ENHANCING ONCOLOGY CARE WITH FEDERATED LEARNING AND FOUNDATION MODELS

Gagan, N[1] and *Sanand, Sasidharan*[2]

[1]GE HealthCare

ABSTRACT

Millions of people worldwide are battling cancer, and personalised care plans are essential for effective diagnosis, treatment, and monitoring of this disease. Recently, Large Language Models (LLMs) have proven valuable in cancer treatment, for instance, extracting key information from Electronic Medical Records (EMRs). This study presents a transformer encoder based LLM, that is domain adapted for Oncology, and outperforms generic models in recognising critical oncology related elements from clinical text. We observe that the development of such domain specific LLMs demands a huge amount of data and computational resources, which is a deterrent to the sustainability development goal of equitable health. To address this problem, we propose a federated learning approach for model development that will eliminate data sharing and centralised computational resource costs. Our evaluations show that the federated approach outperforms the generic base model, highlighting the advantages of collaborative learning in capturing domain specific knowledge and enhancing performance in oncology related NLP tasks. Our work is in line with the United Nations Sustainable Development Goals (SDGs) which are aimed at promoting equitable health and narrowing down the differences in access to advanced cancer treatment.

Keywords - Pre-tuning, Domain Adaptation, Federated Learning, NER, Fine-tuning, BERT, SDG's, Embedding

1. INTRODUCTION

Cancer is a global health issue affecting millions of people worldwide. The recent progress in artificial intelligence (AI) and machine learning (ML) has proved to be very effective for oncology care by means of data-driven insights and decision support systems [1]. Natural Language Processing (NLP), particularly Named Entity Recognition (NER), is a very useful tool in oncology care. It identifies and extracts the vital elements like cancer types, treatments, drugs, etc. from unstructured medical texts such as clinical notes and pathology reports. NER assists oncologists in these tasks so that they can efficiently access patients key information in a very short time, which leads to better diagnosis, treatment planning, and overall patient management.

The transformer based language models such as BERT have proven to be very effective in different NLP tasks [2]. However, the performance of these systems in specialised fields like oncology is not optimal because of the domain specific nature of medical terminologies and concepts. The domain adaptation methods address this issue. BioBERT [3] is an example of a biomedical domain adapted BERT model, that has demonstrated better performance in NLP tasks within this domain compared to the original BERT model. Nevertheless, acquiring good quality medical data is quite difficult due to privacy, data governance, and the difficulties in handling sensitive patient data.

Federated learning which is a privacy-preserving ML technique, has been created as the possible solution that can help in collaborative model training across decentralised data sources without actually sharing the data [4]. Particularly in the area of sensitive data like healthcare, federated learning has a lot of benefits over traditional centralised ML methods. Federated learning allows the collaborative sharing of model training among decentralised data sources without compromising data privacy, which is a very important issue as it involves data governance, computational resource limitations, and models. Our method relies on the combined knowledge of many healthcare facilities thus, direct data sharing is not needed. Also, federated methods are essential to eliminating biases in AI models towards developed demographics by addressing the two major barriers to model development: data exchange and computational facilities.

This study develops a method that unites transformer based language models, domain adaptation, and federated learning to improve oncology care. We introduce a language model specifically designed for the oncology domain, which outperforms generic models in NLP tasks related to oncology. Our evaluation was mainly based on the NER task as a primary metric, which showed the model's capacity to recognise and extract significant entities in the oncology field. Nevertheless, the scope of this oncology-specific foundational model is not limited to NER, it can be modified for other downstream tasks like relation extraction, text classification, and text generation, thus paving the way for further development in oncology text mining and analysis. To overcome the issues related to data sharing and the lack of computational resources, we propose a federated learning based model to perform collaborative model building without compromising data privacy.

