



HAL
open science

Self-supervised Segment Contrastive Learning for Medical Document Representation

Waheed Ahmed Abro, Hanane Kteich, Zied Bouraoui

► **To cite this version:**

Waheed Ahmed Abro, Hanane Kteich, Zied Bouraoui. Self-supervised Segment Contrastive Learning for Medical Document Representation. 22nd International Conference on Artificial Intelligence in Medicine, AIME 2024, Jul 2024, Salt Lake City, United States. pp.312-321, 10.1007/978-3-031-66538-7_31 . hal-04665027

HAL Id: hal-04665027

<https://hal.science/hal-04665027v1>

Submitted on 30 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Self-Supervised Segment Contrastive Learning for Medical Document Representation

Waheed Ahmed Abro¹[0000-0001-5878-3448], Hanane Kteich^{1,2}[0009-0003-5271-1103], and Zied Bouraoui^{1,2}[0000-0002-1662-4163]

¹ CDEP - UR 2471, Université d'Artois, France

² CRIL - CNRS & Université d'Artois, France

Abstract. Learning high-quality text embedding is vital for biomedical topic classification and many other NLP tasks. Contrastive learning has shown remarkable performance in generating high-quality text embeddings. However, existing methods typically generate anchor-positive pairs through discrete augmentations, simplifying the task of distinguishing positive from negative examples and limiting the learning of meaningful representations. In this paper, we present a self-supervised segment contrastive learning (SCL) approach designed for contrastively fine-tuning pre-trained language models. Our method randomly divides documents into anchor and positive segments, facilitating the learning of document embeddings by maximizing agreement between these segments. The proposed model contrastively fine-tune pre-trained Clinical-BioBERT language model to generate document embedding for medical documents. We evaluate our method on two publicly available medical datasets, MIMIC and Bioasq. Extensive experiments show that our proposed SCL approach outperforms baseline models, achieving superior performance in medical classification tasks.

Keywords: Document representation · Medical text · Contrastive learning · Language models.

1 Introduction

Medical data processing plays a vital role in developing intelligent healthcare services. It involves various techniques for acquiring, managing, and analyzing health-related data. The biomedical literature is characterized by the continuous generation of textual data such as daily research articles, clinical notes, and healthcare summaries. Medical documents have quite different linguistic characteristics from normal documents, which can be a problem when classifying medical topics. It is important to develop a quality document embedding to encode these special medical entities. Document embedding is the process of encoding the documents into dense vector representation such that similar documents appear closer in the vector space. Document embedding plays an important role in classifying biomedical documents [9] and several other NLP tasks like information retrieval [15, 2], multiple-choice QA [12], etc.

Traditional document embedding approaches are based on fixed-length features such as bag-of-words and TF-IDF [19]. However, these methods fail to capture word order and ignore semantic word distance, which limits their power on complex tasks. Recent advancements in transformer-based language models have led to a paradigm shift in document embedding, establishing them as the primary method for learning document representations from unlabelled corpora [13, 5, 4]. These models utilize vast amounts of unlabeled text data with a masked language modelling (MLM) objective to learn rich contextual sentence representations, which are then utilized by specific downstream tasks.

Despite their success, sentence representations generated by pre-trained language models suffer from anisotropy problems, where sentence representations occupy a narrow cone in the embedding space [6]. To alleviate this problem, contrastive learning [7, 28] has been proposed to bring similar sentences, i.e., anchor-positive pairs closer in the vector space, and dissimilar sentences, i.e., anchor-negative pairs far away in the vector space to uniform whole vector space.

The state-of-the-art models generate anchor-positive pairs by randomly augmenting the same sentence (e.g., Synonym replacement, delete one word, word repetition, dropout-noise), and anchor-negative pairs are chosen from the same mini-batch known as in batch-negative. Despite achieving success, employing discrete augmentations on sentences to generate positive pairs simplifies the task of distinguishing positive examples from negative ones; therefore, it does not lead to learning meaningful representations.

To address the above issues, we propose a self-supervised contrastive learning approach based on document-level objectives that can be used to contrastively fine-tune the pre-trained language model. Our proposed model randomly splits the document into two portions: a smaller text segment and a larger portion serving as an anchor text. The model learns document embeddings by training an encoder to maximize the agreement between anchor text and a positive text segment taken from the same document. As we are working with biomedical documents, we employed pre-trained ClinicalBioBERT [3] for contrastively fine-tuning.

