MambaHealth: A Lightweight Foundation Model for Efficient Drug Recommendation

Yuda Wang
School of Computing and Data Science
The University of Hong Kong
Pokfulam, Hong Kong
yuda_wang@connect.hku.hk

Xuxin He
School of Medicine
The Chinese University of Hong Kong, Shenzhen
2001 Longxiang Blvd, Longgang District, 518172, Shenzhen, China
xuxinhe@link.cuhk.edu.cn

Shengxin Zhu*
Advanced Institute of Natural Science
Beijing Normal University
Zhuhai 519087, P.R.China
Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science
BNU-HKBU United International College
Zhuhai 519087, P.R.China
Shengxin.Zhu@bnu.edu.cn

Abstract

Advancements in medical foundation models are revolutionizing healthcare by enabling more personalized and interpretable patient care. We introduce MambaHealth, a lightweight and efficient foundation model designed to address complex healthcare challenges through task-specific adaptation. MambaHealth's pretraining integrates diagnostic and procedural data to derive detailed patient state representations, which are then fine-tuned for applications such as drug recommendation, multi-diagnosis management, and temporal prescription optimization. A key feature of MambaHealth is its emphasis on explainability, providing transparent and interpretable insights into its decision-making processes, thereby enhancing trust and reliability in clinical environments. Moreover, MambaHealth offers personalized recommendations based on individual patient data, ensuring adaptability to each patient's unique characteristics. By continuously refining its parameters with updated clinical data, MambaHealth consistently outperforms existing models in both accuracy and efficiency, making it a valuable tool for advancing intelligent healthcare management and supporting informed clinical decision-making.

*Corresponding author

AIM-FM Workshop, 38th Conference on Neural Information Processing Systems (NeurIPS 2024).

1 Introduction

1.1 Background and motivation

Large Foundation Models (FMs) represent the latest advancement in the field of artificial intelligence (AI). These models are trained on massive and diverse datasets, enabling them to excel in a variety of tasks, including text generation, image recognition, and complex decision-making. This broad applicability marks a significant leap in AI technology, offering more flexible and comprehensive solutions compared to earlier models that focused on solving specific tasks. For example, GPT-4 [Achiam *et al.*, 2023], as a language model, achieved unprecedented levels in natural language processing, demonstrating fluent text generation and deep language understanding, which significantly enhanced AI's performance in text-based tasks.

In the healthcare sector, the introduction of large foundation models is driving profound changes. These models can integrate and analyze data from various sources, such as medical images, electronic health records (EHRs), and genomic data, showcasing their robust capabilities in handling complex medical tasks. For instance, they can be applied to intelligent diagnosis, personalized treatment recommendations, and disease prediction, providing more accurate and personalized healthcare services through in-depth analysis of patients' historical and real-time data.

Despite the tremendous potential of large foundation models in healthcare, their application still faces several challenges. First, these models typically require substantial computational resources and training time, which may limit their deployment and efficiency in practical settings. Second, while large foundation models perform exceptionally well in many tasks, the transparency and explainability of their decision-making processes remain critical issues. This is particularly important in medical decision-making, where the model's ability to provide clear explanations is essential for building trust with healthcare professionals and patients. Additionally, achieving truly personalized healthcare remains challenging, with ongoing efforts needed to effectively integrate individualized data into the model's predictive capabilities to provide precise medical recommendations.

1.2 Contribution

This paper presents MambaHealth, a foundation model tailored for healthcare applications, with the following key contributions:

- **Task-Specific Pre-training for Healthcare:** We employ a targeted pre-training approach that integrates diagnostic and procedural data, generating rich patient health representations that form the basis for various downstream medical tasks.
- **Adaptive and Explainable Healthcare Solutions:** MambaHealth combines adaptability with explainability, providing personalized and transparent predictions that enhance trust and usability in clinical environments.
- **Efficient and Scalable Performance:** Despite its lightweight design, MambaHealth demonstrates superior accuracy and reliability across multiple healthcare tasks, offering a scalable solution for intelligent healthcare management.

