EasyChair Preprint

№ 14738

BIOGAN-BERT: BioGPT-2 Fine Tuned and GAN-BERT for Extracting Drug Interaction Based on Biomedical Texts

Made Arbi Parameswara and Rila Mandala

September 6, 2024

BIOGAN-BERT: BioGPT-2 Fine Tuned and GAN-BERT for Extracting Drug Interaction Based on Biomedical Texts

Made Arbi Parameswara*
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
23522002@mahasiswa.itb.ac.id

Rila Mandala
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rila@itb.ac.id

***Abstract*—Drug-drug interactions (DDIs) occur when two or more drugs are used together, leading to unexpected and potentially harmful effects. Identifying DDIs requires manual annotations, but the increasing volume of research publications and the slow data annotation process make this challenging. Machine learning, especially deep learning, can efficiently extract and identify DDIs from biomedical literature. However, class imbalance in datasets reduces model performance. This study introduces BIOGAN-BERT, which combines data augmentation using the Pretrained Language Model (PLM) BioGPT-2 and Generative Adversarial Network (GAN) to address class imbalance in DDI extraction tasks. It identifies gaps in existing imbalance handling studies and proposes enhancements through PLM-based data augmentation and semi-supervised learning with GAN. BioGPT-2 generates additional data from labeled and unlabeled sources, enriching the training dataset. This data is then processed using GAN-BERT, allowing the model to learn from more complex data distributions, thereby improving data quality and model generalization. Traditional methods like sampling only increase the number of data instances, and loss functions merely assign greater representation to the loss values. While these methods expand the learning space for models, they do not enhance data representation. In contrast, this novel approach uses data augmentation to increase both the quantity and the diversity of data. Evaluation results show that BIOGAN-BERT outperforms several baselines, significantly increasing the micro F1-Score for minor classes to 0.85 compared to 0.83 for the best baseline model, demonstrating its effectiveness in handling class imbalance and contextual variations in biomedical data.**

***Index Terms*—Drug-drug interactions (DDI), Machine learning, Data augmentation, Deep learning, Generative Adversarial Network (GAN)**

I. Introduction

DDIs occur when two or more drugs are used simultaneously, leading to potentially harmful effects such as reduced efficacy, toxic effects, or exacerbation of side effects. Identifying DDIs typically involves using specific datasets that have been annotated by experts, such as PharmGKB and DrugBank.com. However, the increasing number of drugs and continuous publication of new research makes maintaining up-to-date DDI data challenging. The manual annotation process is time-consuming and costly, often resulting in a significant amount of unlabelled DDI data [1].

To address the need for efficient DDI extraction, machine learning, especially deep learning techniques, has become essential. These models predict drug interactions effectively, helping healthcare professionals and researchers. One of the datasets used to train these models is the DDI Extraction 2013 dataset. The DDI Extraction Challenge 2013 dataset, a gold standard for developing extraction models, consists of texts from DrugBank and MedLine, annotated with five types of DDI relationships. Despite improvements in model performance, class imbalance remains a significant challenge, leading to poor performance, especially for minor classes. Some researchers have used solutions like weighted cross-entropy and sampling, but with limited success [2]–[4].

Traditional methods like sampling only increase the number of data instances, and loss functions merely assign greater representation to the loss values. While these methods expand the learning space for models, they do not enhance data representation. In contrast, this novel approach uses data augmentation to increase both the quantity and the diversity of data. One of the best traditional methods, using PLMs, leverages only labeled data for augmentation.

This study proposes a novel approach that combines a fine-tuned generative pre-trained language model, such as BioGPT-2, with a generative probabilistic architecture. This method not only increases the number of instances to expand the learning space but also leverages both labeled and unlabeled data to enhance the model's extraction capabilities. The goal is to improve machine learning model performance on the DDI Extraction Challenge 2013 dataset. This approach is expected to surpass the results of previous methods like sampling, data augmentation, and adapting loss functions, providing a more robust solution for DDI extraction from biomedical texts.

*Corresponding author: M. A. Parameswara (email: 23522002@mahasiswa.itb.ac.id)

II. RELATED WORKS

This section will discuss all related works that are used as the reference and used in developing solutions.