Our primary contributions are:

- We propose an unsupervised self-supervised learning objective to contrastively fine-tune pre-trained ClinicalBioBERT model to induce high-quality document embeddings for medical documents.
- We conducted extensive experiments to highlight the advantages of learning document representation using proposed contrastive learning loss against state-of-the-art SimCSE and ESimCSE methods.
- We evaluated the quality of document embedding by training and evaluating multi-layer perceptron (MLP) classifier on top of document embedding on three biomedical classification datasets.

2 Related works

2.1 Medical document processing

Recently, transformer-based models are widely used for processing biomedical documents. To this end, [10] and [27] introduce sentence-aligned multilingual text simplification dataset for the medical domain, covering English, Spanish, French, Farsi, and Chinese. [11] and [1] introduce and investigate systems for generating medical reports from chest X-ray images. This includes integrating disease classification, transformer-based report generation, and an interpretation module to ensure consistency and clinical accuracy or evaluation metrics in text summarization and generation. [14] and [26] evaluated the capacity of LLM in medical domains, for clinical relation extraction or for medical systematic reviews, by adding medical knowledge into pre-trained models. [20] and [25] propose methods for medical prediction and automated medical report generation. [21] introduces MedCLIP, a framework for medical image-text contrastive learning that addresses challenges in pre-training on medical domain data. [17] survey discusses the application of pre-trained language models (PLMs) in the biomedical domain, highlighting their potential to improve performance on various natural language processing (NLP) tasks through pre-training on vast text corpora for universal language representation learning. In contrast, we propose to contrastive train these PLMs on the biomedical domain to generate high-quality document representation for medical documents.

2.2 Document Representation

Document representation learning employing self-supervised contrastive learning methods is a highly active research field [24, 7, 23, 18, 28]. In this direction, ConSERT [24] proposed to solve an anisotropy issue of BERT-derived sentence representation by contrastive training with adversarial attack, token shuffling, cutoff, and dropout augmentation methods. Similarly, SimCSE [7] uses a dropout version of the same sentence as a positive pair. In a similar context, ESIMCSE [23] proposed word repetition as an augmentation method to generate positive pairs. On the other hand, SNCSE [18] applied negation to produce a negative pair. Similarly, DCLR [28] used noise-based negatives. Unlike the mentioned works, our work maximizes the agreement between the anchor and positive text segments taken from the same document to generate document representation.

3 Methodology

3.1 Self-supervised contrastive learning

We propose a Segment Contrastive Learning (SCL) framework to train a pre-trained language model, i.e., ClinicalBioBERT and BERT-Base, using self-supervised contrastive learning. Our method maximizes the agreement between anchor text and a positive text segment taken from the same document. The model uses

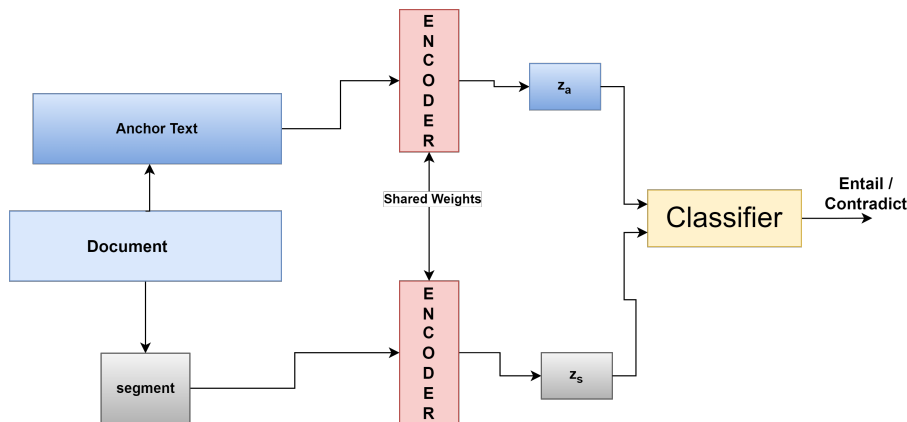


Fig. 1. Architecture of our proposed self-supervised learning objective. For each document in mini-batch, we construct a positive pair by partitioning it into two segments: a smaller text segment and a larger portion serving as an anchor text. Anchor text and positive segment text are passed through the same encoder to generate embeddings z_a and z_s , respectively. The encoder is trained to minimize the distance between embeddings of text vector z_a and segment vector z_s of the same document and maximize the distance with segments of other documents that serve as negative samples (not shown here due to simplicity).

siamese network architecture to learn embeddings of text documents. We form a positive pair (a, s^+) by splitting the document into two portions: a smaller text segment (s^+) and a larger portion serving as an anchor text (a). Anchor text and positive segment text are passed through the same encoder based on a pre-trained language model such as ClinicalBioBERT. The proposed method utilizes the [CLS] token representations to produce the anchor text (z_a) embedding and text segment (z_s^+) embedding; then the classifier is trained to maximize the agreement between the segment and anchor text, both sampled from the same document and minimize the agreement between anchor text and segment taken from another document.