Our work is in line with the United Nations Sustainable Development Goals (SDGs), particularly SDG 3 (Good Health and Well being), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 10 (Reduced Inequalities) as we promote balanced health and limit the disparity in the availability of advanced oncology care. The enhanced performance of our domain adaptable and federated models in capturing oncology specific language and knowledge will conversely help in various aspects of oncology care, propagating the delivery of personalised, equitable, and high quality oncology care to all patients across the demographics.

Even though the strategies mentioned in this paper are designed for encoder based transformers, they can also be applied to decoder based transformer architectures, thus laying a foundation for further exploration and application in various natural language processing tasks. The rest of the paper is organised as follows: Section 2 discusses related work, Section 3 discusses the methodology, Section 4 presents the Results, and Section 5 concludes the paper.

2. RELATED WORK

This section provides an overview of relevant literature and research advancements in two key areas: Domain adaptation of transformer-based language models and federated learning. We underline the progress made in these areas, and at the same time, we point out the gaps and opportunities that motivated our work.

2.1 Domain Adaptation for Healthcare Applications

The performance of transformer based language models like BERT [2], in natural language processing has motivated researchers to consider their application in many fields, including healthcare. Nevertheless, the complexity of medical terminology and concepts creates difficulties due to model application in the context of domain adaptation. Many studies have investigated domain adaptation techniques to enhance the performance of pre-trained language models in the biomedical and clinical fields. This can be seen in ClinicalBERT [5], which was fine-tuned on clinical notes from the MIMIC-III dataset [6] and did better at clinical natural language inference and relation. BlueBERT [7] has been fine-tuned on electronic health records (EHRs) and showed much better performance than BioBERT [3] and other baselines in clinical named entity recognition and relation extraction tasks. Among other domain adapted models are PubMedBERT [8], fine-tuned on PubMed abstracts and full-text articles, and SciBERT [9], which is fine-tuned on a large corpus of scientific literature. In [10] Zhang Et.al trained BERT on Chinese medical diagnostic and treatment texts. Liu Et.al have proposed Med-BERT [11], medical dictionary enhanced BERT model. These models performed better in biomedical information extraction, text classification, and question answering tasks. Although, these domain adapted models have demonstrated potential in their specific medical domains, their suitability for special areas such as oncology is quite restricted. Terms, concepts, and contexts specific to oncology can be extremely subtle and often require special domain adaptation to help understand the specifics of cancer language and knowledge.

2.2 Federated Learning in Healthcare

Federated learning is a technique that deals with privacy and data governance problems in AI applications. It permits the collaboration of model training across different decentralised data sources without actually sharing any sensitive information, hence, it is a privacy-preserving substitute to the traditional way of centralised machine learning [4]. For instance, researchers have proposed a federated learning framework that allows collaboration among multiple medical institutions on medical image analysis tasks like finding COVID-19 using chest X-ray images [12]. These studies show the federated learning is capable of being used to create models in collaboration while at the same time, maintaining data privacy.
In the context of natural language processing, federated learning has been applied for different tasks, such as clinical entity recognition with EHR data from various healthcare institutions. These methods prove the efficiency of federated learning in destroying data silos and enhancing model accuracy by means of collaborative training. Moreover, the scientists have also examined federated transfer learning for medical relation extraction, in which pre-trained models are adjusted on distributed data sources and better performance is achieved than that of centralised training [13]. Also, medical relation extraction tasks have been solved using federated learning as well, thus proving the possibility of privacy preserving collaborative learning [14]. Even though these studies show positive outcomes, the application of federated learning for domain adaptation of transformer based language models to a specific healthcare area such as oncology remains unexplored [15].

2.3 Research Contributions

This research provides an integration approach of domain adaptation and federated learning approaches to improve oncology practice through the development of a stable, privacy preserving base model specific to oncology. Specifically, our contributions are as follows :

1. **Domain Adaptation:** We utilise a set of oncology-related datasets that encompass cancer-specific language nuances and semantics to adapt the BioBERT model for the oncology domain.

2. **Federated Learning:** We employ federated learning to address data collection and computation challenges, training models at source sites, and aggregating weights to distribute costs and maintain privacy.

