SiaKey: A Method for Improving Few-shot Learning with Clinical Domain Information

Zhuochun Li
School of Computing and Information
University of Pittsburgh
Pittsburgh, USA
zhl163@pitt.edu

Khushboo Thaker
School of Computing and Information
University of Pittsburgh
Pittsburgh, USA
k.thaker@pitt.edu

Daqing He
School of Computing and Information
University of Pittsburgh
Pittsburgh, USA
dah44@pitt.edu

***Abstract*—Supervised Natural Language Processing (NLP) models can achieve high accuracy, but they often require a significant amount of annotated data for training, which can be expensive and time-consuming. This is especially true for clinical NLP, where annotations on large-scale electronic health records (EHRs) and online posts necessitate specialists with clinical expertise. On the other hand, fine-tuning Pretrained Language Models (PLMs) may yield poor performance due to limited training data. Few-Shot Learning (FSL) methods offer a promising solution as they can significantly improve clinical NLP with only a small amount of labeled data. In this paper, we introduce a novel FSL technique named SiaKey, which utilizes Siamese Networks and integrates Keyphrases Extraction and Domain Knowledge, for the task of online post classification. This task is challenging since online posts typically contain a greater amount of irrelevant information compared to traditional EHRs. By incorporating Keyphrases using domain knowledge, we extract essential information and reduce distractions, enhancing the classification process. To evaluate SiaKey's performance, we conducted tests with 5, 10, 15, and 20-shot learning on health-related online post-classification tasks. The results of our experiments demonstrate SiaKey's effectiveness in capturing text features, showcasing its superior performance compared to BioBERT on similar FSL tasks.**

***Index Terms*—few-shot learning, keyphrases extraction, clinical domain knowledge, social media**

I. Introduction

The abundant annotated data is considered the necessary element to train deep neural networks for clinical NLP. However, the annotation process could be very expensive due to the huge amount of related data and the high demand for professional clinical knowledge [1]. Even though there are several high quality and publicly available clinical datasets, such as i2b2 datasets [2], MIMIC-III datasets [3], and BioNLP datasets [4], they are not equipped to handle the online noisy health data generated by patients through blogs, posts, and comments. Considering the fact that state-of-the-art supervised deep learning neural networks always have poor performance when the training data is in shortage, implementing a clinical NLP system with few or no annotated data becomes the focus in clinical informatics research.

The fine-tuning approach is applied to solve the traditional neural networks' disadvantage on NLP tasks. Pre-trained Language Models (PLMs) [5] are deep neural network models trained on unlabeled large-scale datasets, such as Wikipedia or PubMed data. This training process is called pre-training, which always takes a long time and vast amount of computational resources [6]. Once this pre-training process is completed, these language models always have a general ability for NLP tasks. Subsequently, for each new task, it is necessary to annotate related data to retrain the models [7]. The process that retraining the PLMs on a task-specific labeled dataset is called fine-tuning. For example, the state-of-the-art medical and clinical PLMs include BioBERT [8] and Clinical BERT [9], which have been trained on millions of EHRs and unlabeled clinical text datasets like MIMIC. This pre-training process enables them to learn general medical and clinical linguistic characteristics. Later, this general knowledge could be transferred for specific downstream tasks like biology text mining or clinical Named Entity Recognition (NER) [10] tasks, by fine-tuning the PLMs with a lesser amount of task-specific annotated data.

Unfortunately, the fine-tuning PLMs procedure also requires a considerable amount of data and has unsatisfied performance on FSL. Thus, the FSL approach with a small amount of annotated data provides a solution to these scenarios. With the aid of prior knowledge, FSL can quickly adapt to new tasks with only a limited number of samples that contain supervised information [11]. A Siamese neural network is an artificial neural network that employs the same weights to operate in parallel on two distinct input vectors, producing comparable output vectors [12]. This network was initially used for computer vision tasks, but the same idea could be extended to text classification. To the best of our knowledge, the utilization of Siamese Neural Networks (SNN) on noisy forum text data created by patients and caregivers has not been explored. This model can be trained to compute feature vectors, which can subsequently be employed for a variety of clinical NLP tasks, including Named-Entity Recognition (NER) and Sentiment Analysis.