A. Drug-drug interactions

Relation extraction is an Natural Language Processing (NLP) technique used to identify and extract relationships between entities or concepts in a text, enhancing understanding and analysis. In DDIs, this technique processes sentences to extract relevant DDI information. This process is crucial as DDIs can reduce drug efficacy or increase toxicity, posing significant clinical risks, especially for the elderly. Studies show that DDIs can decrease the quality of life and increase healthcare costs. With the rise of new drugs, deep learning models have become essential for efficiently extracting and identifying DDIs from biomedical literature [5].

B. GAN-BERT

GAN are machine learning models consisting of a generator that creates synthetic data and a discriminator that distinguishes between real and synthetic data. Through competitive training called adverserial learning, GANs improve until the generated data is indistinguishable from real data [6]. A research combining GANs with BERT performs well in text classification tasks, especially with limited labeled data, by using a mix of labeled and unlabeled data [7]. A semi-supervised learning approach, SS-GAN, trains the discriminator with $c+1$ classes where $(1, \ldots, c)$ are the "true" classes based on the training data targets, and $c+1$ is the "false" class generated by the generator [8]. A GAN-BIOBERT model applied for sentiment detection in clinical data, using 108 annotated and 2,000 unlabeled data points, achieved an accuracy of 0.91 and an F1-macro score of 0.92, demonstrating the effectiveness of GANs in improving performance with limited labeled data [9].

C. Fine-Tuned PLM

This research uses two PLMs, BIOBERT and GPT-2, to enhance text representation and perform data augmentation for DDI detection. BIOBERT captures contextual information from biomedical texts, crucial for relation extraction, while GPT-2, being freely available, is used for data augmentation. The data augmentation involves fine-tuning BioGPT-2 with specific datasets to generate relevant augmented data, improving the model's performance in DDI extraction. This approach addresses the issue of limited labeled data and helps the model recognize complex patterns, enhancing accuracy and effectiveness in detecting DDIs from biomedical texts.

D. Imbalance Class Handling Methods

Based on a review addressing class imbalance in machine learning, particularly in NLP, several main approaches can be identified: data sampling, data augmentation, and loss function adaptation [10].

1) **Sampling**: This technique increases instances of the minority class by randomly duplicating them to help the model learn better from underrepresented classes.
2) **Data Augmentation**: This method enriches training data by restructuring syntax and adding lexical items to create a more balanced dataset without needing new data collection.
3) **Loss Function Adaptation**: This approach adjusts the loss function to assign a higher penalty for misclassifications in the minority class, encouraging the model to better learn patterns from the underrepresented data.

E. Augmentation Filter

Data augmentation and unlabeled data in training must be meticulously filtered to avoid degrading model performance. Inaccurate or irrelevant data can lead to substantial bias or inaccuracies in predictions. Ensuring the quality and relevance of the data is paramount to developing an effective model. Preliminary research by [11] highlights two key metrics that are suitable for filtering training data in this research.

1) **Diversity (Bilingual Evaluation Understudy (BLEU) Score**: This metric evaluates text quality by comparing n-gram matches to ensure the model learns from varied patterns, including a brevity penalty for balanced evaluation. The BLEU metric is used in this research due to its efficiency in measuring the diversity between augmented data and the original data.
2) **Accuracy**: This metric assesses how well the additional data aligns with given labels, preventing the model from being misled and ensuring relevance to the provided labels.

F. Related Research

The related research explores various techniques for handling data imbalance in drug-drug interaction (DDI) extraction from biomedical texts. Models like the one employing a Dependency Graph Enhanced Module and Sequential Representation Generation Module, which uses an attention network to differentiate connected nodes and under-sampling for balanced data distribution, have improved accuracy in detecting drug interactions [12]. Similarly, combining under-sampling with a drug knowledge graph (KG) from sources like DrugBank and KEGG, along with Bi-GRU for contextual information and BioBERT as a text encoder, achieved an F1-score of 0.81 [13]. Another approach integrating BioBERT with Bi-LSTM yielded an F1-macro score of 0.83 [4]. Several studies also explored combining imbalance handling methods: random under-sampling and weighted cross-entropy with PLM BIOBERT and BioGPT-2 achieved an F1-score of 0.84 [2], while under-sampling and weighted cross-entropy with deep neural networks resulted in an F1-score of 0.78 [14]. CNN with SMOTE improved performance by 0.04 [3], and SMOTE combined with GAN enhanced minor class samples, boosting performance from 0.84 to 0.96 [15]. A combination of under-sampling and over-sampling achieved an AUC of 0.99 [16], and a new model, Prompt Tuning and Data Augmentation (PTDA), leveraging GPT-2 for data augmentation and prompt tuning, achieved an F1-micro score of 0.85 on the DDI Extraction 2013 dataset [17].