3. **Extensive Evaluation:** To demonstrate the effectiveness of our approach in capturing domain specific semantics and improving oncology based NLP tasks, we perform evaluations including embedding visualisation, clustering analysis, and named entity recognition (NER) tasks.

This study intends to improve AI in oncology care through the use of transformer-based language models, domain adaptation, and federated learning methods. In this paper, we address the challenges concerning data sharing, privacy concerns, and computational resource constraints.

3. METHODOLOGY

Our approach to enhancing oncology care through federated learning and transformer based foundation models involves four key components: data processing and preparation, domain adaptation, federated learning, and comprehensive evaluation. In this section, we provide a detailed explanation of each component, along with the underlying techniques and methodologies.

3.1 Data Processing and Preparation

To create a domain-specific language model for oncology, we pre-train the BioBERT model on different kinds of datasets related to oncology using masked language modelling (MLM) and next sentence prediction (NSP) tasks. The datasets for this task include:

1. **Cancer related trials:** This dataset encompasses 100,000 cancer trial samples, providing comprehensive information on cancer clinical trials, including trial descriptions, eligibility criteria, and treatments. [1]

2. **PubMed Hallmarks of Cancer Dataset:** This dataset comprises 1,852 publication abstracts related to the hallmarks of cancer. [2]

3. **Cancer Document Classification:** This dataset consists of 7,569 cancer document samples, Research Paper Text field in this dataset was used for training. [3]

4. **Oncology Patient Medical Reports:** To further enhance the model's understanding of oncology specific language, we incorporated 19,253 anonymized medical reports belonging to cancer patients. This dataset provides valuable insights into the language and structure of clinical documentation in oncology.

The MLM task involved randomly masking 15% of the input tokens in each sentence and replacing them with the [MASK] token, without masking special tokens such as [CLS] and [SEP]. The MLM task's goal was to identify the original masked tokens from the context, which allowed the model to learn domain specific representations of the oncology [2]. The NSP task required sentence pairs to be generated by sampling consecutive sentences (positive examples) or non-consecutive sentences (negative examples) from the dataset. The NSP task made the model learn the sequential nature of oncology related texts and improved its understanding of document structure [16].
We used the BioBERT tokenizer to tokenize sentence pairs and built input tensors for the model. The tokenizer truncated or padded the sequences to a length of 512 tokens. The input tensors consisted of input IDs, token type IDs, attention masks, next sentence labels, and the labels for the MLM task. To enhance the BioBERT model, we loaded the pre-trained weights and trained it with AdamW optimizer using a learning rate of 5e-5 for 3 epochs. BioBERT was pre-trained on the oncology-specific dataset using MLM and NSP tasks which resulted in a domain-focused language model that is able to catch the subtleties and details of oncology-related language. This model can now more accurately perceive and express domain specific concepts, terms, and relationships. The MLM task enabled the model to acquire contextual representations, while the NSP task assisted in comprehension of the coherence and sequential order of oncology related text. The domain adaptation is targeted to improve the model's performance on oncology specific natural language processing tasks like named entity recognition, relation extraction, and text classification.

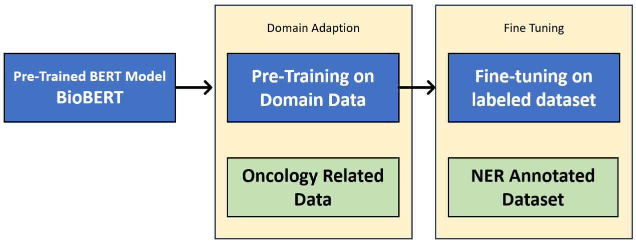


Figure 1 – BERT model training process

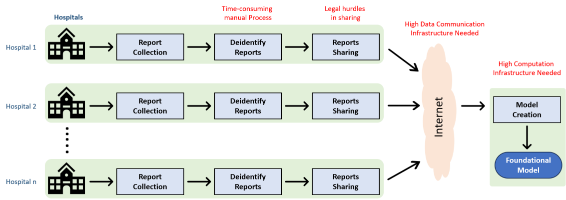
3.2 BERT Model Adaptation and Fine-tuning