Domain knowledge is the understanding of a specific industry, discipline or activity [13]. However, it is often overlooked in clinical NLP studies. Related works tend to solely focus on collected contents of EHRs that contain symptom descriptions and diagnoses from health specialists. We suggest that online health-related posts on social media should also be considered, as they may contain patient emotions, feelings, and other contextual words. Incorporating this information is much more challenging than the content found in EHRs. We utilized keyphrase extraction to retrieve the most critical information.

This paper investigates the use of Few-Shot Learning (FSL) techniques to address the lack of annotated datasets for clinical NLP tasks, specifically the task of online health-related posts text classification. We propose the innovative SiaKey system, which combines post titles, contents, and keyphrases extracted as medical entities from the post, as model input. With the aid of related domain knowledge, we extracted medical entities and utilized titles as summarizations of the posts. We then conduct the FSL procedure on Siamese Networks. Our results demonstrate that this novel method retrieves critical clinical information from noisy post contents and achieves better performance than BioBERT on similar FSL tasks.

II. Related Work

A. Pre-trained Language Models

Various language model architectures have been proposed to solve different NLP tasks. The popular first-generation PLMs incorporate GloVe [14] and Word2Vec [15]. The features of first-generation PLMs are word embedding based on occurrences of different words in documents. Although these models are efficient in comprehending semantic information, they usually ignored the linguistic meaning of words and the underlying contexts behind the embeddings. The second-generation PLMs are BERT [16], GPT-2 [17], and T5 [18]. The main improvement of these PLMs is that they take the context information into consideration, which could reveal and understand complicated concepts of words. Consequently, these models have achieved state-of-the-art performance for diverse NLP tasks.

B. Siamese Networks in Few-Shot Learning

There have been studies assessing the effectiveness of Siamese Networks (SNN) for image classification. Zhang [19] used SNN to capture the spatial information for object tracking tasks via multiscale spatial attentions. And Hunt [20] applied SNN for the classification of electrograms. In the context of FSL, SNN has been used by Koch [21] for one-shot image recognition, which was based on the convolutional architecture to retrieve discriminative features from only one single example of each new class. Droghini [22] employed SNN for few-shot human-fall detection objectives by using audio signals. Nevertheless, none of these studies used SNN in the NLP field.

A recent study by Oniani [23] explored SNN for FSL in Clinical NLP and demonstrated good performance on Text Classification and Named-Entity Recognition tasks. However, their dataset only contained clinical narratives from professional experts, which is much more straightforward and does not need domain knowledge such as keyphrase extraction to abstract keywords from noisy inputs.

To the best of our knowledge, none of the studies referenced have utilized Siamese Networks in the FSL clinical NLP field. Therefore, our work fills the gap in the application of Siamese Networks in the complex clinical NLP field. Specifically, our method provides a solution for noisy clinical text generated on social platforms classification tasks and demonstrates the efficiency of incorporating titles and keyphrases.

TABLE I
Samples of our dataset.

Title	Body	Label
Anybody else feel like crap when they switch from Kpins to Valium?	I switched last week to Valium to start my taper off a 4 year use of 2m per day k use and I feel like holy hell.	Addicted
One month off suboxone.	Hello everyone. One month free and clear. Feels good man!	E-Recovery
9 months sober, replacing one addiction with another.	I've been sober for almost 9 months now. When I stopped using substances, I began to binge eat.	M-Recovery
Today Marks my 20th year of Recovery.	I just want to say recovery is possible even for a low bottom junkie like me. Good luck to you.	A-Recovery
Help	Somebody calm me please.	Others

III. Datasets

For our experiments, we consider the Drug Abuse Data - Reddit Dataset [24], which was published in the paper "Utilizing Social Media for Identifying Drug Addiction and Recovery Intervention" in 2020. The dataset contains posts from drug addiction-related Subreddits which are topic-specific communities within the Reddit online social media (https://www.reddit.com/). Therefore, the posts consist of a significant amount of noisy text, enabling us to thoroughly test the ability of SiaKey to capture critical features from them. It annotated 3151 posts out of all the posts collected as one of the 5 classes (Stage of addiction): 'Addicted', 'E(early)-Recovery', 'M(maintaining)-Recovery', 'A(advanced)-Recovery', 'Others'. In our NLP classification task, we used the 'title' (Each post is associated with a title), 'body'(The main descriptive part of the post), and 'label_classification'(The label given to a post for the classification task) three fields.

