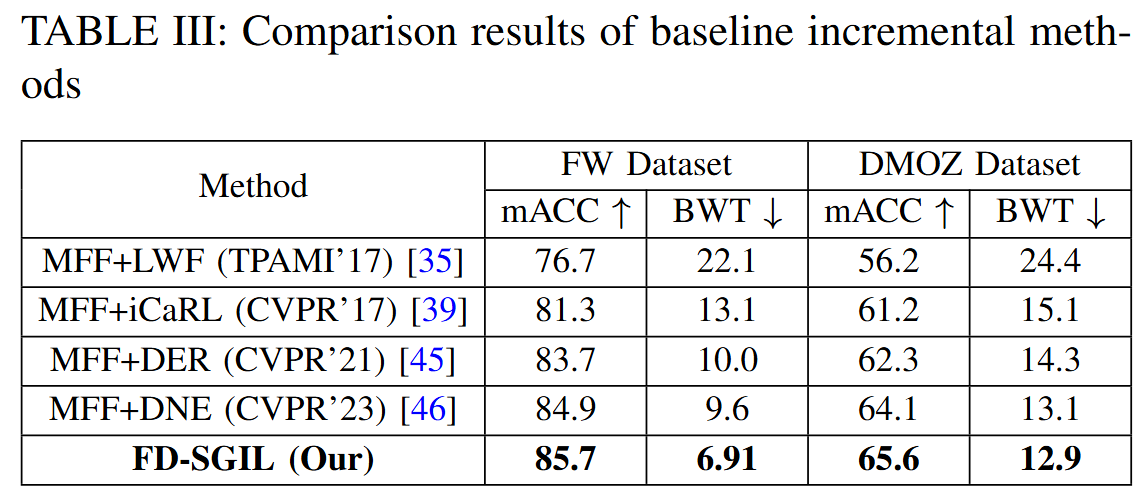


The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

Reviewer’s comment 4:

Shallow Treatment of Multi-Class Detection The paper claims superior multi-class performance but: No Discussion of Class Overlap: Many fraudulent types (e.g., phishing vs. scam) share similar features, risking misclassification. Limited Dataset Diversity: Only three datasets are used, likely lacking coverage of rare or emerging fraud types. Impact: Overstated generalization claims; the model may struggle with fine-grained classification.

We would like to thank the reviewer very much for this helpful comment. We acknowledge that distinguishing between multiple fraudulent website types can be challenging, especially when certain categories share overlapping features or when new types differ significantly from previously seen ones. Our proposed FD-SGIL framework addresses this by combining a dynamically weighted transfer loss with experience replay, which helps preserve past knowledge while learning new classes. Furthermore, we constructed the FW dataset containing seven of the most harmful fraud categories and conducted extensive experiments. As shown in Table Ⅲ, FD-SGIL consistently achieves superior overall performance, maintaining stable mACC values across incremental tasks, which demonstrates its effectiveness in handling diverse and evolving fraud types. For newly emerging categories not covered in this study, FD-SGIL can readily incorporate them through incremental training in future work.

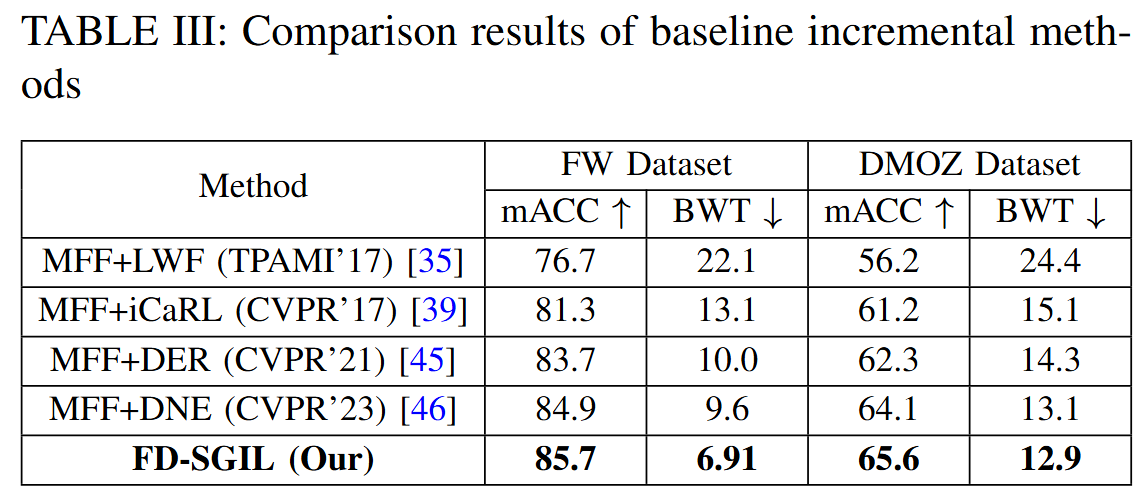


The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reviewer’s comment 5:

Incremental Learning Limitations Underplayed The proposed "transfer loss" and "dynamic weight adjustment" mechanisms are not compared rigorously to state-of-the-art (SOTA) IL methods like: Replay-Based Methods: Experience replay (e.g., [30]) is more proven but may be omitted due to memory constraints. Meta-Learning: Could better handle few-shot new classes but is not explored. Impact: Catastrophic forgetting may still occur for long-tailed or dissimilar new classes.

We would like to thank the reviewer very much for this helpful comment. In subsection **C. Incremental Learning Methods**, we categorize existing approaches into three families: regularization based, experience replay based, and dynamic structure adjustment based methods [34]. To ensure a fair evaluation, we compared FD-SGIL against representative baselines from each category, namely LWF [35] (regularization based), iCaRL [39] (experience replay based), and DER [45], DNE [46] (dynamic structure adjustment based). As shown in Table Ⅲ, FD-SGIL consistently achieves the best overall performance in terms of both mACC and BWT, demonstrating its ability to effectively balance stability and plasticity in incremental phishing detection. We acknowledge that meta-learning approaches are often applied to few-shot scenarios; however, the focus of this work is on long-term incremental detection of evolving fraudulent website categories, where our structure-growing design combined with transfer loss and experience replay is more suitable.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reference:

[34] G. M. Van de Ven, T. Tuytelaars, and A. S. Tolias, “Three types of incremental learning,” Nature Machine Intelligence, vol. 4, no. 12, pp.1185–1197, 2022.

[35] Z. Li and D. Hoiem, “Learning without forgetting,” IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 2935–2947, 2017.

[39] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 2001–2010.

[45] S. Yan, J. Xie, and X. He, “Der: Dynamically expandable representation for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3014–3023.

[46] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp.11 858–11 867.

Reviewer’s comment 6:

Ignoring Real-World Deployment Constraints Critical IoT deployment challenges are overlooked:

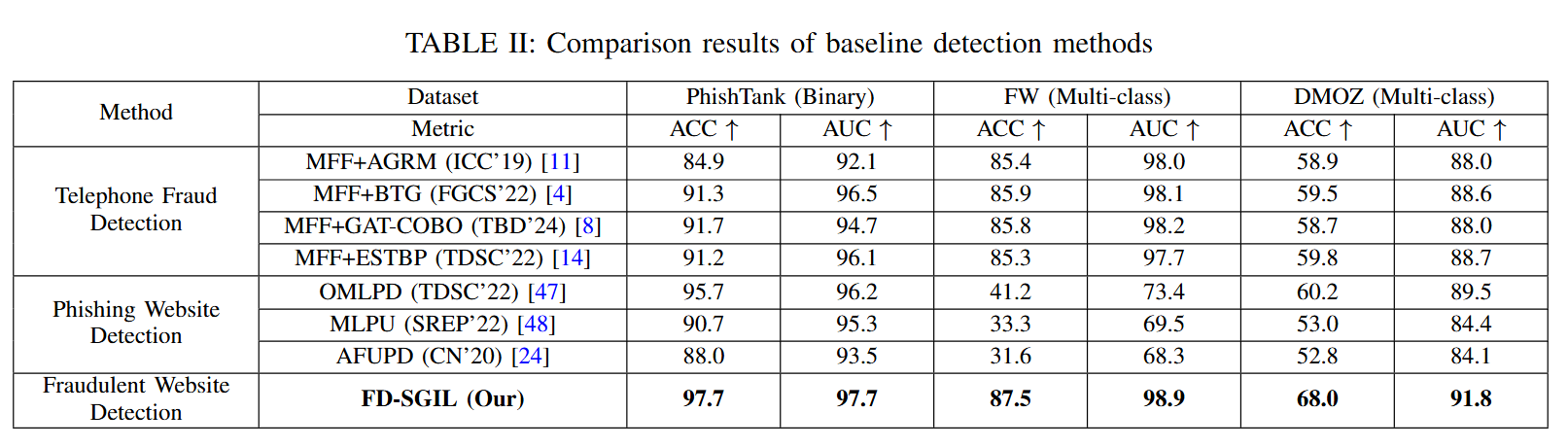
Latency Requirements: No metrics on inference speed for real-time detection. Data Privacy: WHOIS records may contain sensitive user data; no privacy-preserving feature extraction is proposed. Impact: The framework may be unusable in privacy-sensitive or low-latency scenarios.