The overall flow of the BERT model for domain adaptation on an oncology corpus and fine-tuning for the NER task is illustrated in Figure 2. The process starts with the pre-trained BioBERT model, which was trained on a large corpus of biomedical text. To adapt the model to the oncology domain, we do some extra pre-training on oncology-related data. This in turn helps the model acquire domain specific language patterns and vocabulary. This domain adaptation step makes the weights of the model more precise and better at capturing the nuances and characteristics of oncology. Subsequently, we fine-tune the domain adapted model on a labelled NER annotated dataset specific to the oncology domain. At this stage, we further adjust the model's weights to capture the specific patterns and features necessary for accurately identifying named entities within the oncology context. The fine-tuning process leverages the information obtained from both the general pre-training (
BioBERT) and domain specific pre-training (oncology related data). The fine-tuned model obtained in the end can be used to automatically extract and annotate named entities from new, unseen oncology text data, thus making it possible for efficient information extraction and analysis in oncology.

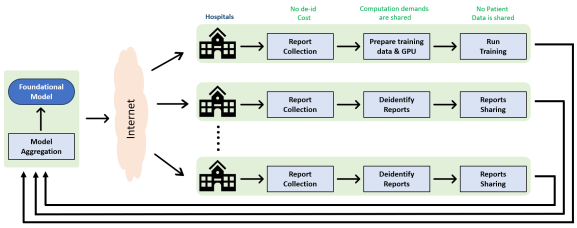
3.3 Federated Learning

To enhance the domain specific language model and address the challenges of data privacy and centralised training, we

[1]ClinicalTrials.gov
[2]huggingface.co/datasets/qanastek/HoC
[3]kaggle.com/datasets/falgunipatel19/biomedical-text-publication



(a) Conventional Approach



(b) Federated Approach

Figure 2 – Model Training Approaches

employed a federated learning approach, as illustrated in Figure 2. Earlier, machine learning models for the healthcare domain were usually created by gathering data from various hospitals and then putting it together in a central repository to make models. This centralised data set was used to build AI models. As shown in Figure 2a health care facilities have to open up their patients confidential data to an external model development repository. The storage and operation of data that is not under the control of individual hospitals have made people worry about privacy, possible breaches, and legal as well as ethical issues. This reduces the volume and diversity of the data made available for model training, which could result in biassed or less generalizable models.

Federated learning allows collaborative model training without hospitals having to reveal their patients data. As demonstrated in Figure 2b. Every hospital has its own data set and trains the local model on their site. Only model updates, like weights or gradients, are exchanged with a central server for aggregation. Federated learning addresses privacy and security issues by decentralising data and granting each hospital autonomy over its own data. This approach guarantees that confidential patient data will not be accessed by unauthorised people, and it complies with stringent regulations and standards, such as HIPAA, which governs the management of health information in the healthcare industry. Apart from that, federated learning encourages data governance and ownership as the hospitals keep their own data and make a choice of when to participate in collaborative model training. This approach encourages other hospitals to exchange their trained updates that are more diverse and reflective of the model's development. The increased diversity of data enables the development of stronger and more generalizable models that can reflect the differences in patient populations and clinical practices among different hospitals.

Moreover, federated learning optimises the computational resources, and this is effective relative to the standard approach. The common strategy is that centralised model training typically requires high computational power and storage, which are not always available in individual hospitals. Federated learning enables every hospital to use its local computational resources for local model training, eliminating the need for expensive infrastructure and allowing even hospitals with fewer resources to participate in the process.

