# Multi-classification of Linguistic Steganography Driven by Large Language Models

Jie Wang, Yihao Wang, Yihao Li, Ru Zhang*, Jianyi Liu

***Abstract*—In linguistic steganalysis (LS), the fundamental requirement is to detect steganographic text (stego) effectively, and existing LS methods have shown excellent detection performance even in complex scenarios. However, different steganography employs various grammatical rules and syntactic structures, which result in distinct text representations. This diversity complicates the identification of stego in mixed datasets. A major challenge in LS is pinpointing the specific steganography used, which is crucial for developing effective extraction algorithms. Thus, this paper proposes a multi-classification method based on the Large Language Models (LLMs) called LSMC. This approach utilizes the strengths of LLMs in semantic understanding and contextual analysis, allowing for accurate classification. Experimental results show that the LSMC method can efficiently perform multi-classification, precisely discerning the steganography used in mixed datasets with up to 20 categories. This provides a feasible solution for fully deciphering steganography in subsequent stages.**

***Index Terms*—Linguistic steganalysis, Large Language Models, Multi-Classification, Steganography identification.**

## I. Introduction

In the current digital era, network attacks and data leakage incidents occur frequently, making it essential to securely transmit information carriers, including images, text, and audio[1], [2]. Traditional encryption techniques are susceptible to active attacks due to the distinctive characteristics of their ciphertexts[3]. Steganography, as a technique in hidden systems, embeds secret information into the redundant parts of multimedia, creating media that appear normal but contain hidden information[4], [5]. Only authorized individuals can detect and correctly extract secret information, preventing illegal listeners from intercepting and decrypting it. However, steganography can also cause unpredictable and severe consequences if it is abused by wrongdoers. Thus, in order to effectively counter steganography, steganalysis methods have received widespread attention[6], [7], [8]. The lossless transmission of text on social platforms has made it a central focus of steganography research [9], [10]. LS detects hidden information by capturing the statistical properties of the cover and stego texts, such as semantic, word dependency, syntactic, and other content-related features. Existing LS methods can be broadly categorized into manually crafted and automatic extraction. The former techniques typically rely on constructing specific features, such as word frequency [11], [12] and word association [13], to assist in identifying and interpreting specific steganography schemes. These methods perform well in certain scenarios. The latter design deep learning models to capture high-dimensional features, enhancing the detection of steganographic embedding disturbances and generally achieving superior detection performance [14]. In the early stages, Yang et al. [15] and Zou et al. [16] designed various sequence and convolutional models, significantly improving the detection performance of stego. Recently, scholars have shifted their focus from designing model architectures to addressing complex scenarios. For instance, Wang et al. [17] and Wang et al. [18] employed techniques like self-training, and user profiling to enhance detection accuracy in few-shot scenarios. Moreover, with the development of LLMs, their powerful feature extraction capabilities offer new possibilities for LS. Bai et al. [19] reframed LS as a generative task, improving detection performance, while Tang et al. [20] proposed both generative and classification-based LS models, demonstrating outstanding performance in challenging scenarios. Although significant progress has been made in LS, existing methods mainly focus on content detection. They are failing to comprehensively classify different steganography. In real-world applications, such as social networks, steganographers often use various steganography schemes to transmit secret information. These schemes effectively mimic the semantic characteristics of natural text, generating high-quality stego. Thus, it poses challenges for LS in fully deciphering hidden information. Such composite texts exhibit strong perceptual and semantic concealment, resulting in highly similar distributions, as shown in Fig. 1.

However, in such complex scenarios, the stego generated

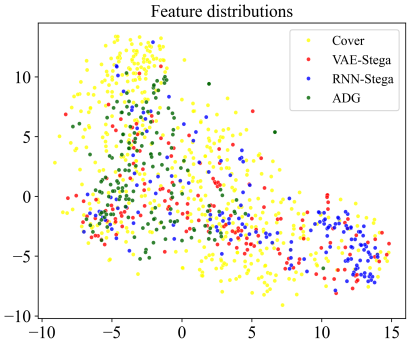


Fig. 1. Distribution of mixed stego features (t-SNE [21] visualization of steganalysis features extracted by TS_BiRNN [15])

This work was supported in part by the Natural Science Foundation of China under Grant U21B2020.