2 Related Work

2.1 Development of Large Foundation Models

The evolution of Large Foundation Models (FMs) has marked a transformative shift in the field of artificial intelligence, significantly enhancing capabilities across a variety of domains. The development of these models has been driven by advancements in neural network architectures, increased computational resources, and the availability of vast datasets. The concept of foundation models builds upon earlier advancements in deep learning and neural network architectures. The introduction of architectures such as the Transformer model [Vaswani, 2017]revolutionized the field by enabling scalable and efficient processing of sequential data. Transformers, with their attention mechanisms, provided a robust framework for handling long-range dependencies in text, setting the stage for the development of large-scale models.

The advent of large language models (LLMs) marked a significant milestone in the development of FMs. Models like GPT-3 [Floridi and Chiriatti, 2020] by OpenAI demonstrated the potential of scaling up neural network architectures to achieve state-of-the-art performance on various natural language processing (NLP) tasks. GPT-3, with its 175 billion parameters, showcased the capability of large FMs to perform a wide range of tasks, including text generation, translation, and question answering, with minimal task-specific fine-tuning. Recent advancements have seen the emergence of multimodal models that integrate text, images, and other data types. For instance, CLIP [Carlsson *et al.*, 2022] and DALL-E [Dayma *et al.*, 2021], developed by OpenAI, combine vision and language to generate and understand complex multimodal content. These models illustrate the potential of large FMs to bridge different types of data and enhance tasks like image captioning and visual question answering.

The impact of large foundation models extends across various domains, from enhancing natural language understanding and generation to revolutionizing fields such as healthcare, finance, and autonomous systems. Their ability to generalize across diverse tasks and domains demonstrates the transformative potential of large-scale AI models.

2.2 Impact and Advancements of Large Foundation Models in Healthcare

The application of Large Foundation Models (FMs) in healthcare has significantly advanced medical research, diagnosis, and patient care. In supplementary treatment and diagnosis, models such as MedGPT [Kraljevic *et al.*, 2021], LLM-Mini-CEX [Shi *et al.*, 2023] , and SkinGPT-4 [Zhou *et al.*, 2023] enhance diagnostic accuracy and decision-making by providing valuable insights from extensive clinical data. DoctorGLM [Xiong *et al.*, 2023] , for example, offer diagnostic assistance and simulate clinical scenarios, leveraging its comprehensive training datasets. In drug design, models like the PanGu Drug Model [Lin *et al.*, 2022] and HelixFold-Single Fang *et al.* [2022] are revolutionizing the field by predicting molecular interactions and designing new drug compounds, thereby accelerating the drug discovery process and reducing associated costs. Advanced models such as DSI-Net [Zhu *et al.*, 2021] , MedLSAM [Lei *et al.*, 2023] , and Lvit [Li *et al.*, 2023b] are transforming medical image segmentation, crucial for accurate disease detection and assessments in radiology and pathology.

In the domain of doctor-patient communication, models such as BioMedLM [Bolton *et al.*, 2024] and ChatDoctor [Li *et al.*, 2023a] improve interactions by understanding and generating medical dialogue, which enhances communication and patient education. Multimodal integration is further advanced by models like OpenMEDLab [Wang *et al.*, 2024], which combine text, images, and other medical data to provide holistic insights into patient care. In health management, models such as GatorTron [Yang *et al.*, 2022] utilize data from electronic health records and real-time monitoring to optimize treatment plans and predict outcomes, supporting proactive and personalized care strategies. These advancements underscore the transformative potential of Large FMs in healthcare, reflecting their ability to address complex medical challenges and improve various facets of patient care. Despite their promising capabilities, ongoing challenges related to computational resources, data privacy, and the need for model explainability and interpretability continue to necessitate further research and development.

3 Methodology

3.1 Foundation Model Architecture: MambaHealth

MambaHealth is a lightweight foundation model designed for efficient and accurate drug recommendation in healthcare settings. Unlike traditional large-scale models, MambaHealth focuses on delivering robust performance with minimal parameters, making it suitable for resource-constrained environments. The model integrates diagnosis and procedure data into a unified patient representation, leveraging specialized embeddings and a state space model to capture the dynamic health states of patients.