We would like to thank the reviewer very much for this helpful comment. The main contribution of this work is to demonstrate the methodological effectiveness of FD-SGIL in enabling continuous detection of newly emerging fraudulent website types while retaining robust detection ability for previously seen ones, rather than explicitly focusing on optimizing latency or addressing privacy constraints. In our multi-source feature fusion design, URL and webpage content are represented by BERT, while WHOIS records are encoded through a lightweight feature mapping, and the three representations are concatenated for fusion. This provides strong feature representations with only minor computational overhead compared to baseline methods and does not significantly slow down real-time detection. We acknowledge that challenges such as strict latency requirements in IoT environments and privacy concerns in handling WHOIS data are important for practical deployment, and we regard them as valuable and complementary directions for future work.

Reviewer’s comment 7:

Weak Baseline Comparison in Related Work The literature review (Section II) critiques non-ML and ML methods but: Superficial Analysis of Deep Learning Gaps: Fails to acknowledge hybrid approaches (e.g., GNNs for URL graphs) that could outperform BERT. No Mention of Zero-Shot Learning: For entirely new fraud types, zero-shot detection might be more viable than IL. Impact: The proposed method’s novelty is overstated relative to unexplored alternatives.

We would like to thank the reviewer very much for this helpful comment. In the revised manuscript, we have reorganized the literature review to cover the latest studies related to telephone fraud detection, phishing website detection, and incremental learning methods. In particular, we acknowledge the importance of graph-based approaches such as AGRM [11], BTG [4], GAT-COBO [8] and ESTBP [14], which have been applied to multi-source fusion for fraud detection. As shown in Table Ⅱ, FD-SGIL outperforms them by achieving the highest ACC and AUC values, confirming the effectiveness of our proposed framework.



Regarding the suggestion on zero-shot learning, while this paradigm is promising for handling entirely unseen categories, its performance often relies on strong semantic priors or auxiliary information, which may not always be available for emerging fraudulent website categories. In contrast, our work focuses on incremental learning, which allows the model to continuously adapt to newly observed categories of fraudulent websites while preserving knowledge of previously learned ones. This setting better reflects the real-world fraud detection scenario, where labeled data of new categories gradually becomes available and must be integrated into the detection system without retraining from scratch.

The above modified content is placed in section **Ⅱ. RELATED WORK of Pages 2~4** and subsection **B. Comparison to Baseline Detection Methods of Page 9** in the revised version.

Reference:

[4] X. Hu, H. Chen, S. Liu, H. Jiang, G. Chu, and R. Li, “Btg: A bridge to graph machine learning in telecommunications fraud detection,” Future Generation Computer Systems, vol. 137, pp. 274–287, 2022.

[8] X. Hu, H. Chen, J. Zhang, H. Chen, S. Liu, X. Li, Y. Wang, and X. Xue, “Gat-cobo: Cost-sensitive graph neural network for telecom fraud detection,” IEEE Transactions on Big Data, vol. 10, no. 4, pp. 528–542, 2024.

[11] M. Liu, J. Liao, J. Wang, and Q. Qi, “Agrm: Attention-based graph representation model for telecom fraud detection,” in ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019, pp. 1–6.

[14] G. Chu, J. Wang, Q. Qi, H. Sun, S. Tao, H. Yang, J. Liao, and Z. Han, “Exploiting spatial-temporal behavior patterns for fraud detection in telecom networks,” IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 6, pp. 4564–4577, 2022.

Reviewer’s comment 8:

Ethical and Legal Risks The paper does not discuss: Bias in WHOIS Data: Regional biases (e.g., flagging sites from "weak oversight" countries) could lead to discriminatory false positives. Misuse Potential: The model could be weaponized to flag legitimate sites as fraudulent. Impact: Unintended harm to businesses or individuals in developing regions.