3.4 Evaluation Strategy

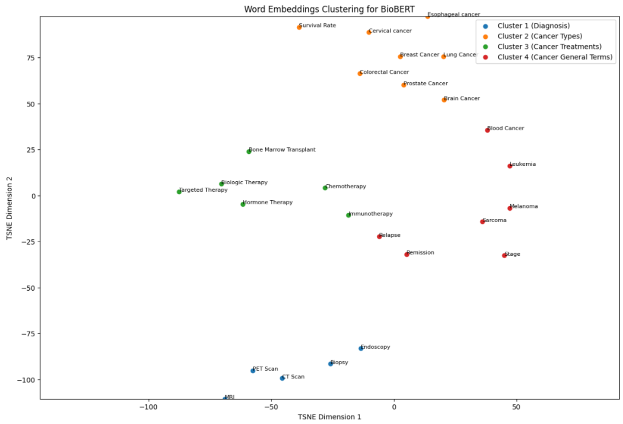
To evaluate the effectiveness of our approach and demonstrate the benefits of domain adaptation and federated learning, we conducted evaluations, including embedding visualisation, clustering analysis, and named entity recognition (NER) tasks. We assessed the domain adaptation of our pre-tuned models and BioBERT by visualising the semantic relations of the oncology related terms that the models could comprehend. We took the embeddings of the key words related to cancer treatment, diagnosis, and general terms from each model. To visualise the high dimensional word embeddings in a two dimensional space, t-SNE was used. To maintain t-SNE projection standardisation, we used the same projection martrix for all three models. We applied K-means clustering after dimensionality reduction to group similar words together based on their proximity in the reduced space. The embedding visualisations and clustering analyses gave qualitative insights into the model's capacity to represent semantic similarities. Furthermore, we also tested the models on the NER task using a manually annotated clinical reports dataset. The dataset was preprocessed, tokenized, and the NER labels were aligned with the token sequences. The models were fine-tuned to predict the labels for each token. To measure the contribution of domain adaptation and federated learning to the exact identification and classification of cancer related entities in the clinical texts, we compared the precision, recall, and F1 scores of pre-trained models and BioBERT on the NER task.

4. RESULTS AND DISCUSSION

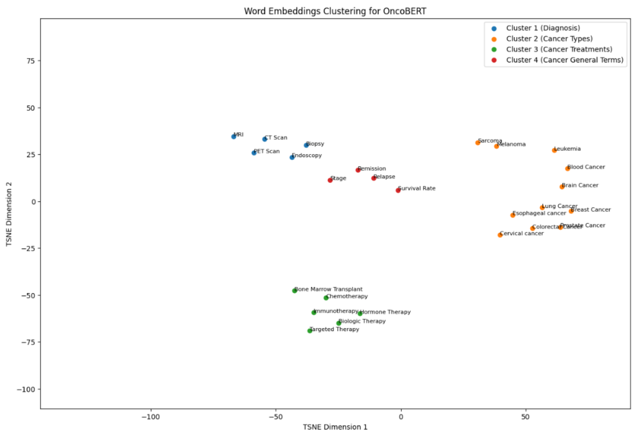
In this section, we discuss the outcomes of our evaluations, focusing on two major aspects: (1) visualisation and clustering of embedding, and (2) the named entity recognition task. The first step is to check how domain adaptation can really capture the semantics of oncology related terms by means of embedding visualisation and clustering. After that, we analyse the performance of all three models on the NER task, showing quantitative results and talking about the enhancements made by domain adaptation and federated learning.