III. Method & Implementation

This section details the research methods and implementation, offering a comprehensive overview of the approaches and techniques used to investigate the study objectives.

A. Dataset

This research uses two types of datasets: the primary DDI Extraction 2013 dataset, manually annotated by pharmacists and linguists to identify harmful drug interactions, and additional unlabeled datasets. Each label as shown in Table: I is the primary dataset provides specific information about the interactions.

TABLE I
Labels and their Functions in DDI Extraction

Label	Function
Mechanism	Describes the pharmacokinetic mechanism of a drug interaction.
Effect	Indicates the effects resulting from drug interactions.
Advise	Offers recommendations or advice regarding drug interactions.
Int	Indicates an interaction without detailed information.
Negative	States that no interaction occurs.

The TAC 2019 dataset, created by the FDA, is used in this research to extract drug-drug interaction (DDI) information from structured product labels of 406 drugs. This dataset supports automated health information exchange systems and is employed as unlabeled data for augmentation to enhance model performance. Detailed information about this dataset can be found in the article "Overview of the TAC 2019 Track on Drug-Drug Interaction Extraction from Drug Labels" [18]. The labels from the TAC 2019 data are intentionally removed, as it is used as unlabeled data for training purposes. All the TAC 2019 data is utilized solely for training because it serves as an augmentation dataset.

B. Experiment Settings

This study uses an experimental approach to test the proposed hypothesis, allowing systematic and controlled research to evaluate the effectiveness of various algorithms in addressing class imbalance in DDI extraction from biomedical texts. By using an experimental approach, researchers can objectively compare the performance of different methods and algorithms. The dependent variable is the F1-Micro metric for minor classes in the DDI Extraction 2013 dataset, such as advise, effect, int, and mechanism. Stratified cross-validation with k = 5 is used to ensure robust evaluation of these metrics.

The independent variable in this study is the variation of class imbalance handling algorithms [10], selected to save research time and reduce computational costs. The algorithms tested and compared for improving DDI extraction performance as described in Table II include under-sampling, over-sampling, hybrid-sampling, data augmentation with GPT-2, weighted cross-entropy (WCE), Auto Dice Loss (ADL), Synthetic Minority Over-sampling Technique (SMOTE), and the novel method BIOGAN-BERT. The experimental hardware

TABLE II
Description of Methods Used for Data Imbalance Handling

Method	Description
No Handler	Does not apply any method to address data imbalance.
Under-sampling	Reduces the number of majority class instances to balance the dataset.
Over-sampling	Increases the number of minority class instances to balance the dataset.
Hybrid-sampling	Combines both under-sampling and over-sampling techniques to balance the dataset.
SMOTE	Synthesizes new instances for the minority class using k-nearest neighbors.
GPT-2	Uses data augmentation through synthetic text generation via the GPT-2 model.
ADL	Implements Auto Dice Loss to optimize model training and focus on minority classes.
WCE	Uses Weighted Cross Entropy to give higher importance to the minority classes during training.
BIOGAN-BERT	Combines a GAN model with BERT to generate synthetic data and improve the performance of the DDI extraction task.

includes the Tambora Academic Research Server at STEI ITB, featuring an Intel® Xeon® Silver 4208 CPU, 64 GB of RAM, and a combination of Nvidia Quadro RTX 5000 and Nvidia RTX A5000 GPUs. GPU usage is managed through a FIFO booking system, ensuring availability for the training process.

C. BIOGAN-BERT Architecture

The BioGAN-BERT architecture has three key components: processing unlabeled data, fine-tuning BioGPT-2 for data augmentation, and the classifier model. Each component is detailed in this section.