We would like to thank the reviewer very much for this insightful comment. First, regarding potential bias in WHOIS data, we note that all WHOIS information used in our study was parsed directly from real-world fraudulent websites. While false positives may occasionally occur, this setup allows us to better capture realistic fraud scenarios and ensures that the evaluation reflects practical challenges faced by real-world detection systems.

Second, we would like to reiterate that the main contribution of this work lies in demonstrating the methodological effectiveness of FD-SGIL in achieving accurate and continual detection of fraudulent websites, rather than explicitly addressing robustness against WHOIS spoofing or other adversarial manipulations. Nevertheless, we have investigated the effect of different feature combinations in subsubsection **2) Impact of Multi-source Fused Features**. As shown in Table Ⅴ, FD-SGIL can still maintain strong detection performance even when certain feature sources (e.g., WHOIS) are unavailable, indicating that the framework inherently provides a degree of adaptability to adversarial variations.

Finally, we agree that ethical and legal considerations, including potential bias and misuse, are critical for future deployment. We regard adversarial robustness and bias mitigation as important and complementary research directions for extending the present work toward trustworthy real-world applications.

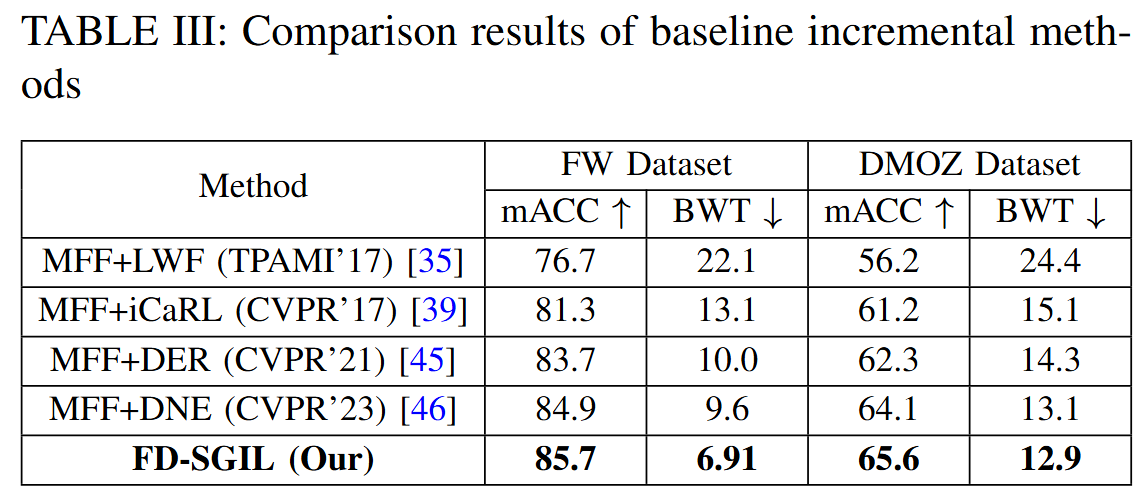
**Response to Comments from Reviewer #2 on the Introduction (Section VI)**

Reviewer’s comment 1:

Overstated Claims Without Caveats

Issue: The conclusion claims FD-SGIL "excels" in multi-class detection and catastrophic forgetting but omits: Limitations in the experiments: No discussion of failure cases (e.g., specific fraud types where performance drops). Real-world constraints: Computational costs of BERT or latency on edge devices are ignored.

We would like to thank the reviewer very much for this insightful comment. To comprehensively evaluate model performance and capture potential weaknesses, we report two widely used metrics in incremental learning: mean accuracy (mACC), which measures the average detection performance across all tasks, and backward transfer (BWT), which quantifies the effect of learning new tasks on the retention of previously learned ones. As shown in Table Ⅲ, FD-SGIL achieves the best results on both metrics compared with all baselines, indicating its effectiveness in balancing detection of emerging fraud types with preservation of prior knowledge. Regarding computational cost, while URL and webpage content are encoded with BERT, WHOIS features are processed by a lightweight mapping module, and the fused representation introduces only minor overhead relative to baseline methods. We acknowledge that challenges such as strict latency requirements in IoT applications and privacy concerns in handling WHOIS data are important for practical deployment, and we consider them valuable directions for future research.



Reviewer’s comment 2:

Superficial Future Work

Issue: Future directions are generic (e.g., "more superior structures") and lack specificity. Example: No mention of adversarial robustness (e.g., testing against AIGC-generated phishing sites) or privacy-aware training (critical for WHOIS data).