3.1.1 Patient Health State Modeling

MambaHealth employs compact embedding matrices to represent both diagnosis and procedure codes:

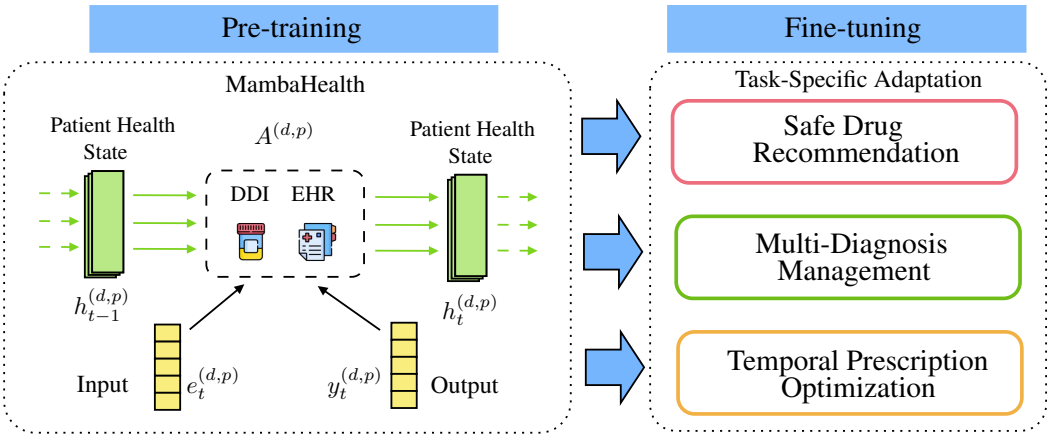


Figure 1: MambaHealth.

Diagnosis Embedding The diagnosis embedding matrix $\mathcal{E}_{\mathcal{D}}$ maps each diagnosis code to a low-dimensional vector space. For a given diagnosis vector $\mathbf{d}(t)$, the corresponding embedding $\mathbf{e}_{\mathcal{D}}(t)$ is computed as:

$$\mathbf{e}_{\mathcal{D}}(t) = \mathbf{d}(t) \cdot \mathcal{E}_{\mathcal{D}}.$$

Procedure Embedding Similarly, the procedure embedding matrix $\mathcal{E}_{\mathcal{P}}$ maps procedure codes to embedding vectors, producing $\mathbf{e}_{\mathcal{P}}(t)$ as:

$$\mathbf{e}_{\mathcal{P}}(t) = \mathbf{p}(t) \cdot \mathcal{E}_{\mathcal{P}}.$$

The combined embedding $\mathbf{e}_t^{(\mathcal{D},\mathcal{P})}$ is used as input for modeling the patient's health state.

3.1.2 State Space Modeling

To capture the evolution of patient states, MambaHealth uses a state space model (SSM). The latent state $\mathbf{h}_t^{(\mathcal{D},\mathcal{P})}$ is updated based on:

$$\overline{\mathbf{h}}_t^{(\mathcal{D},\mathcal{P})} = \overline{\mathbf{A}}^{(\mathcal{D},\mathcal{P})}\overline{\mathbf{h}}_{t-1}^{(\mathcal{D},\mathcal{P})} + \overline{\mathbf{B}}\mathbf{e}_t^{(\mathcal{D},\mathcal{P})},$$

where $\overline{\mathbf{A}}^{(\mathcal{D},\mathcal{P})}$ and $\overline{\mathbf{B}}$ are discrete matrices that control the state transitions. The model incorporates drug-drug interaction (DDI) and electronic health record (EHR) data through adaptive weighting to refine the state transitions.

4 Experiments

The experiments are designed to validate MambaHealth's effectiveness in drug recommendation, focusing on its ability to balance accuracy, safety, and efficiency while maintaining a small parameter footprint. We compare MambaHealth with several state-of-the-art models, including SafeDrug [Yang *et al.*, 2021] , CycleTrans [Zheng *et al.*, 2024] , MedGPT [Kraljevic *et al.*, 2021] , and BioGPT [Luo *et al.*, 2022] , across various tasks. All experiments are conducted using the MIMIC-III dataset, with standard preprocessing steps including data cleaning, normalization, and splitting into training, validation, and test sets.