Jie Wang, Yihao Wang, Yihao Li, Ru Zhang, and Jianyi Liu are with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China. (The corresponding author is Ru Zhang.) (Email: WangJie2023@bupt.edu.cn, yh-wang@bupt.edu.cn, yihaoli@bupt.edu.cn, zhangru@bupt.edu.cn, and liujy@bupt.edu.cn)

by different algorithms still exhibit distinct features. Thus, the fundamentals of the LS method can be approximated for multi-classification tasks. Nevertheless, due to current technological limitations, existing LS methods perform poorly in multi-classification tasks.

In fact, extensive research has been devoted to classifying different security algorithms, such as image steganography and traditional cryptographic algorithms in the field of information security [22], [23]. These methods are typically optimized for specific algorithms or techniques, but they still face challenges in addressing the multi-classification problem of LS. Text steganography embeds patterns into the semantics and structure of natural language. This makes it hard for traditional methods to accurately identify and distinguish it. To address this issue, we propose a multi-classification method based on LLMs, termed LSMC. This approach comprises two key components: a feature extraction module and a feature mapping module. In the feature extraction module, this method primarily relies on the deep processing and understanding capabilities of LLMs in terms of text semantics and context. We fine-tune LLMs using a dataset composed of various steganography and extract feature vectors from the final hidden layer to capture distinctions between diverse algorithms. The feature mapping module then processes the feature vectors through multi-classifiers to achieve precise classification of steganography. Experimental results show that the LSMC method significantly improves multi-classification performance in scenarios involving hybrid algorithms.

## II. Proposed Method

In this paper, we propose a multi-classification approach for scenarios involving multiple steganography, the main framework is shown in Fig. 2. In this section, we will describe the LSMC method in detail.

### A. Feature Extraction Module

To achieve accurate classification in complex multiple steganography scenarios, we adopt LLMs due to their strong capabilities in semantic understanding and contextual analysis. In our classification model, the input is the complex text to be analyzed. These texts $X_i$ undergo word embedding, converting discrete textual data into continuous vector representations $V_i$. The above describes the formula as Eq. 1. The $V_i$ are then fed into pre-trained LLMs, where the model processes the text deeply through its multi-layer neural network. Particularly, we extract the output $H^L$ from the last hidden layer $L$ as the feature representation of the text. This output captures deep semantic information,which can be expressed as Eq. 2 :

$$\mathbf{V} = \{\mathrm{E}(x_1), \mathrm{E}(x_2), \ldots, \mathrm{E}(x_n)\} \tag{1}$$

$$\mathbf{H}^L = \mathrm{Trm}_{\text{model}}(\mathbf{V}) \tag{2}$$

where, $\mathbf{E}$ performs word embedding for each $x_i$ the and $\mathbf{V} = \{v_1, v_2, \ldots, v_n\}$ is the embedding of the input data.

To improve training efficiency, we employed the LoRA (Low-Rank Adaptation) technique [24] to fine-tune this model. The formula is shown as follows:

$$\mathbf{W}_0 + \Delta\mathbf{W} = \mathbf{W}_0 + BA, B \in \mathbb{R}^{d\times r}, A \in \mathbb{R}^{r\times k} \tag{3}$$

where, $r \ll \min(d, k)$. Then the parameters of the original model and the fine-tuning parameters obtained by LoRA are merged to construct a fine-tuned model that can be adapted to advanced steganalysis tasks.