4.1 Embedding Visualisation and Clustering

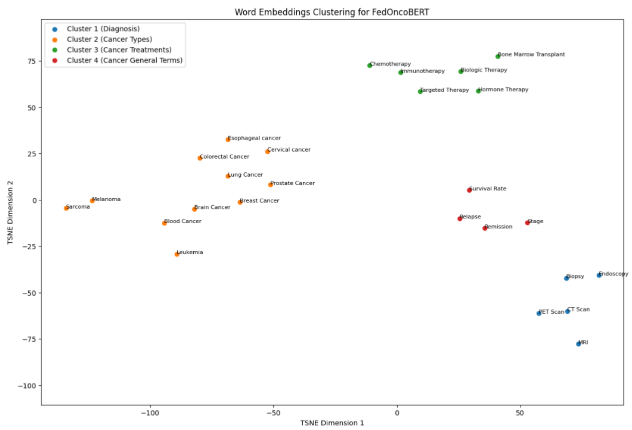
To assess the domain adaptation in acquiring semantic relations and similarities between oncology-related terms, we performed embedding visualization and clustering analysis on BioBERT and our two domain-adapted models (one developed without federation and the other with federation). t-SNE (t-distributed Stochastic Neighbor Embedding) is a



(a) BioBERT Embeddings



(b) Oncology Pre-tuned BERT Embeddings



(c) Federated Oncology Pre-tuned BERT Embeddings

Figure 3 – Embedding visualisations for different BERT-based models.

machine learning algorithm for dimensionality reduction, particularly useful for visualizing high-dimensional data. In our context, t-SNE helps to reduce the high-dimensional word embeddings to a 2D space, allowing us to visualize and analyze the relationships between different oncology terms. Figure 3 presents the visualisation of word embeddings clustered by t-SNE dimensionality reduction for all three BERT models. Each point in the visualisations is an entity (a word or term), and colours indicate different clusters formed by the k-means clustering algorithm. These visualizations are important as they allow us to observe how different oncology terms are grouped or separated in the embedding space, reflecting the models' understanding of semantic relationships. From Figure 3a, we notice that in BioBERT clustering, cancer treatment, diagnosis, and general medical terminology terms are scattered across different clusters instead of being closely grouped together. For instance, *Leukemia, Melanoma*, *Sarcoma* are moved to different clusters with general terms. This suggests, that the generic BERT model is not able to catch the semantic relations between oncology specific entities properly.

In contrast, the clustering results of domain adapted models (Figures 3b and 3c) demonstrate improved clustering of oncology related terms. Cancer types, treatments, and diagnostic procedures are properly grouped into different clusters, which implies that these models have successfully caught the semantic similarities and relationships in oncology. The proximity and direction of the word embeddings in the vector space reflect an improved understanding of the relationships between oncology related terms.

This enhanced clustering and alignment of word embeddings in our domain-adapted models provide a foundation for downstream tasks such as named entity recognition, relation extraction, and text classification, where a deep understanding of oncology concepts is crucial. The semantic relationships and similarities among oncology terms are better reflected by the domain-adapted models compared to the generic BERT model, suggesting they are more competent in addressing NLP tasks in oncology.

4.2 Named Entity Recognition (NER) Task

To assess the impact of federated learning and domain adaptation on named entity recognition in oncology, we evaluated our domain-adapted BERT models and BioBERT using a manually annotated dataset of 1550 private clinical reports. This dataset provided a robust test of the models' ability to detect and classify oncology-related entities. Table 1 presents the NER task results for each model over three fine-tuning epochs, including precision (the proportion of correctly identified entities among all predicted entities), recall (the proportion of correctly identified entities among all actual entities), F1-score (the harmonic mean of precision and recall), and accuracy. These metrics collectively offer a comprehensive view of how domain adaptation and federated learning influence model performance, with precision and recall specifically highlighting the models ability to correctly

Model	Ep.	Prec.	Rec.	F1	Acc.
BioBERT	1	0.579 761	0.561 087	0.570 271	0.562 193
	2	0.613 650	0.602 766	0.608 159	0.609 599
	3	0.621 776	0.606 939	0.614 268	0.619 746
Finetuned BERT	1	0.606 918	0.584 869	0.595 690	0.592 137
	2	0.623 128	0.626 762	0.624 940	0.629 791
	3	0.625 666	0.624 729	0.625 197	0.637 718
Federated BERT	1	0.611 512	0.580 348	0.595 523	0.590 274
	2	0.616 274	0.621 599	0.618 925	0.626 179
	3	0.623 578	0.623 311	0.623 445	0.634 450

Table 1 – Performance metrics of BERT models over three epochs.

identify entities while minimizing false positives and false negatives.