1) Unlabeled Data Processing: This section discusses the process of handling unlabeled TAC 2019 data through various stages of data processing to ensure the best quality data is used in the training process, as shown in Fig: 2, thereby improving the extractor's capabilities. The TAC 2019 data is transformed into CSV format and filtered to match the characteristics of the DDI Extraction 2013 dataset. This involves padding, tokenizing, masking, and entity-tagging sentences that contain drug pair names. Pseudo-labeling is used to generate additional labeled data for training and data augmentation, optimizing the use of unlabeled data more effectively than previous methods by addressing confidence levels between the "unknown" class and other classes.

2) Generate Augmented Data with BioGPT-2: The augmentation stage, as shown in Fig: 3, involves combining pseudo-labeled data from the unlabeled TAC 2019 dataset with the training data from the DDI Extraction 2013 dataset. This is followed by pre-processing and fine-tuning the model using metrics like perplexity and hyperparameters such as the AdamW optimizer and early stopping. After fine-tuning, the model generates new data using top-k and top-p sampling algorithms (k=50, p=0.8) to increase the amount of data in the minor class. The labels are ensured to match the DDI Extraction dataset because the generated data from the fine-tuned model uses prompt labels, ensuring consistency

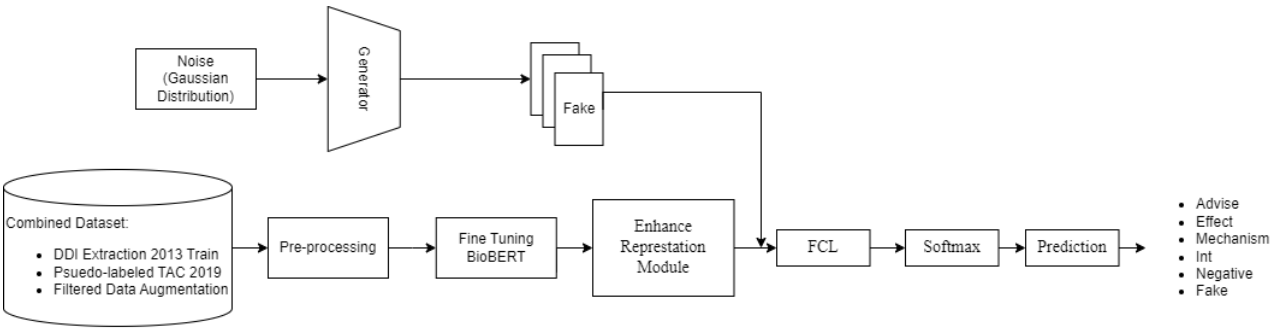

Fig. 1. Architecture of Classifier

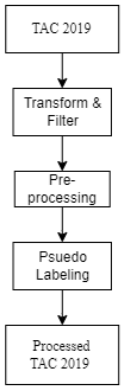

Fig. 2. Unlabelled Data Process

throughout. The use of labels as the first word is also chosen to produce more focused and contextual augmented data [19]. The generated data undergoes filtering to maintain accuracy, using a filtering protocol based on the best baseline machine learning model, the oversampling model, to ensure that sentences contain specific entity token pairs, have a minimum of three words, and correctly enclose entity names. Quality metrics such as diversity (BLEU) and correctness, benchmarked against existing models, ensure the effectiveness of the synthetic data.

3) Classifier: The classifier architecture, shown in Fig: 1, utilizes unlabeled datasets, data augmentation, and the DDI Extraction 2013 training data with BioGPT-2. Details of the data instances can be seen in Table III. The study uses the training and testing datasets from the DDI Extraction 2013 dataset, following the approach of other researcher [2], and employs 5-fold stratified cross-validation for performance evaluation. This approach combines and modifies the methods of GAN-BERT [7] and EGFI [2] to enhance text representation for DDI extraction. The generator uses noise input to produce synthetic data, improving the model's generalization. Key hyperparameters include a Micro F1-score metric, the AdamW optimizer, cross-entropy loss, a batch size of 8, 50 epochs, and the pre-trained model `monologg/biobert_v1.1_pubmed`.