### B. Feature Mapping Module

We have designed a feature mapping module specifically for feature transformation in multi-classification tasks. The core function is to efficiently map the extracted features $f_{LM}(t)$ so that it can be input into the classifier. Here, a multi-classifier is required, where the extracted textual features are processed through a fully connected layer to ensure that these features can be effectively distinguished in a high-dimensional space. Then, the processed features are fed into a softmax layer for probability normalization, thereby achieving multi-class output. The formula is as follows:

$$\begin{aligned}\text{softmax}\left(W_{fc} f_{\text{final}}(t) + b_{fc}\right) = [&p(S_1|f_{\text{final}}(t)), p(S_2|f_{\text{final}}(t)),\\ &\ldots, p(S_m|f_{\text{final}}(t))]\end{aligned}, \tag{4}$$

where, $W_{fc}$ and $b_{fc}$ are the weight matrix and bias vector of the fully connected layer, respectively, and $p(S_i|f_{\text{final}}(t))$

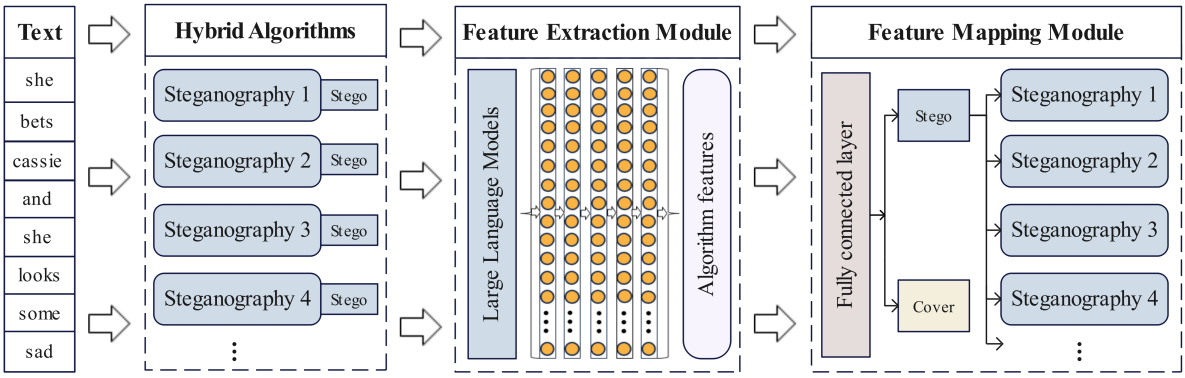


Fig. 2. The overall framework of the LSMC method.

TABLE I
THE COMPARISON OF ALGORITHMS (FIXED BPW) FOR MULTI-CLASSIFICATION DETECTION, **BOLD** VALUES REPRESENT THE BEST RESULTS.

| Methods | | TS_BiRNN [15] | | R_BiLSTM_C [29] | | SSLS [30] | | Zou [16] | | GS-Llama [19] | | **Ours (LSMC)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | Algorithms | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| Movies | V+R | 68.54 | 65.17 | 67.08 | 60.70 | 88.02 | 87.61 | 91.67 | 91.62 | 90.60 | 90.64 | **94.68** | **94.64** |
| | V+R+A | 60.83 | 53.47 | 62.29 | 50.75 | 69.37 | 65.52 | 77.81 | 76.31 | 76.97 | 76.10 | **82.81** | **82.65** |
| | V+R+A+T | 63.75 | 58.80 | 66.46 | 56.92 | 76.56 | 74.84 | 78.85 | 78.02 | 75.66 | 75.10 | **85.72** | **85.51** |
| Twitter | V+R | 68.75 | 64.17 | 69.27 | 59.90 | 81.87 | 81.51 | 84.79 | 85.02 | 83.02 | 81.46 | **89.37** | **89.36** |
| | V+R+A | 63.44 | 54.23 | 63.33 | 51.24 | 67.60 | 65.43 | 68.23 | 67.15 | 66.37 | 64.89 | **69.37** | **69.73** |
| | V+R+A+T | 56.56 | 45.80 | 58.02 | 47.10 | 59.58 | 57.60 | 63.65 | 57.45 | 60.52 | 51.85 | **65.20** | **64.23** |
| News | V+R | 67.40 | 65.99 | 70.94 | 62.30 | 93.95 | 93.94 | 94.69 | 94.68 | 89.47 | 89.30 | **96.35** | **96.34** |
| | V+R+A | 69.79 | 63.52 | 69.38 | 64.90 | 92.91 | 92.89 | 92.50 | 92.41 | 90.20 | 90.43 | **94.37** | **94.31** |
| | V+R+A+T | 62.19 | 59.56 | 64.58 | 57.43 | 76.66 | 75.35 | 81.67 | 81.07 | 78.12 | 77.74 | **83.12** | **83.26** |