To test our FSL methods, we randomly picked 200 samples for each of the five classes to build our test set with 1000 samples in total. In addition, we randomly selected 20 samples for each of the five classes out of the test datasets to build our FSL training sets. Specifically, we conducted 5-shot(S_5), 10-shot(S_{10}), 15-shot(S_{15}) and 20-shot(S_{20}) training process. The relationship between these training sets is: $S_5 \subset S_{10} \subset S_{15} \subset S_{20}$. Therefore, the larger shots training set always contains fewer shots training sets, which guarantees the fairness of comparing different shots training results. Finally, to evaluate the supervised training of BioBERT as the upper bound, we utilized all available samples except for the test cases, resulting in a total of 2151 samples (3151-1000=2151). Samples of our dataset are shown in Table I.

IV. Methods

In this section, we first present how we build the Siamese Networks on the FSL of NLP. Then, we propose a keyphrase extraction method to improve the few-shot performance.

A. Sentence Embeddings

To meet the speed and efficiency requirements of FSL, we utilized the universal-sentence-encoder [25], which is a sentence encoding module of TensorFlow-hub. This encoder is capable of encoding sentences into high-dimensional embeddings that can be used for various NLP tasks, such as semantic similarity and text classification. The length of the resulting embedding vector is always 512, regardless of the input's length. By leveraging transfer learning, we employed this pre-trained universal sentence encoder to obtain a more robust representation of sentences.

B. Siamese Networks Architecture

The primary concept of Siamese Networks is to compute the triple loss between anchor (A), positive (P), and negative (N) input text as the loss function. In a triplet, the anchor and positive input text belong to the same output class, while the negative input text belongs to a different output class. Once we obtain the sentence embeddings from the universal sentence encoder, we preprocess these embeddings using dense and normalization layers before feeding them into the final triple loss layer. This preprocessing step helps decrease the variance and dimensionality of input embeddings. Fig. 1 illustrates our entire system architecture. To calculate the triple loss, we aim to project the embeddings such that the distance between the anchor and negative samples, $d(A_i, N_i)$, is α greater than the distance between the anchor and positive samples, $d(A_i, P_i)$. We define α as the margin point because we typically define non-negative loss functions in neural networks. The loss will be zero if the difference between $d(A_i, N_i)$ and $d(A_i, P_i)$ is greater than the margin. Otherwise, the difference in distance is considered as the triplet loss, which is back-propagated through the entire Siamese Network. The mathematical definition of the triple loss is shown below.

$$L(A,P,N) = \frac{1}{N}(\sum_{i=1}^{N} max(d(A_i,P_i) - d(A_i,N_i) + \alpha, 0) \quad (1)$$

where $d(X_i, Y_i) = ||\vec{x_i} - \vec{y_i}||_2^2$

C. Domain Knowledge

Upon closer analysis of the posts in each class, we discovered that titles contained valuable information and acted as a summary or hint for the subsequent post content. Consequently, the function of the title is similar to manual keyphrase extraction, as it is derived from the patient's own understanding and can aid our model in extracting critical features from the post body. Ultimately, we decided to integrate the title as a manual keyphrase extraction method and medical entity as an automatic extraction method in our experiments.