4.1 Task 1: Safe Drug Recommendation

The first task evaluates MambaHealth's ability to recommend drug combinations while minimizing harmful drug interactions (DDI Rate) [Yang *et al.*, 2021] and maximizing accuracy (PRAUC)

[Boyd *et al.*, 2013] . In clinical practice, it is crucial to recommend drug combinations that avoid adverse drug-drug interactions. The DDI Rate measures the proportion of harmful interactions within the recommended combinations (lower is better), while PRAUC evaluates the accuracy of the recommendations across varying thresholds, with higher values indicating better performance. Table 1 demonstrates that MambaHealth achieves the lowest DDI Rate and the highest PRAUC, showcasing its superior ability to recommend safe and effective drug combinations.

Table 1: Comparison of DDI Rate and PRAUC for drug recommendation.

Model	**DDI Rate↓**	**PRAUC↑**
SafeDrug (2021)	0.0677 ± 0.0039	0.7448 ± 0.0062
CycleTrans (2024)	0.0315 ± 0.0080	0.6952 ± 0.0259
MedGPT (2023)	0.0412 ± 0.0042	0.7610 ± 0.0084
BioGPT (2023)	0.0455 ± 0.0050	0.7512 ± 0.0078
MambaHealth (Ours)	**0.0265 ± 0.0028**	**0.7764 ± 0.0051**

4.2 Task 2: Multi-Diagnosis Management

This task assesses MambaHealth's performance in managing complex clinical scenarios where patients have multiple concurrent diagnoses. The evaluation is based on the Jaccard Index [Niwattanakul *et al.*, 2013] and F1-Score [Yacouby and Axman, 2020] , two key metrics for multi-label classification. In real-world settings, patients often present with multiple diagnoses, requiring effective management of overlapping medical conditions. The Jaccard Index measures the similarity between predicted and actual diagnosis sets (higher is better), while the F1-Score balances precision and recall in multi-label predictions. As shown in Table 2, MambaHealth outperforms other models in both the Jaccard Index and F1-Score, indicating its superior ability to handle multi-diagnosis management tasks.

Table 2: Comparison of Jaccard Index and F1-Score for multi-diagnosis management.

Model	**Jaccard↑**	**F1-Score↑**
SafeDrug (2021)	0.4839 ± 0.0023	0.6441 ± 0.0040
CycleTrans (2024)	0.3955 ± 0.0854	0.5143 ± 0.0060
MedGPT (2023)	0.5010 ± 0.0052	0.6547 ± 0.0061
BioGPT (2023)	0.4882 ± 0.0048	0.6490 ± 0.0057
MambaHealth (Ours)	**0.5128 ± 0.0042**	**0.6146 ± 0.0039**

4.3 Task 3: Temporal Prescription Optimization

This task demonstrates MambaHealth's capability to optimize drug prescriptions over time while maintaining treatment efficacy. The evaluation focuses on the average number of drugs prescribed across multiple patient visits and how much that number is reduced over time. In clinical practice, minimizing unnecessary medication while maintaining treatment efficacy is crucial. Table 3 illustrates how MambaHealth significantly reduces the average number of drugs prescribed, emphasizing its efficiency in minimizing unnecessary medication.

Table 3: Comparison of average number of drugs prescribed for temporal sequence modeling. Ground-truth Avg. # of Drugs: 11.44. $|\Delta|$ Avg # of Drugs represents the average difference between the number of recommended drugs and the actual ground-truth number of drugs.