represents the probability that the text was generated by the $i$-th steganographic algorithm.

## III. EXPERIMENTS

To ensure the comprehensiveness and reliability of the comparison, all experiments are run on the NVIDIA GeForce RTX 3090 GPUs.

### A. Settings

***Dataset:*** To fully validate the effectiveness of the LSMC method, we construct six hybrid datasets. The cover comes from three classic text types: Twitter, Movies, and News. The stego are generated by the VAE-Stega (V) [25], RNN-Stega (R) [26], ADG (A) [27], and Tina-Fang (T) [28] algorithms. The ratio of cover to stego is set at 1:1, with a total of 4800 texts. We randomly divided these into training, validation, and test sets with a ratio of 6:2:2. ***Baselines:*** To comprehensively evaluate the performance of the LSMC method, we selected five baseline models for comparison. These baselines include: Non-BERT-based (TS_BiRNN [15] and R_BiLSTM_C [29]), BERT-based (SSLS [30] and Zou [16]), and LLMs-based (GS-Llama [19]).

***Hyperparameters:*** In the proposed method, the language model is ChatGLM-6B [31]. The experimental parameters are set as follows: batch size = 5, LoRA rank = 64, LoRA alpha = 128, epochs = 4, and the learning algorithm used is AdamW [32] with an initial learning rate of 2e-4. In the ablation experiments, we added two other models namely ChatGLM3-6B [33] and Llama2-7B [34] for exploring the effect of different models on the classification effect.

***Evaluation Metrics:*** During the evaluation process, we use detection accuracy (Acc) and F1 score, which are common metrics in classification tasks, as the evaluation criteria.

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{5}$$

$$\mathrm{F1} = 2 \times (\mathrm{P} \times \mathrm{R})/(\mathrm{P} + \mathrm{R}), \tag{6}$$

where, TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative. P and R represent the Precision and Recall of detection.

### B. Multi-classification and binary classification performance of hybrid algorithms with a fixed bpw comparison

Here, we perform binary and multi-classification on three types of mixed steganographic datasets and compare the detection performance with other baselines. The results are shown in Table I and Table II.

As shown in Tables, the method significantly outperforms existing baseline methods on different datasets for both LS and multi-classification tasks. In addition, Fig. 3 visually demonstrates the performance of the LSMC method in extracting text features generated by different steganography. These results show that the method achieves better LS and multi-classification performance in complex scenarios.

### C. Multi-classification performance of hybrid algorithms with mixed bpw comparison

In this dataset, each algorithm is represented with five different embedding rates. It is important to note that ADG is an adaptive generation algorithm, so we configured its parameters to produce text data with five classes bpw. The variation in embedding rates can indirectly lead to differences in steganography, resulting in a complex dataset containing up to 20 algorithms. We performed multi-classification tests on these datasets and compared the performance with baseline methods, as shown in Table III.

Our method still consistently outperforms other approaches in terms of detection accuracy and F1 score, effectively achieving basic classification across multiple algorithms.