Suppose we have a mini-batch of N documents, denoted as $D = \{(d_i)\}_1^N$. For each document d_i , we randomly select a text segment s^+ and the remaining text as anchor text a to form a positive pair (a_i, s_i^+) . We then select a negative segment s^- from the remaining $N - 1$ documents of the batch to serve as the negative pair, (a_i, s_j^-) . One concern that could be seen here is the text segments could be similar and fit on the anchor text of several documents. However, this is not an issue as in the training objective, we have multiple negatives, so our model is forced to optimize most dissimilar documents than most similar ones. It is important to note that the segment predictive contrastive learning process can be viewed as an unsupervised natural language inference task, where a positive segment sample represents an entailment of a document, and negative samples

from other documents represent a contradiction of the document. The multiple negatives ranking loss [8] function is used to optimize the model.

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n \frac{e^{(sim(z_{a_i}, z_{s_i}^+))}}{\sum_{k=1}^k e^{(sim(z_{a_i}, z_{s_k}^-))}}, \quad (1)$$

where z_{a_i} and $z_{s_i}^+$ are the vector representations of the anchor text segment and positive segment taken from the same document. The k negative sample of segments, vector representation is denoted by $z_{s_k}^-$, and sim is the cosine similarity. Multiple negative ranking loss compares the positive pair representation with the negative pair samples in mini-batch. The generalized architecture is shown in Figure 1.

3.2 Document Classification

After producing robust document representation, we feed this representation as input to the classifier. The MLP classifier consists of three hidden layers and an output layer. We utilize the TANH activation function in the hidden layers and the sigmoid function in the output layer.

$$\hat{y} = \sigma(W * d + b) \quad (2)$$

where σ denotes the sigmoid activation function, d is the document representation, and W and b are the weights and biases of the hidden and output layers, respectively. Furthermore, the model minimizes the binary cross-entropy loss between true document labels and predicted document labels.

4 Experimental Setup

4.1 Datasets

We used two publicly available medical datasets namely, Medical Information Mart for Intensive Care (MIMIC-III) and Bioasq.

(MIMIC-III) [9] The MIMIC-III dataset comprises 50K discharge summaries from US hospitals, with each summary mapped to one or more (International Classification of Diseases, Ninth Revision) ICD-9 taxonomy labels. We utilized labels from the first level of the ICD-9 hierarchy. The dataset is partitioned into two categories: train and test. The training set encompasses 40000 summaries, while the test set consists of 10000, respectively. MIMIC-III is a multi-label classification dataset featuring 19 labels.

(BIOASQ) [16] The BIOASQ dataset comprises biomedical articles sourced from PubMed. Each article is annotated with concepts from the Medical Subject Headings (MeSH) taxonomy. We employed the 1st levels of the MeSH taxonomy. The dataset is divided into train and test categories. The training set contains 80000 summaries, and the test set consists of 20000, respectively. The BIOASQ is a multi-label classification dataset featuring 15 labels. Table 1 provides statistics of the datasets.

Table 1. Task type, number of classes, train and test examples statistics for each dataset.

Dataset	Task Type	No. Classes	Train	Test
MIMIC	Muti-label Classification	19	40000	10000
Bioasq	Muti-label Classification	15	80000	20000

4.2 Training Setup

We employ the ClinicalBioBERT³ and Bert-Base model⁴ pretrained-models. The proposed self-supervised SCL model is trained for 2 epochs with a batch size of 16 training samples. The length of the positive segments is set to 64 tokens. Adam optimizer with learning rates of $2e - 5$ and weight decay of 0.001 is used to optimize the training loss. We used the transformers [22] library to train our model. The MLP models are trained for 16 epochs and finetuning models are trained for 4 epochs. We truncate and pad the document text to align with a maximum sequence input length of 512 tokens.