Model	**Avg. # of Drugs**	$\|\Delta\|$ **Avg. # of Drugs↓**
SafeDrug (2021)	19.9123 ± 0.0763	74.13%
CycleTrans (2024)	5.086 ± 0.1458	55.52%
MedGPT (2023)	15.512 ± 0.0924	48.34%
BioGPT (2023)	14.235 ± 0.0887	44.67%
MambaHealth (Ours)	**13.020 ± 0.0894**	**13.65%**

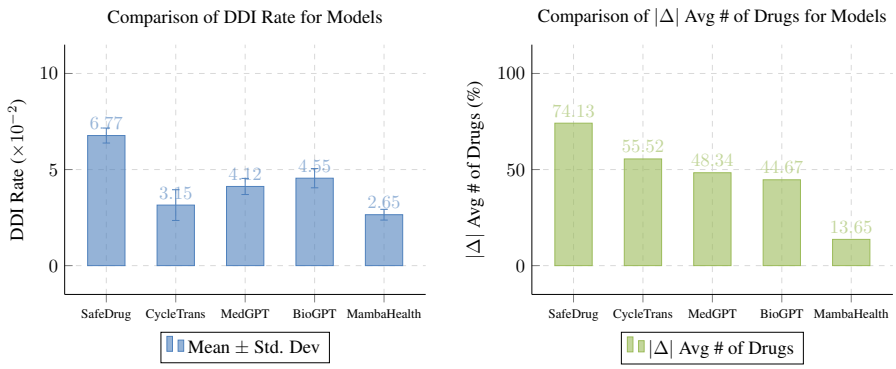


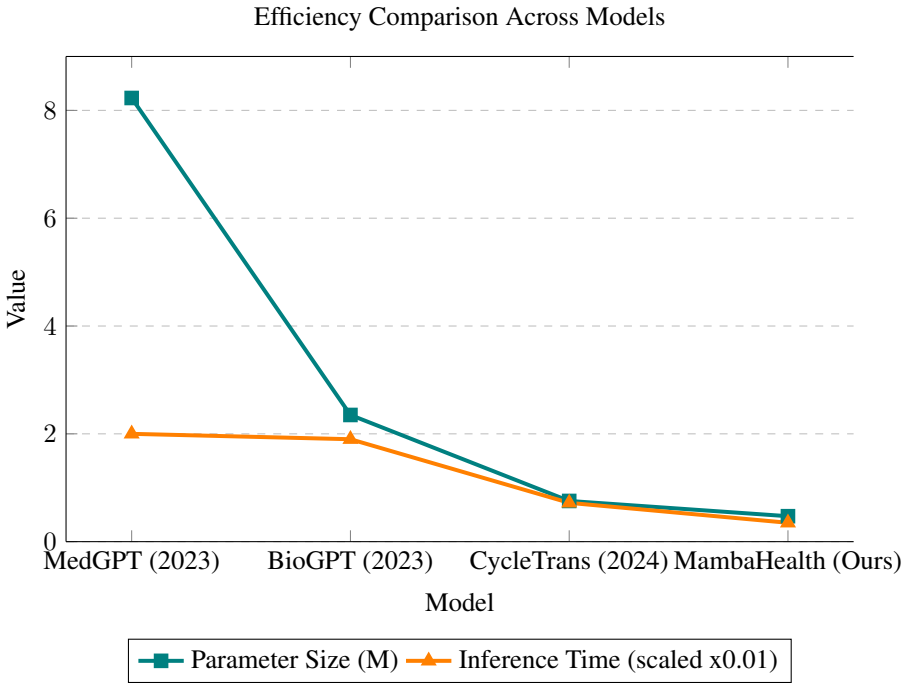
Figure 2: Comparative analysis of DDI Rates and $|\Delta|$ Average Number of Drugs across various models.

4.4 Efficiency and Parameter Evaluation

This task evaluates the efficiency metrics of MambaHealth compared to other models, including parameter size, inference time, and resource utilization. The results in Table 4 and Figure 4.4 highlight MambaHealth's significant advantages in parameter efficiency and faster inference times while maintaining comparable or superior performance.

Table 4: Efficiency Comparison Across Models.