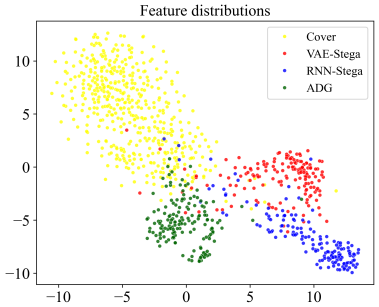


Fig. 3. Distribution of datasets features (t-SNE visualization of steganalysis features extracted by LSMC)

TABLE II
THE COMPARISON OF ALGORITHMS (FIXED BPW) FOR LS DETECTION, **BOLD** VALUES REPRESENT THE BEST RESULTS.

| Methods | | TS_BiRNN [15] | | R_BiLSTM_C [29] | | SSLS [30] | | Zou [16] | | GS-Llama [19] | | **Ours (LSMC)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | Algorithms | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| Movies | V+R | 70.00 | 68.70 | 71.67 | 70.56 | 92.18 | 92.18 | 94.69 | 94.59 | 93.54 | 93.53 | **96.87** | **96.89** |
| | V+R+A | 61.15 | 61.35 | 59.68 | 57.46 | 85.72 | 85.66 | 87.08 | 86.78 | 81.66 | 81.64 | **92.39** | **92.30** |
| | V+R+A+T | 64.58 | 63.91 | 64.79 | 62.19 | 85.20 | 85.13 | 87.92 | 87.74 | 83.43 | 83.32 | **93.43** | **93.38** |
| Twitter | V+R | 90.31 | 90.34 | 90.21 | 90.37 | 96.87 | 96.87 | 97.81 | 97.81 | 96.81 | 96.80 | **98.64** | **98.65** |
| | V+R+A | 60.42 | 59.75 | 61.88 | 53.44 | 73.85 | 73.79 | 77.19 | 77.90 | 72.70 | 72.13 | **82.81** | **82.83** |
| | V+R+A+T | 66.04 | 65.83 | 71.35 | 65.50 | 76.25 | 75.48 | 81.77 | 81.40 | 74.16 | 72.47 | **86.35** | **86.07** |
| News | V+R | 66.77 | 65.21 | 67.71 | 58.11 | 94.89 | 94.88 | 94.46 | 96.47 | 92.29 | 92.29 | **98.33** | **98.34** |
| | V+R+A | 67.08 | 64.09 | 70.31 | 65.12 | 94.27 | 94.27 | 96.15 | 96.19 | 95.00 | 94.98 | **98.02** | **98.03** |
| | V+R+A+T | 71.04 | 70.98 | 73.65 | 73.56 | 92.29 | 92.27 | 95.21 | 95.19 | 94.72 | 94.70 | **96.97** | **97.00** |

TABLE III
THE COMPARISON OF ALGORITHMS (MIXED BPW) FOR MULTI-CLASSIFICATION DETECTION,**BOLD** VALUES REPRESENT THE BEST RESULTS.

| Methods | | TS_BiRNN [15] | | R_BiLSTM_C [29] | | SSLS [30] | | Zou [16] | | GS-Llama [19] | | **Ours (LSMC)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | Algorithms | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| Movies | V+R | 60.54 | 60.55 | 60.21 | 60.14 | 68.47 | 68.76 | 67.81 | 68.60 | 66.54 | 64.02 | **72.29** | **72.38** |
| | V+R+A | 58.90 | 57.61 | 58.38 | 49.36 | 66.35 | 63.51 | 69.17 | 64.75 | 67.55 | 63.18 | **72.64** | **72.12** |
| | V+R+A+T | 54.62 | 48.24 | 55.15 | 48.14 | 64.68 | 63.40 | 65.42 | 64.06 | 64.02 | 64.10 | **70.20** | **69.74** |
| Twitter | V+R | 61.33 | 61.55 | 61.17 | 60.93 | 67.39 | 67.12 | 68.75 | 68.55 | 66.98 | 67.36 | **70.41** | **70.46** |
| | V+R+A | 55.98 | 49.04 | 54.10 | 53.11 | 62.70 | 61.59 | 63.54 | 63.14 | 61.89 | 62.77 | **69.27** | **69.79** |
| | V+R+A+T | 54.31 | 52.03 | 53.17 | 52.48 | 59.87 | 58.61 | 61.56 | 60.44 | 60.13 | 60.11 | **66.56** | **56.63** |
| News | V+R | 63.04 | 61.23 | 63.31 | 62.95 | 73.54 | 73.49 | 74.69 | 73.87 | 73.15 | 73.54 | **76.04** | **75.79** |
| | V+R+A | 57.85 | 53.52 | 57.85 | 57.51 | 71.14 | 68.07 | 74.90 | 74.60 | 73.58 | 72.91 | **79.27** | **78.76** |
| | V+R+A+T | 56.17 | 52.92 | 54.10 | 52.80 | 68.02 | 68.22 | 67.60 | 66.64 | 67.55 | 66.37 | **75.62** | **75.03** |