We would like to thank the reviewer very much for this insightful comment. We acknowledge that the discussion of future research directions in the original submission was somewhat general. In the revised manuscript, we have made this section more concrete. Specifically, we highlight the importance of adversarial robustness by noting that emerging threats such as AIGC-generated phishing sites, voice cloning, and deepfake-based scams may introduce new attack vectors that conventional features (e.g., URLs and WHOIS) cannot fully address. We also emphasize the necessity of privacy-aware training strategies when leveraging sensitive data sources such as WHOIS records. Furthermore, we outline that future work should investigate novel multimodal indicators (e.g., deepfake traces in media, logical inconsistencies in user interactions) and explore the potential of large language models (LLMs) for fraudulent website detection. These directions complement the methodological contributions of FD-SGIL and provide a concrete roadmap toward more robust, adaptive, and privacy-conscious detection frameworks.

The above modified content is placed in the section **Ⅵ. CONCLUSION of Page 12** in the revised version.

Reviewer’s comment 3:

Ignored Ethical and Deployment Challenges

Issue: No reflection on: Bias risks: WHOIS-based features may disproportionately flag sites from certain regions. Scalability: Deploying BERT-based models on IoT devices remains impractical.

We would like to thank the reviewer very much for this insightful comment. Regarding potential bias in WHOIS data, all information used in this study was parsed directly from real-world fraudulent websites. While false positives may occur, this setup allows us to better capture realistic fraud scenarios and ensures that evaluation reflects practical challenges. In terms of deployment, we acknowledge that BERT is relatively resource-intensive and may not be optimal for latency-sensitive IoT environments. However, our primary objective in this work is to validate the methodological effectiveness of FD-SGIL for continual fraudulent website detection, for which BERT provides rich semantic representations. We agree that bias mitigation, privacy-aware training, and deployment efficiency are critical for real-world applications, and we regard these as valuable and complementary directions for extending this research.

Reviewer’s comment 4:

4. No Link to Broader Impact

Issue: Fails to contextualize how FD-SGIL advances cybersecurity beyond academia. Example: Could highlight potential integration with browser extensions or ISP-level detection systems.

We would like to thank the reviewer very much for this insightful comment. We agree that it is important to highlight the broader impact of FD-SGIL beyond the academic setting. The proposed framework is designed to continuously adapt to newly emerging fraudulent website types while preserving detection ability for known ones, which makes it naturally suited for real-world deployment scenarios. For example, FD-SGIL could be integrated into browser extensions to provide users with real-time protection against malicious websites, or incorporated into ISP-level monitoring systems to block large-scale fraud attempts at the network layer. In addition, the multi-source fusion design allows flexible adaptation to the heterogeneous data available in practical environments, such as URL blacklists, DNS logs, and WHOIS records. We also emphasize that fraudulent website governance requires a multidisciplinary effort that combines technical detection, legal enforcement, and public education. In this context, FD-SGIL contributes to the technical foundation of trustworthy cybersecurity solutions and offers a pathway for scalable integration into practical defense systems.

Reviewer’s comment 5:

Incremental Learning Gaps Unaddressed

Issue: While catastrophic forgetting is mentioned, the conclusion ignores: Stability-Plasticity Dilemma: Trade-offs between adapting to new classes and retaining old knowledge. Few-Shot Scenarios: Handling new fraud types with minimal labeled examples.

We would like to thank the reviewer very much for this helpful comment. First, FD-SGIL addresses the stability–plasticity trade-off through a dynamic structure adjustment strategy. Specifically, when new fraudulent website categories emerge, FD-SGIL grows new detection branches dedicated to these classes while freezing parameters of previously learned branches. This design allows the model to adapt to new fraud types (plasticity) while preserving knowledge of earlier ones (stability). We acknowledge that this strategy increases the number of model parameters as the structure grows, but this trade-off is essential for achieving precise detection of previously unseen fraudulent websites without sacrificing performance on known ones.

Second, regarding few-shot scenarios, we note that FD-SGIL is designed primarily for long-term incremental detection rather than few-shot learning. Our framework combines a structure growing design with transfer loss and experience replay, making it more suitable for evolving fraudulent website categories in practical long-term deployments. Finally, we would like to once again thank **Reviewer #2** for the constructive feedback.