1) Medical Entity Extraction: In our health-related tasks, we consider keyphrases as critical medical entities in text. For medical entity extraction, we utilized Scispacy [26], which extracts the entities in the text and links them with the Unified Medical Language System (UMLS) [1], an official medical meta-thesaurus. However, UMLS contains both medical as well as non-medical terms for example emotions, body anatomy, and activities. As we are interested in only health or medical-related terms mentioned in the post by the user, we filtered these entities based on their semantic types. The entities related to drug, disease, treatment procedures, and diagnosis. Scispacy provides all the entities in the text, but for our purpose, we only wanted top N entities that are more representative of the post. So we ranked the entities based on EmbedRank [27] automatically and utilized the top N entities which have more similarity with the user-generated text. For this purpose, we utilized BioBERT embeddings [8]. As a service tool, this API can be accessed at: http://ngrok.luozm.me:8395/keyphrase/kweigt

[1]https://uts.nlm.nih.gov/uts/umls/home

V. Experiments and Results

To evaluate the effectiveness of our SiaKey system, we conducted text classification tasks on the Drug Abuse Dataset. Initially, we fine-tuned BioBERT on all available samples except for the test set to establish an upper bound for our experiments. Subsequently, we explored the feasibility of combining the title with the text as the input and compared their results. Finally, we trained model using different combinations of title, text, and keyphrases (extracted from the title, text, or both) to analyze their performance. The method of combining the title, text, and keyphrases was a simple concatenation.

After extracting text features from our SiaKey system, we employed the simple K-nearest neighbors (KNN) algorithm with K=5 to accomplish the downstream task of text classification. The FSL was done on 1 GPU (NVIDIA gtx1080), with a batch size of 32. We trained the model for 50 epochs with 10 steps for each epoch. The optimizer is "Adam" with a learning rate of $1\mathrm{e}{-3}$. Although we used the metrics Precision, Recall, and F1 score of all five classes to evaluate the text classification task, we decided to show the final performance via Accuracy and Macro-Averaged F1 metrics due to the number of results.

A. Upper Bound Experiments

BioBERT was employed as the upper bound for fine-tuning the PLMs method, which was trained on all 2151 available samples except the test cases in our dataset. Considering that our FSL method only trains a small portion of all samples in a shorter time, this full training result functions as an upper bound for the subsequent FSL evaluations. Table II shows the performance of fine-tuning on sufficient data volumes.

TABLE II
BioBERT fine-tuning.

Model	Accuracy	Average-F1
BioBERT	80.83%	80.74%

B. FSL Experiments with Title and Text

Our SiaKey system was trained on 5-shot, 10-shot, 15-shot, and 20-shot FSL processes using various input combinations, including title, text, and title+text. We also compared the performance of our system with BioBERT in each input combination, and the results are presented in Table III.

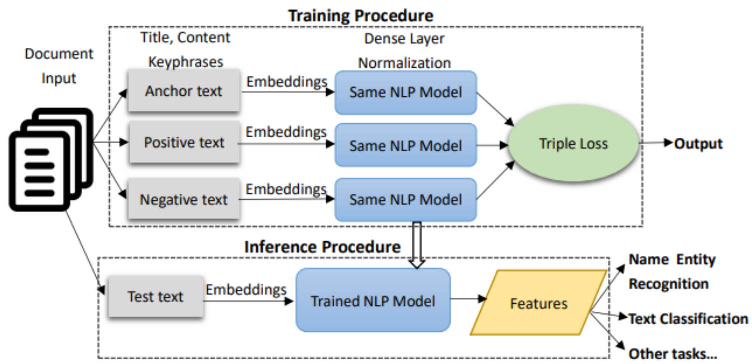

Fig. 1. Overview of our SiaKey system. The NLP Model is a neural network with simple architecture: Input(512), dense_layer1(256), dropout_layer1(256), batch normalization layer(256), dense_layer2(64), dropout_layer2(64), dense_layer3(128), norm_layer(128).

TABLE III
FSL EXPERIMENTS WITH TITLE AND TEXT.