### *D. Ablation experiments*

To better demonstrate the generality of LSMC in multi-classification tasks, we included experimental tests using the ChatGLM3-6B[33] and LLaMA2-7B[34] models in this section. The results are shown in Table IV and Table V. The three LLMs exhibit similar performance in classification, with minimal differences. These minor differences are likely due to the impact of random seeds on loss reduction in deep learning algorithms. This result suggests that even under different model architectures, the classification performance remains highly accurate, highlighting the robustness and broad applicability of LLMs in this task.

TABLE IV
ABLATION EXPERIMENTS WITH DIFFERENT MODELS FOR DATASETS WITH A FIXED BPW

| Methods | | ChatGLM3-6B [33] | | LLaMA2-7B [34] | |
|---|---|---|---|---|---|
| Dataset | Algorithms | Acc | F1 | Acc | F1 |
| Movies | V+R | 93.22 | 93.16 | 94.37 | 94.32 |
| | V+R+A | 83.54 | 83.20 | 82.91 | 82.80 |
| | V+R+A+T | 85.41 | 85.08 | 83.54 | 83.02 |
| Twitter | V+R | 89.27 | 89.28 | 89.89 | 89.99 |
| | V+R+A | 72.29 | 72.00 | 75.52 | 74.47 |
| | V+R+A+T | 67.91 | 66.44 | 67.29 | 65.39 |
| News | V+R | 96.87 | 96.87 | 97.39 | 97.39 |
| | V+R+A | 95.31 | 95.29 | 94.68 | 94.63 |
| | V+R+A+T | 85.00 | 84.98 | 83.33 | 83.11 |

TABLE V
ABLATION EXPERIMENTS WITH DIFFERENT MODELS FOR DATASETS WITH THE MIXED BPW.

| Methods | | ChatGLM3-6B [33] | | LLaMA2-7B [34] | |
|---|---|---|---|---|---|
| Dataset | Algorithms | Acc | F1 | Acc | F1 |
| Movies | V+R | 73.12 | 73.02 | 73.29 | 73.26 |
| | V+R+A | 75.10 | 75.43 | 75.93 | 75.39 |
| | V+R+A+T | 71.14 | 71.14 | 71.97 | 71.95 |
| Twitter | V+R | 72.81 | 72.73 | 70.93 | 69.67 |
| | V+R+A | 70.10 | 69.66 | 72.29 | 71.87 |
| | V+R+A+T | 66.56 | 66.41 | 68.22 | 67.09 |
| News | V+R | 76.77 | 76.62 | 77.18 | 76.78 |
| | V+R+A | 79.58 | 79.44 | 80.41 | 80.23 |
| | V+R+A+T | 78.02 | 77.80 | 75.83 | 75.63 |

## IV. CONCLUSION

Current LS methods focus on detecting the presence of stego, but lack the ability to classify steganography in complex scenes. Thus, this paper proposes a multi-classification method, called LSMC. This method uses LLMs to extract their hidden layer vectors containing rich textual features, which are subsequently mapped to a multi-classification module to achieve precise classification. Experimental results show that in complex scenarios, the LSMC method improves about 2%-18% in multi-classification accuracy compared to other methods. With the wide application of LLMs, in the future work, we will further focus on the problem of detecting and classifying advanced steganography based on LLMs, while striving to balance model complexity and accuracy.