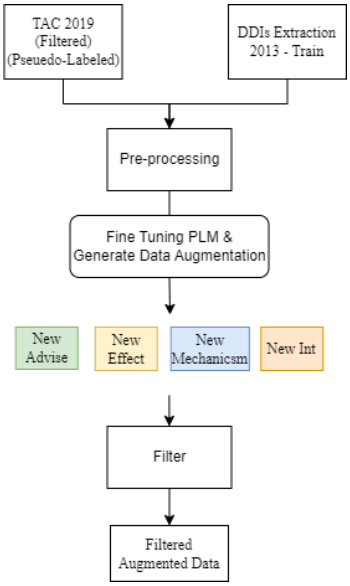

Fig. 3. Data Augmentation Process

IV. Research Results & Discussion

This section details the results and discussion of the proposed method.

A. Unlabaled Data

To transform a dataset from .XML format, special tokens are first added based on recognized drug names from the TAC 2019 dataset. Unrecognized entities are adjusted and assigned types based on the DDI Extraction 2013 dataset. Sentences that do not match the DDI Extraction 2013 structure are removed, resulting in 3112 instances. These instances are labeled using the pseudo-labeling method to enhance model accuracy during training. Pseudo-labeling information is then used in the data augmentation process by BioGPT-2, improving fine-tuning. The pseudo-labeling process uses the baseline model with the best performance, with data augmentation statistics shown in Table III.

TABLE III
DATA DISTRIBUTION ACROSS DIFFERENT LABELS

Label	Train DDI Extraction 2013	Pseudo Labeled TAC 2019	Augmented BioGPT-2	Total
Advise	728	1218	1649	3601
Effect	1565	659	232	2487
Int	177	27	1098	1303
Mechanism	1153	422	98	1692
Negative	8987	757	0	9752

B. Augmented Data

The fine-tuning process using BioGPT-2 involves generating data for each label, approximately doubling the instances of the major class. Following this, a filtering process is applied in three stages: structure, accuracy, and diversity. The structure stage ensures that the augmented data matches the format of the DDI Extraction 2013 data, with one drug pair per sentence. The accuracy stage ensures the augmented data matches the assigned labels, discarding incorrect instances. The diversity stage ensures the data is both accurate and varied, enhancing the model's generalization capability. The resulting augmented data instances are detailed in Table III.

C. Model Performance

TABLE IV
MICRO F1-SCORE VALIDATION AND TEST RESULTS FOR DIFFERENT IMBALANCE HANDLERS

Imbalance Handler	Micro F1-Score Validation	Micro F1-Score Test
No Handler	0.87	0.82
Under-sampling	0.71	0.68
Over-sampling	**0.89**	0.83
Hybrid-sampling	0.86	0.81
SMOTE	0.56	0.47
GPT-2	0.88	0.82
ADL	0.89	0.82
WCE	0.88	0.80
BIOGAN-BERT	0.88	**0.85**

As shown in Table IV all of the dependent variable algorithms are presented here. Several algorithms were compared in this study.

Under-sampling resulted in low F1 scores (0.71 for validation and 0.68 for testing) due to significant information loss from reducing majority class instances, which decreased model performance. Hybrid sampling performed better (0.86 for validation and 0.81 for testing) but was still less effective compared to no imbalance handling, as random data reduction in these methods diminished the model's generalization ability.

Conversely, over-sampling significantly improved performance, with F1 scores of 0.89 for validation and 0.83 for testing. This technique increased the number of minority class instances, allowing the model to learn from a more balanced dataset, thereby enhancing overall performance.

WCE also performed well (0.88 for validation and 0.80 for testing) by assigning higher weights to minority classes, though it required careful hyperparameter tuning to optimize performance. ADL outperformed WCE by dynamically adjusting weights and data distribution during training, leading to balanced data representation and improved model accuracy.

BIOGAN-BERT achieved the highest scores on the test set (0.88 for validation and 0.85 for testing), demonstrating superior data imbalance handling. Its low overfitting and high generalization capability resulted from data augmentation PLMs like BioGPT-2, which generated realistic biomedical text to enhance model training. The adaptive learning mechanisms in BIOGAN-BERT further optimized data distribution handling, confirming the efficacy of PLM-based data augmentation in producing robust models.