Model-Shots	Input	Accuracy	Average-F1
SiaKey-5	Title	46.80%	46.31%
BioBERT-5	Title	18.61%	12.10%
SiaKey-5	Text	45.50%	45.32%
BioBERT-5	Text	15.45%	10.24%
SiaKey-5	Title+Text	**49.10%**	**48.63%**
BioBERT-5	Title+Text	21.33%	13.71%
SiaKey-10	Title	52.50%	52.01%
BioBERT-10	Title	42.63%	43.08%
SiaKey-10	Text	50.40%	50.51%
BioBERT-10	Text	39.87%	41.14%
SiaKey-10	Title+Text	**53.20%**	**53.23%**
BioBERT-10	Title+Text	45.67%	46.74%
SiaKey-15	Title	59.30%	59.18%
BioBERT-15	Title	48.06%	47.15%
SiaKey-15	Text	51.40%	51.70%
BioBERT-15	Text	44.13%	43.52%
SiaKey-15	Title+Text	**59.80%**	**59.72%**
BioBERT-15	Title+Text	49.17%	48.57%
SiaKey-20	Title	**64.50%**	**64.60%**
BioBERT-20	Title	61.18%	61.53%
SiaKey-20	Text	53.10%	52.99%
BioBERT-20	Text	52.23%	51.87%
SiaKey-20	Title+Text	59.60%	59.52%
BioBERT-20	Title+Text	57.83%	58.30%

TABLE IV
SIAKEY SYSTEM EXPERIMENTS WITH TITLE, TEXT, AND KEYPHRASES.

Shots	Input	Accuracy	Average-F1
5	Title	46.80%	46.31%
5	Title+Keyphrases(Text)	47.90%	46.44%
5	Text+Keyphrases(Text)	45.30%	44.75%
5	Title+Text	49.10%	48.63%
5	Title+Text+Keyphrases(Title)	50.50%	50.80%
5	Title+Text+Keyphrases(Text)	50.00%	49.64%
5	Title+Text+Keyphrases(Both)	**51.70%**	**51.54%**
10	Title	52.50%	52.01%
10	Title+Keyphrases(Text)	51.90%	50.59%
10	Text+Keyphrases(Text)	50.20%	49.99%
10	Title+Text	53.20%	53.23%
10	Title+Text+Keyphrases(Title)	50.90%	53.59%
10	Title+Text+Keyphrases(Text)	54.00%	52.48%
10	Title+Text+Keyphrases(Both)	**54.50%**	**54.20%**
15	Title	59.30%	59.18%
15	Title+Keyphrases(Text)	55.60%	55.44%
15	Text+Keyphrases(Text)	51.30%	51.32%
15	Title+Text	59.80%	59.72%
15	Title+Text+Keyphrases(Title)	59.90%	59.03%
15	Title+Text+Keyphrases(Text)	59.30%	58.50%
15	Title+Text+Keyphrases(Both)	**60.30%**	**59.75%**
20	Title	**64.50%**	**64.60%**
20	Title+Keyphrases(Text)	53.10%	53.19%
20	Text+Keyphrases(Text)	53.10%	53.19%
20	Title+Text	59.60%	59.52%
20	Title+Text+Keyphrases(Title)	61.00%	60.97%
20	Title+Text+Keyphrases(Text)	60.50%	60.48%
20	Title+Text+Keyphrases(Both)	60.50%	60.11%

C. SiaKey Experiments with Title, Text, and Keyphrases

This section examines the performance of our SiaKey system with various combinations of input text, including title, text, and keyphrases, using 5-shot to 20-shot training sets. Specifically, keyphrases are extracted from the title, text, or both. To facilitate comparison, we only include the training input of ”Title” and ”Title+Text” in Table IV, with Table III providing additional information. Using the title as input can reduce training time by 28% compared to the method using text alone. Moreover, incorporating domain knowledge by adding keyphrases barely increases the training time.

VI. CONCLUSIONS

In this paper, we explored the integration of domain knowledge into the FSL framework and present the Siakey system. We conducted several text classification experiments to evaluate our model's performance, focusing on combinations of the post title, text, and keyphrases for better feature extraction. Our results demonstrate the usefulness of our Siakey system in FSL tasks and its effectiveness in extracting critical information from noisy text data. Our work has significant implications as we have introduced an innovative and efficient method to automatically perform text identification tasks with only a few training samples. This approach can greatly assist healthcare systems by efficiently classifying patients into disease trajectories and providing timely resources for their care. Furthermore, the advantage of feature extraction extends its usability to a wide range of NLP tasks.