## REFERENCES


[1] Z. Saeidi, A. Yazdi, S. Mashhadi, M.Hadian, and A. Gutub, "High performance image steganography integrating IWT and Hamming code within secret sharing," *IET Image Processing*,vol.18,no.1,pp.129-139, Sep.2023.

[2] T. AlKhodaidi, and A. Gutub, "Refining image steganography distribution for higher security multimedia counting-based secret-sharing,"*Multimedia Tools and Applications*, vol.80, pp.1143-1173, Jan. 2021.

[3] J. Wang, R. Zhang, and J. Liu, "Partial-privacy image encryption algorithm based on time-varying delayed exponentially controlled chaotic system," *Nonlinear Dynamics*, vol. 112, pp. 10633-10659, May. 2024.

[4] W. Peng, T. Wang, Z. Qian, S. Li, and X. Zhang, "Cross-modal text steganography against synonym substitution-based text attack," *IEEE Signal Processing Letters*, vol. 30, pp. 299-303, Mar. 2023.

[5] R. Thabit, N. I. Udzir, S. M. Yasin, A. Asmawi, and A.Gutub, "CSNTSteg: Color Spacing Normalization Text Steganography model to improve capacity and invisibility of hidden data," *IEEE Access*, vol. 10, pp.65439-65458, Jun. 2022.

[6] J. Hemalatha, M. Sekar, C. Kumar, A. Gutub, and A. Sahu, "Towards improving the performance of blind image steganalyzer using third-order SPAM features and ensemble classifier," *Journal of Information Security and Applications*,vol.76, pp.103541, Aug. 2023.

[7] A. Aljarf, H. Zamzami, and A.Gutub, "Integrating machine learning and features extraction for practical reliable color images steganalysis classification," *Soft Computing*, vol.27, no.19, pp. 13877-13888, Jul. 2023.

[8] A. Aljarf, H. Zamzami, and A. Gutub, "Is blind image steganalysis practical using feature-based classification?," *Multimedia Tools and Applications*, vol.83, no.2, pp.4579-4612, Jan.2024.

[9] Y. Wang, L. Li, Y. Tang, R. Zhang, J. Liu, "Toward Copyright Integrity and Verifiability via Multi-Bit Watermarking for Intelligent Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems*,Jan. 2025.

[10] N. A. Roslan, N. I. Udzir, R. Mahmod, A. Gutub, "Systematic literature review and analysis for Arabic text steganography method practically," *Egyptian Informatics Journal*, vol. 23, no. 4, pp. 177-191, Dec. 2022.

[11] H. Yang, and X. Cao, "Linguistic steganalysis based on meta features and immune mechanism," *Chinese Journal of Electronics*, vol. 19, no.4, pp. 661-666, Oct. 2010.

[12] L. Xiang, X. Sun, G. Luo, and B. Xia, "Linguistic steganalysis using the features derived from synonym frequency," *Multimedia tools and applications*, vol. 71, pp. 1893-1911, Dec. 2014.

[13] C. Qi, S. Xingming, and X. Lingyun, "A secure text steganography based on synonym substitution," *in Proceeding of the IEEE Conference Anthology*, pp. 1-3, 2013.

[14] W. Peng, J. Zhang, Y. Xue, and Z. Yang, "Real-time text steganalysis based on multi-stage transfer learning," *IEEE Signal Processing Letters*, vol. 28, pp. 1510-1514, Jul. 2021.

[15] Z. Yang, K. Wang, J. Li,Y. Huang, and Y. Zhang, "TS-RNN: Text steganalysis based on recurrent neural networks," *IEEE Signal Processing Letters*, vol.26, no.12, pp.1743-1747, Jun. 2019.