Model	**Parameter Size (M)**	**Inference Time (ms)**
MedGPT (2023)	8.23M	200ms
BioGPT (2023)	2.35M	190ms
CycleTrans (2024)	0.755M	72ms
MambaHealth (Ours)	**0.472M**	**35ms**



5 Limitations

While MambaHealth demonstrates promising results in various healthcare tasks, several limitations need to be acknowledged:

- **Scalability Concerns:** Although MambaHealth is designed as a lightweight model compared to traditional large-scale foundation models, its scalability in extremely resource-constrained environments, such as mobile devices or edge computing scenarios, remains to be fully tested. Further optimization may be required to ensure its applicability in these settings without compromising performance.
- **Data Privacy and Security:** The use of patient data in training and deploying MambaHealth raises concerns regarding data privacy and security. Ensuring that MambaHealth adheres to strict data protection standards, such as GDPR [Voigt and Von dem Bussche, 2017] and HIPAA [English and Ford, 2004] , is essential, but practical implementation of these standards remains a challenge, especially when handling large-scale, multi-institutional data.

6 Conclusion

In this paper, we presented MambaHealth, a lightweight foundation model designed for efficient and accurate drug recommendation within healthcare settings. Unlike traditional large-scale models, MambaHealth demonstrates that high-performance drug recommendation is achievable even with minimal parameters, making it highly suitable for resource-constrained environments. By leveraging specialized embeddings, state space modeling, and integrating crucial data such as diagnoses, procedures, DDI, and EHR, MambaHealth effectively captures dynamic patient health states.

Our extensive experiments on the MIMIC-III dataset validate MambaHealth's superior ability in several key tasks, including safe drug recommendation, multi-diagnosis management, and temporal prescription optimization. The model consistently outperformed state-of-the-art models like MedGPT, BioGPT, and CycleTrans in terms of safety (DDI Rate), accuracy (PRAUC, Jaccard Index, F1-Score), and efficiency (parameter size and inference time). Specifically, MambaHealth achieves the lowest DDI Rate and highest PRAUC while maintaining a parameter size of only 0.472 million, significantly smaller than competing models.

Furthermore, the parameter efficiency and rapid inference times showcased in our results highlight the potential of MambaHealth to deliver robust drug recommendation solutions without compromising on performance, even in environments with limited computational resources.

In conclusion, MambaHealth sets a new standard for foundation models in healthcare, demonstrating that efficient, accurate, and safe drug recommendations can be achieved with minimal computational overhead. Future work will focus on extending MambaHealth's applications to broader clinical tasks and exploring further optimizations to enhance scalability and robustness in real-world deployments.

7 Acknowledgement

The research is supported by Natural Science Foundation of China (12271047); Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College (2022B1212010006); UIC research grant (UICR0400036-21C, UICR0400008-21; R04202405-21); Guangdong College Enhancement and Innovation Program (2021ZDZX1046).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Elliot Bolton, Abhinav Venigalla, Michihiro Yasunaga, David Hall, Betty Xiong, Tony Lee, Roxana Daneshjou, Jonathan Frankle, Percy Liang, Michael Carbin, et al. Biomedlm: A 2.7 b parameter language model trained on biomedical text. *arXiv preprint arXiv:2403.18421*, 2024.

Kendrick Boyd, Kevin H Eng, and C David Page. Area under the precision-recall curve: point estimates and confidence intervals. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13*, pages 451–466. Springer, 2013.

Fredrik Carlsson, Philipp Eisen, Faton Rekathati, and Magnus Sahlgren. Cross-lingual and multilingual clip. In *Proceedings of the thirteenth language resources and evaluation conference*, pages 6848–6854, 2022.

Boris Dayma, Suraj Patil, Pedro Cuenca, Khalid Saifullah, Tanishq Abraham, Phuc Le Khac, Luke Melas, and Ritobrata Ghosh. Dall· e mini. *HuggingFace. com. https://huggingface. co/spaces/dallemini/dalle-mini (accessed Sep. 29, 2022)*, 2021.

Abigail English and Carol A Ford. The hipaa privacy rule and adolescents: legal questions and clinical challenges. *Perspectives on sexual and reproductive health*, 36(2):80–86, 2004.

Xiaomin Fang, Fan Wang, Lihang Liu, Jingzhou He, Dayong Lin, Yingfei Xiang, Xiaonan Zhang, Hua Wu, Hui Li, and Le Song. Helixfold-single: Msa-free protein structure prediction by using protein language model as an alternative. *arXiv preprint arXiv:2207.13921*, 2022.