However, there are also limitations to our work. First, only the BioBERT model is compared in our experiments, future work could introduce comparisons between more PLMs. Second, the lack of public datasets for patient-generated social media data [28] makes it difficult to compare our work to other FSL strategies on benchmark datasets. Thus, comparing systems and benchmarking is one of the directions for future studies in FSL.

REFERENCES

[1] S. M. Yimam, C. Biemann, L. Majnaric, Š. Šabanović, and A. Holzinger, "An adaptive annotation approach for biomedical entity and relation recognition," *Brain Informatics*, vol. 3, no. 3, pp. 157–168, 2016.

[2] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2010 i2b2/va challenge on concepts, assertions, and relations in clinical text," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 552–556, 2011.

[3] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[4] C. Nédellec, R. Bossy, J.-D. Kim, J.-J. Kim, T. Ohta, S. Pyysalo, and P. Zweigenbaum, "Overview of bionlp shared task 2013," in *Proceedings of the BioNLP shared task 2013 workshop*, 2013, pp. 1–7.

[5] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, 2020.

[6] Y. Tian, Y. Wan, L. Lyu, D. Yao, H. Jin, and L. Sun, "Fedbert: when federated learning meets pre-training," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 4, pp. 1–26, 2022.

[7] H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and T. Zhao, "Smart: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization," *arXiv preprint arXiv:1911.03437*, 2019.

[8] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, p. 1234, 2020.

[9] E. Alsentzer, J. R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, and M. McDermott, "Publicly available clinical bert embeddings," *arXiv preprint arXiv:1904.03323*, 2019.

[10] C. Sun and Z. Yang, "Transfer learning in biomedical named entity recognition: an evaluation of bert in the pharmaconer task," in *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, 2019, pp. 100–104.

[11] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM computing surveys (csur)*, vol. 53, no. 3, pp. 1–34, 2020.

[12] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[13] T. Dash, S. Chitlangia, A. Ahuja, and A. Srinivasan, "A review of some techniques for inclusion of domain-knowledge into deep neural networks," *Scientific Reports*, vol. 12, no. 1, p. 1040, 2022.

[14] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[15] Y. Goldberg and O. Levy, "word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[17] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[18] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer." *J. Mach. Learn. Res.*, vol. 21, no. 140, pp. 1–67, 2020.

[19] J. Zhang, B. Huang, Z. Ye, L.-D. Kuang, and X. Ning, "Siamese anchor-free object tracking with multiscale spatial attentions," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[20] B. Hunt, E. Kwan, D. Dosdall, R. S. MacLeod, and R. Ranjan, "Siamese neural networks for small dataset classification of electrograms," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.

[21] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2. Lille, 2015, p. 0.

[22] D. Droghini, F. Vesperini, E. Principi, S. Squartini, and F. Piazza, "Few-shot siamese neural networks employing audio features for human-fall detection," in *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*, 2018, pp. 63–69.

[23] D. Oniani, S. Sivarajkumar, and Y. Wang, "Few-shot learning for clinical natural language processing using siamese neural networks," *arXiv preprint arXiv:2208.14923*, 2022.

[24] S. Ghosh, J. Misra, S. Ghosh, and S. Podder, "Utilizing social media for identifying drug addiction and recovery intervention," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 3413–3422.

[25] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar *et al.*, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.

[26] M. Neumann, D. King, I. Beltagy, and W. Ammar, "ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing," in *Proceedings of the 18th BioNLP Workshop and Shared Task*. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 319–327. [Online]. Available: https://www.aclweb.org/anthology/W19-5034

[27] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl, and M. Jaggi, "Simple unsupervised keyphrase extraction using sentence embeddings," in *Proceedings of the 22nd Conference on Computational Natural Language Learning*, 2018, pp. 221–229.

[28] Y. Ge, Y. Guo, Y.-C. Yang, M. A. Al-Garadi, and A. Sarker, "Few-shot learning for medical text: A systematic review," *arXiv preprint arXiv:2204.14081*, 2022.