4.3 Baseline Models

To assess the quality of the representation of the document learned through the self-supervised method, we compare model performance against the pre-trained language model document Embedding, state-of-the-art contrastive learning of unsupervised sentence embedding models in feature-based and end-to-end fine-tuning settings:

- Embedding + MLP Classification: In this setting, we have frozen the parameters of the language model (ClinicalBioBERT and BERT) and applied an MLP classification layer for predicting the medical taxonomy labels.
- SimCSE [7] and ESIMCSE Embedding [23] + MLP Classification: In this setting, we trained ClinicalBioBERT and BERT in a contrastive manner using the SimCSE and ESIMCSE objectives. After training, we freeze the parameters and applied an MLP classification layer for predicting the medical taxonomy labels.
- BERT Classifier (Fine-tune): In the pre-trained BERT model [5], we employ a linear layer atop the final encoder layer’s [CLS] token to classify medical taxonomy labels. The model is fine-tuned for 4 epochs with a batch size of 16.
- ClinicalBioBERT Classifier (Fine-tune): In the pre-trained ClinicalBioBERT model [3], we incorporate a linear layer on top of the last encoder layer [CLS] token for classifying medical taxonomy labels. The model is fine-tuned for 4 epochs with a batch size of 16.

³ <https://huggingface.co/Dinithi/ClinicalBioBERT>

⁴ <https://huggingface.co/google-bert/bert-base-uncased>

4.4 Feature-based Document Classification

Table 2 presents medical taxonomy classification results in the feature-based embedding setting where the ClinicalBioBERT and BERT encoder features are not updated during training. The table illustrates the results of various models across two distinct datasets, namely MIMIC and Bioasq, in terms of micro and macro F1 scores. The top rows display the performance of models utilizing ClinicalBioBERT Feature-based embedding. The ClinicalBioBERT Embedding_{SCL} + MLP model produces the highest macro and micro F1 scores, achieving 54.65 and 71.28 macro F1-score on MIMIC and Bioasq datasets, respectively. This indicates that self-supervised *SCL* learning produces high-quality embeddings. Conversely, the state-of-the-art ClinicalBioBERT Embedding_{SimCSE} + MLP and ClinicalBioBERT Embedding_{ESimCSE} + MLP models does not enhance the performance of the baseline model Embedding + MLP. This observation suggests that relying solely on dropout augmentation or simple repetition of words to construct positive pairs and generate text embeddings may not yield significant improvements at the document or paragraph level embeddings. This stands in contrast to their demonstration of strong performance for sentence embeddings. Results demonstrate that the proposed SCL method improves embedding derived from ClinicalBioBERT and ClinicalBioBERT Embedding_{ESimCSE} by around 4% and 3% macro-F1 score on the MIMIC and Bioasq dataset, respectively.

Furthermore, shifting the focus to model’s performance utilizing BERT-base Feature-based embeddings, we observe similar trends in performance. The BERT Embedding_{SCL} + MLP model maintains its superiority with the highest macro F1 scores of 50.73 and 70.51 on MIMIC and Bioasq datasets, respectively. The proposed model outperforms all methods by producing approximately 8% and 3% better results in terms of macro-F1 on MIMIC and Bioasq datasets, respectively. Furthermore, results show that models trained on BERT Embedding models produce inferior results than models trained on ClinicalBioBert models. This indicates that training on domain-specific embeddings designed for biomedical text yields superior performance in medical classification tasks. Overall, the proposed SCL method generates high-quality embeddings which leads to model superior performance across different datasets and settings.

4.5 End-to-end Document Classification

To assess the quality of the embeddings learned by the proposed SCL model, we additionally perform end-to-end fine-tuning on the MIMIC and Bioasq datasets. The results are presented in Table 3. It is clear from the results that the SCL model outperforms other methods, achieving an improvement of approximately 1.2% and 1.6% in macro F1 score compared to the ClinicalBioBERT classifier and ClinicalBioBERT SimCSE classifier, respectively, on the MIMIC dataset. Additionally, there is an increase in performance of approximately 1% on the Bioasq dataset. It is apparent from the results that the BERT ESimCSE produces second-best results as it utilizes word repetition to de-bias the length for positive pairs.