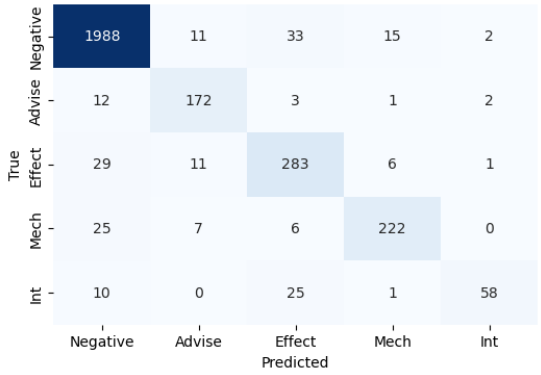

Fig. 4. Confusion Matrix of BIOGAN-BERT

Based on confusion matrix on Figure 4 shows the performance of the BIOGAN-BERT model in classifying five different classes: Negative, Advise, Effect, Mechanism (Mech), and Interaction (Int). The matrix highlights how well the model predicted each class, with diagonal values representing correct predictions (true positives). For instance, the model correctly identified 1988 instances of the "Negative" class and 172 instances of the "Advise" class. Misclassifications are represented in the off-diagonal values, indicating areas where the model confused one class for another, such as 33 "Negative" instances being classified as "Effect." Overall, the matrix demonstrates that BIOGAN-BERT performs well across most classes, with some confusion in more complex categories like "Int" and "Mechanism"

BIOGAN-BERT outperformed other methods due to its ability to increase data instances using effective data representation based on BioGPT-2 augmentation and generating synthetic data with GAN-BERT. This provided the model with better generalization capabilities across all classes. However, due to its reliance on PLM architecture, the performance of BIOGAN-BERT is heavily dependent on the chosen PLM. Selecting an appropriate PLM is crucial, as models like GPT-2 (not BioGPT-2) still have 0.01 lower performance than oversampling method.

D. Model Performance

The ablation study presented in Table V highlights the contribution of each component in the BIOGAN-BERT ar-

TABLE V
ABBLATION STUDY

Imbalance Handler	Micro F1-Score Validation	Micro F1-Score Test
BIOGAN-BERT	0.88	**0.85**
No *GAN-BERT*	0.87	0.83
No *BioGPT-2*	**0.89**	0.83
No *BIOGAN-BERT*	0.87	0.82

chitecture, specifically BioGPT-2-based augmentation, GAN-BERT, and the full BIOGAN-BERT model. The results show that BIOGAN-BERT with all components achieves the highest Micro F1 score on the test data (0.85), while the best validation score is obtained without BioGPT-2 (0.89). Removing GAN-BERT reduces the Micro F1 score to 0.87 for validation and 0.8311 for testing, indicating its critical role in handling data imbalance by generating synthetic data. Without BioGPT-2 augmentation, the Micro F1 score drops to 0.83 on testing, emphasizing its importance in improving model generalization through data variability. The model without BIOGAN-BERT scores 0.87 for validation and 0.82 for testing, demonstrating the significant contributions of both GAN-BERT and BioGPT-2 in the overall performance.

V. CONCLUSION & FUTURE WORK

This research developed a new model architecture, BIOGAN-BERT, combining Fine-Tuned BioGPT-2 and GAN-BERT for DDI extraction tasks. BioGPT-2's data augmentation enriched the training dataset with additional labeled and unlabeled data, which was then processed semi-supervised using GAN-BERT. This approach expanded the dataset and improved data quality, resulting in a model with better generalization and accuracy in handling context variations. BIOGAN-BERT demonstrated superior performance, achieving a Micro F1-Score of 0.85, surpassing the 0.83 score of over-sampling, highlighting the effective combination of data augmentation and semi-supervised learning techniques in addressing data imbalance and context variation challenges in biomedical data.

Further development using better methods to utilize unlabeled data can involve several semi-supervised approaches, such as SALTClass. SALTClass enriches limited biomedical texts with unlabeled data through clustering algorithms, which then enhance text representation. This method also integrates various supervised techniques and has been proven to significantly improve biomedical text classification [20]. By optimizing the use of this method, it is expected that more data will pass the filters, thereby significantly enhancing the model's capabilities through the addition of relevant data.