[16] J. Zou, Z. Yang, S. Zhang, S. Rehman, and Y. Huang, "High-performance linguistic steganalysis, capacity estimation and steganographic positioning," *in Proceeding of the International Workshop on Digital Watermarking*, pp. 80-93, 2020.

[17] H. Wang, Z. Yang, J. Yang, C. Chen, and Y. Huang, "Linguistic steganalysis in few-shot scenario," *IEEE Transactions on Information Forensics and Security*, vol.18, pp. 4870-4882, Jul. 2023.

[18] Y. Wang, R. Song, R. Zhang, and J. Liu, "UP4LS: User Profile Constructed by Multiple Attributes for Enhancing Linguistic Steganalysis," *arXiv preprint arXiv:2311.01775*, 2023.

[19] M. Bai, J. Yang, K. Pang, H. Wang, and Y. Huang, "Towards Next-Generation Steganalysis: LLMs Unleash the Power of Detecting Steganography," *arXiv preprint arXiv:2405.09090*, 2024.

[20] Y. Tang, Y. Wang, R. Zhang, and J. Liu, "Rethinking LLM and Linguistic Steganalysis: An Efficient Detection of Strongly Concealed Stego," *arXiv preprint arXiv:2406.04218*, 2024.

[21] V. D. L, "Accelerating t-sne using tree based algorithms," *The journal of machine learning research*, vol. 15, no.1, pp.3221-3245, Jan. 2014.

[22] V. Banoci, G. Bugar, M.Broda, and D. Levicky, "Multi-classification model for image steganalysis," in *Proceedings of the Croatian Society Electronics in Marine International Symposium*, Sep. 2013.

[23] L. Zhao, Y. Chi, Z. Xu, and Z. Yue, "Block Cipher Identification Scheme Based on Hamming Weight Distribution," *IEEE Access*, vol.11, pp. 21364-21373, Feb. 2023.

[24] E. Hu, Y. Shen, P. Wallis, Z. Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," *arXiv preprint arXiv:2106.09685*, 2021.

[25] Z. Yang, S. Zhang, Y. Hu, Z. Hu, and Y. Huang, "VAE-Stega: Linguistic Steganography Based on Variational Auto-Encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880–895, Sep. 2021.

[26] Z. Yang, X. Guo, Z. Chen, Y. Huang, and Y. Zhang,"RNN-stega: Linguistic steganography based on recurrent neural net-works," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1280-1295, Sep. 2019.

[27] S. Zhang, Z. Yang, J. Yang, and Y. Huang, "Provably secure generative linguistic steganography," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 3046–3055, 2021.

[28] T. Fang, M. Jaggi, and K. Argyraki, "Generating steganographic text with LSTMs," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics- Student Research Workshop*, pp. 100–106, 2017.

[29] Y. Niu, J. Wen, P. Zhong and Y. Xue, "A Hybrid R-BILSTM-C Neural Network Based Text Steganalysis," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1907-1911, Nov. 2019.

[30] Y. Xu, T. Zhao, and P. Zhong, "Small-scale linguistic steganalysis for multi-concealed scenarios," *IEEE Signal Processing Letters*, vol.29, pp.130-134, Nov. 2021.

[31] A. Zeng, B. Xu, B. Wang, C. Zhang, D. Yin, et al, "ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools," *arXiv preprint arXiv:2406.12793*, 2024.

[32] I. Loshchilov, and F. Hutter, "Decoupled weight decay regularization," in *Proceedings of International Conference on Learning Representations(ICLR)*, 2019.

[33] A. Zeng, X. Liu, Z. Du, Z. Wang, H. Lai, et al, "GLM-130B: An Open Bilingual Pre-trained Model," in *Proceedings of International Conference on Learning Representations(ICLR)*, 2023.

[34] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, et al, ""Llama 2: Open Foundation and Fine-Tuned Chat Models," *arXiv preprint arXiv:2307.09288*, 2023.