	MIMIC		Bioasq	
	macro-F1	μ -F1	macro-F1	μ -F1
ClinicalBioBERT Embedding + MLP [3]	50.36	65.43	68.05	83.60
ClinicalBioBERT Embedding _{SimCSE} + MLP [7]	48.01	62.61	69.13	83.31
ClinicalBioBERT Embedding _{ESimCSE} + MLP [23]	50.85	64.42	68.77	83.17
ClinicalBioBERT Embedding _{SCL} + MLP	54.65	66.41	71.28	84.43
BERT Embedding + MLP [5]	40.23	58.04	68.64	83.30
BERT Embedding _{SimCSE} + MLP [7]	40.76	57.64	68.05	82.70
BERT Embedding _{ESimCSE} + MLP [23]	42.02	58.09	68.04	82.66
BERT Embedding _{SCL} + MLP	50.73	65.23	70.51	84.08

Table 2. The proposed model performance against baseline methods using feature-based ClinicalBioBERT and BERT-base models on MIMIC and Bioasq datasets in terms of macro and micro F1-score.

	MIMIC		Bioasq	
	macro-F1	μ -F1	macro-F1	μ -F1
ClinicalBioBERT Classifier [3]	65.70	72.52	76.60	86.12
ClinicalBioBERT _{SimCSE} Classifier [7]	65.38	70.89	76.62	85.99
ClinicalBioBERT _{ESimCSE} Classifier [23]	66.35	72.64	76.97	86.32
ClinicalBioBERT _{SCL} Classifier	66.92	72.68	77.35	86.64
BERT Classifier [5]	62.88	70.82	77.00	86.16
BERT _{SimCSE} Classifier [7]	62.45	70.38	76.67	85.94
BERT _{ESimCSE} Classifier [23]	64.04	70.96	76.84	85.91
BERT _{SCL} Classifier	64.77	71.24	77.49	86.26

Table 3. End-to-end classification performance of the proposed model against baseline methods using ClinicalBioBERT and BERT-base models on MIMIC and Bioasq datasets in terms of macro and micro F1-score.

Furthermore, the final rows of Table 3 present the results of fine-tuning the pre-trained BERT-base model. It is evident from the results that fine-tuning the contrastively pre-trained BERT model based on the SCL objective yields superior outcomes compared to the BERT classifier and BERT SimCSE Classifier on both the MIMIC and Bioasq datasets. This demonstrates that the contrastive learning task accelerates the fine-tuning process by learning high-quality document representations and facilitates the learning of a superior model.

5 Conclusion

We introduced a self-supervised segment contrastive learning (SCL) approach to learn document representation for medical documents. The proposed method randomly divides documents into anchor and positive segments, then contrastively fine-tune the pre-trained language models to maximize agreement between these

segments. Through evaluations, we assessed the performance of the document embeddings against state-of-the-art baselines on different medical classification datasets, including MIMIC and Bioasq. Overall, our findings highlight the potential of SCL for refining pre-trained language models and addressing challenges in document representation learning. In future work, we plan to evaluate our method for complex information retrieval from biomedical documents.

Acknowledgments

This work was supported by ANR-22-CE23-0002 ERIANA and ANR Chaire IA Responsable.

References

1. Abacha, A.B., Yim, W.w., Michalopoulos, G., Lin, T.: An investigation of evaluation methods in automatic medical note generation. In: ACL. pp. 2575–2588 (2023)
2. Abro, W.A., Aicher, A., Rach, N., Ultes, S., Minker, W., Qi, G.: Natural language understanding for argumentative dialogue systems in the opinion building domain. *Knowledge-Based Systems* **242**, 108318 (2022)
3. Alsentzer, E., Murphy, J.R., Boag, W., Weng, W.H., Jin, D., Naumann, T., McDermott, M.: Publicly available clinical bert embeddings. arXiv preprint arXiv:1904.03323 (2019)
4. Clark, K., Luong, M.T., Le, Q.V., Manning, C.D.: Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555 (2020)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL. pp. 4171–4186 (2019)
6. Ethayarajh, K.: How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) EMNLP-IJCNLP. pp. 55–65 (2019)
7. Gao, T., Yao, X., Chen, D.: SimCSE: Simple contrastive learning of sentence embeddings. In: Moens, M.F., Huang, X., Specia, L., Yih, S.W.t. (eds.) EMNLP. pp. 6894–6910 (2021)
8. Henderson, M., Al-Rfou, R., Strophe, B., Sung, Y.H., Lukács, L., Guo, R., Kumar, S., Miklos, B., Kurzweil, R.: Efficient natural language response suggestion for smart reply. arXiv preprint arXiv:1705.00652 (2017)
9. Johnson, A.E.W., Pollard, T.J., Shen, L., Wei, H., Lehman, L., Feng, M., Ghassemi, M.M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: MIMIC-III, a freely accessible critical care database. *Scientific Data* **3** (2016), <https://api.semanticscholar.org/CorpusID:33285731>
10. Joseph, S., Kazanas, K., Reina, K., Ramanathan, V.J., Xu, W., Wallace, B.C., Li, J.J.: Multilingual simplification of medical texts. arXiv preprint arXiv:2305.12532 (2023)
11. Nguyen, H.T., Nie, D., Badamdorj, T., Liu, Y., Zhu, Y., Truong, J., Cheng, L.: Automated generation of accurate & fluent medical x-ray reports. arXiv preprint arXiv:2108.12126 (2021)