REFERENCES

[1] E. Spina, C. Hiemke, and J. de Leon, "Assessing drug-drug interactions through therapeutic drug monitoring when administering oral second-generation antipsychotics," *Expert opinion on drug metabolism & toxicology*, vol. 12, no. 4, pp. 407–422, 2016, publisher: Taylor & Francis.

[2] L. Huang, J. Lin, X. Li, L. Song, Z. Zheng, and K.-C. Wong, "Egfi: drug–drug interaction extraction and generation with fusion of enriched entity and sentence information," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab451, 2022.

[3] S. Deepika, M. Saranya, and T. Geetha, "Cross-corpus training with cnn to classify imbalanced biomedical relation data," in *Natural Language Processing and Information Systems: 24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019, Salford, UK, June 26–28, 2019, Proceedings 24*. Springer, 2019, pp. 170–181.

[4] M. KafiKang and A. Hendawi, "Drug-drug interaction extraction from biomedical text using relation biobert with blstm," *Machine Learning and Knowledge Extraction*, vol. 5, no. 2, pp. 669–683, 2023.

[5] T. Zhang, J. Leng, and Y. Liu, "Deep learning for drug–drug interaction extraction from the literature: a review," *Briefings in bioinformatics*, vol. 21, no. 5, pp. 1609–1627, 2020.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[7] D. Croce, G. Castellucci, and R. Basili, "Gan-bert: Generative adversarial learning for robust text classification with a bunch of labeled examples," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 2114–2119.

[8] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

[9] J. J. Myszewski, E. Klossowski, P. Meyer, K. Bevil, L. Klesius, and K. M. Schroeder, "Validating gan-biobert: a methodology for assessing reporting trends in clinical trials," *Frontiers in Digital Health*, vol. 4, p. 878369, 2022.

[10] S. Henning, W. Beluch, A. Fraser, and A. Friedrich, "A survey of methods for addressing class imbalance in deep-learning based natural language processing," *arXiv preprint arXiv:2210.04675*, 2022.

[11] S. Gandhi, R. Gala, V. Viswanathan, T. Wu, and G. Neubig, "Better synthetic data by retrieving and transforming existing datasets," *arXiv preprint arXiv:2404.14361*, 2024.

[12] Y. Shi, P. Quan, T. Zhang, and L. Niu, "Dream: Drug-drug interaction extraction with enhanced dependency graph and attention mechanism," *Methods*, vol. 203, pp. 152–159, 2022.

[13] H. Lu, D. Song, Y. Zhu, and L. Li, "Drug-drug interaction extraction using drug knowledge graph," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 3846–3848.

[14] M. Fatehifar and H. Karshenas, "Drug-drug interaction extraction using a position and similarity fusion-based attention mechanism," *Journal of Biomedical Informatics*, vol. 115, p. 103707, 2021.

[15] J. Wei, G. Feng, Z. Lu, P. Han, Y. Zhu, and W. Huang, "Evaluating drug risk using gan and smote based on cfda's spontaneous reporting data," *Journal of Healthcare Engineering*, vol. 2021, no. 1, p. 6033860, 2021.

[16] H. Yang and C. C. Yang, "Discovering drug-drug interactions and associated adverse drug reactions with triad prediction in heterogeneous healthcare networks," in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2016, pp. 244–254.

[17] K. Wang, X. Fu, Y. Liu, W. Chen, and J. Chen, "Ptda: Improving drug-drug interaction extraction from biomedical literature based on prompt tuning and data augmentation." *IAENG International Journal of Computer Science*, vol. 51, no. 5, 2024.

[18] T. Goodwin, D. Demner-Fushman, K. Fung, and P. Do, "Overview of the tac 2019 track on drug-drug interaction extraction from drug labels," 11 2019.

[19] K. Yue, B.-C. Chen, J. Geiping, H. Li, T. Goldstein, and S.-N. Lim, "Object recognition as next token prediction," pp. 16 645–16 656, 2024.

[20] A. Bagheri, D. Oberski, A. Sammani, P. G. van der Heijden, and F. W. Asselbergs, "Saltclass: classifying clinical short notes using background knowledge from unlabeled data," *bioRxiv*, p. 801944, 2019.