Luciano Floridi and Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30:681–694, 2020.

Zeljko Kraljevic, Anthony Shek, Daniel Bean, Rebecca Bendayan, James Teo, and Richard Dobson. Medgpt: Medical concept prediction from clinical narratives. *arXiv preprint arXiv:2107.03134*, 2021.

Wenhui Lei, Xu Wei, Xiaofan Zhang, Kang Li, and Shaoting Zhang. Medlsam: Localize and segment anything model for 3d medical images. *arXiv preprint arXiv:2306.14752*, 2023.

Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. *Cureus*, 15(6), 2023.

Zihan Li, Yunxiang Li, Qingde Li, Puyang Wang, Dazhou Guo, Le Lu, Dakai Jin, You Zhang, and Qingqi Hong. Lvit: language meets vision transformer in medical image segmentation. *IEEE transactions on medical imaging*, 2023.

Xinyuan Lin, Chi Xu, Zhaoping Xiong, Xinfeng Zhang, Ningxi Ni, Bolin Ni, Jianlong Chang, Ruiqing Pan, Zidong Wang, Fan Yu, et al. Pangu drug model: learn a molecule like a human. *Biorxiv*, pages 2022–03, 2022.

Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. Biogpt: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6):bbac409, 2022.

Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. Using of jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 380–384, 2013.

Xiaoming Shi, Jie Xu, Jinru Ding, Jiali Pang, Sichen Liu, Shuqing Luo, Xingwei Peng, Lu Lu, Haihong Yang, Mingtao Hu, et al. Llm-mini-cex: Automatic evaluation of large language model for diagnostic conversation. *arXiv preprint arXiv:2308.07635*, 2023.

A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555, 2017.

Xiaosong Wang, Xiaofan Zhang, Guotai Wang, Junjun He, Zhongyu Li, Wentao Zhu, Yi Guo, Qi Dou, Xiaoxiao Li, Dequan Wang, et al. Openmedlab: An open-source platform for multi-modality foundation models in medicine. *arXiv preprint arXiv:2402.18028*, 2024.

Honglin Xiong, Sheng Wang, Yitao Zhu, Zihao Zhao, Yuxiao Liu, Linlin Huang, Qian Wang, and Dinggang Shen. Doctorglm: Fine-tuning your chinese doctor is not a herculean task. *arXiv preprint arXiv:2304.01097*, 2023.

Reda Yacouby and Dustin Axman. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In *Proceedings of the first workshop on evaluation and comparison of NLP systems*, pages 79–91, 2020.

Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. *arXiv preprint arXiv:2105.02711*, 2021.

Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Mona G Flores, Ying Zhang, et al. Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records. *arXiv preprint arXiv:2203.03540*, 2022.

Yuhan Zheng, Xiaotao Lin, Kexuan Chen, and Shengxin Zhu. Cycletrans: a transformer-based clinical foundation model for safer prescription. In *AAAI 2024 Spring Symposium on Clinical Foundation Models*, 2024.

Juexiao Zhou, Xiaonan He, Liyuan Sun, Jiannan Xu, Xiuying Chen, Yuetan Chu, Longxi Zhou, Xingyu Liao, Bin Zhang, and Xin Gao. Skingpt-4: an interactive dermatology diagnostic system with visual large language model. *arXiv preprint arXiv:2304.10691*, 2023.

Meilu Zhu, Zhen Chen, and Yixuan Yuan. Dsi-net: Deep synergistic interaction network for joint classification and segmentation with endoscope images. *IEEE Transactions on Medical Imaging*, 40(12):3315–3325, 2021.

A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD be included in the main submission.**

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification:The abstract and introduction clearly states the claims

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: check in papers

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: the paper does not include theoretical results

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: the code will be sumitted together

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: codes and data will be summitted

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification:The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: in appendix

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: we make sure to preserve anonymity

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: not include

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: no such risk

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification:we have cited the original paper that produced the code package or dataset

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification:

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: paper does not involve crowdsourcing nor research with human subjects.