12. Pang, R.Y., Parrish, A., Joshi, N., Nangia, N., Phang, J., Chen, A., Padmakumar, V., Ma, J., Thompson, J., He, H., Bowman, S.: QuALITY: Question answering with long input texts, yes! In: Carpuat, M., de Marneffe, M.C., Meza Ruiz, I.V. (eds.) NAACL. pp. 5336–5358 (2022)
13. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al.: Improving language understanding by generative pre-training (2018)
14. Roy, A., Pan, S.: Incorporating medical knowledge in bert for clinical relation extraction. In: EMNLP. pp. 5357–5366 (2021)
15. Sansone, C., Sperli, G.: Legal information retrieval systems: State-of-the-art and open issues. *Information Systems* **106**, 101967 (2022)
16. Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., et al.: An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. *BMC bioinformatics* **16**(1), 1–28 (2015)
17. Wang, B., Xie, Q., Pei, J., Chen, Z., Tiwari, P., Li, Z., Fu, J.: Pre-trained language models in biomedical domain: A systematic survey. *ACM Computing Surveys* **56**(3), 1–52 (2023)
18. Wang, H., Li, Y., Huang, Z., Dou, Y., Kong, L., Shao, J.: Sncse: Contrastive learning for unsupervised sentence embedding with soft negative samples. *CoRR abs/2201.05979* (2022)
19. Wang, S., Manning, C.: Baselines and bigrams: Simple, good sentiment and topic classification. In: Li, H., Lin, C.Y., Osborne, M., Lee, G.G., Park, J.C. (eds.) ACL. pp. 90–94 (2012)
20. Wang, S., Liu, Z., Peng, B.: A self-training framework for automated medical report generation. In: EMNLP. pp. 16443–16449 (2023)
21. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. *arXiv preprint arXiv:2210.10163* (2022)
22. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Brew, J.: Huggingface’s transformers: State-of-the-art natural language processing. *ArXiv abs/1910.03771* (2019)
23. Wu, X., Gao, C., Zang, L., Han, J., Wang, Z., Hu, S.: ESIMCSE: Enhanced sample building method for contrastive learning of unsupervised sentence embedding. In: Calzolari, N., Huang, C.R., Kim, H., Pustejovsky, J., Wanner, L., Choi, K.S., Ryu, P.M., Chen, H.H., Donatelli, L., Ji, H., Kurohashi, S., Paggio, P., Xue, N., Kim, S., Hahm, Y., He, Z., Lee, T.K., Santus, E., Bond, F., Na, S.H. (eds.) COLING. pp. 3898–3907 (2022)
24. Yan, Y., Li, R., Wang, S., Zhang, F., Wu, W., Xu, W.: ConSERT: A contrastive framework for self-supervised sentence representation transfer. In: Zong, C., Xia, F., Li, W., Navigli, R. (eds.) ACL. pp. 5065–5075. Online (2021)
25. Yang, B., Wu, L.: How to leverage multimodal ehr data for better medical predictions? *arXiv preprint arXiv:2110.15763* (2021)
26. Yun, H.S., Marshall, I.J., Trikalinos, T., Wallace, B.C.: Appraising the potential uses and harms of llms for medical systematic reviews. *arXiv preprint arXiv:2305.11828* (2023)
27. Zhang, N., Zhang, Y., Guo, W., Mitra, P., Zhang, R.: Famesumm: Investigating and improving faithfulness of medical summarization. *arXiv preprint arXiv:2311.02271* (2023)
28. Zhou, K., Zhang, B., Zhao, X., Wen, J.R.: Debaised contrastive learning of unsupervised sentence representations. In: Muresan, S., Nakov, P., Villavicencio, A. (eds.) ACL. pp. 6120–6130 (2022)