The training outcomes indicate a bit of an improvement in all the models, with our domain adapted models being the best in precision, recall, F1-score, and accuracy. Nevertheless, the real influence of domain adaptation and federated learning is most visible when we assess the NER-annotated data with a focus on cancer related entities. Table 2 presents the frequency of recognition for specific cancer-related named entities when comparing the performance of BioBERT and our domain adapted models on the clinical reports dataset. This highlights the improvements achieved by our domain adapted models in recognising critical cancer related entities compared to the generic BioBERT model. For instance, our models demonstrate a notable increase in the recognition frequency of cancer treatment, prosthetic, and drug regimen entities. These improvements are due to the domain specific fine tuning of our models on oncology related datasets, thus, they can better be used for capturing the nuances and terminology that are particular to the oncology field.

Tag	BioBERT	Fine-tuned BERT	Federated BERT
cancer_treatment	7	269	307
prosthetic	504	601	655
drug_regimen	1885	2095	2239
pathological_findings	118	254	356

Table 2 – Number of tagged instances identified.

5. CONCLUSION

This study shows the possibility of using federated learning and LLM's domain adaptation techniques to improve cancer treatment. We addressed the problems of data privacy, governance, and resource limitations using the transformer based network of BioBERT, pre-training it on oncology specific datasets, and introducing federated learning techniques. Our domain adapted models were better than the generic ones in understanding oncology texts and recognising oncology related entities, which led to improved semantic understanding and accuracy. Even though the strategies mentioned in this paper are designed for encoder based transformers, they can also be applied to decoder based transformer architectures. This lays the foundation for further exploration and application in various natural language processing tasks, broadening the impact and utility of our approach. Additionally, our research aligns with and supports several United Nations Sustainable Development Goals, specifically SDG 3 (Good Health and Well-being), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 10 (Reduced Inequalities). By developing advanced AI models for cancer care, applying cutting-edge technologies, and employing federated learning to create AI models using data from different regions, our study contributes to improving global oncology care and healthcare accessibility while promoting a more equitable and sustainable future.

REFERENCES

[1] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[3] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[4] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[5] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

[6] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[7] Yifan Peng, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: an evaluation of bert and elmo on ten benchmarking datasets. *arXiv preprint arXiv:1906.05474*, 2019.

[8] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language

model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.

[9] Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

[10] Bofeng Zhang, Xiuhong Yao, Haiyan Li, and Mirensha Aini. Chinese medical named entity recognition based on expert knowledge and fine-tuning bert. In *2023 IEEE International Conference on Knowledge Graph (ICKG)*, pages 84–90, 2023.

[11] Ning Liu, Qian Hu, Huayun Xu, Xing Xu, and Mengxin Chen. Med-bert: A pretraining framework for medical records named entity recognition. *IEEE Transactions on Industrial Informatics*, 18(8):5600–5608, 2022.

[12] Rajesh Kumar, Abdullah Aman Khan, Jay Kumar, Noorbakhsh Amiri Golilarz, Simin Zhang, Yang Ting, Chengyu Zheng, Wenyong Wang, et al. Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging. *IEEE Sensors Journal*, 21(14):16301–16314, 2021.

[13] Yiqiang Chen, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao. Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, 35(4):83–93, 2020.

[14] Dianbo Sui, Yubo Chen, Jun Zhao, Yantao Jia, Yuantao Xie, and Weijian Sun. Feded: Federated learning via ensemble distillation for medical relation extraction. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 2118–2128, 2020.

[15] Ittai Dayan, Holger R Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J Wood, Chien-Sung Tsai, et al. Federated learning for predicting clinical outcomes in patients with covid-19. *Nature medicine*, 27(10):1735–1743, 2021.

[16] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.