European Journal of Nuclear Medicine and Molecular Imaging (2024) 51:332–335
https://doi.org/10.1007/s00259-023-06450-7

EDITORIAL

Medical image Generative Pre-Trained Transformer (MI-GPT): future direction for precision medicine

Xiaohui Zhang[1,2,3] · Yan Zhong[1,2,3] · Chentao Jin[1,2,3] · Daoyan Hu[1,2,3] · Mei Tian[1,2,3,4] · Hong Zhang[1,2,3,5,6]

Published online: 7 October 2023

Medical imaging has its earliest roots in 1895 when Wilhelm Roentgen discovered X-ray, providing physicians with the first approach to image internal conditions of human body [1]. After that, multiple imaging methods were developed and optimized in succession based on various imaging principles, such as computed tomography (CT) [2], magnetic resonance imaging (MRI) [3], and positron emission tomography (PET) [4]. The advent of these imaging techniques has rendered medical imaging a crucial pillar of clinical practice and a fundamental domain for the realization of precision medicine.

With the ongoing advancements in biological and instrumental science, medical imaging technologies have made remarkable progress in recent decades. The overall structural, functional, and molecular alterations of the individuals could be obtained non-invasively through multiple imaging methods [5]. Especially, with the development of molecular imaging, pathophysiological processes at the cellular and molecular levels can be precisely visualized, characterized, and quantified [6–8]. The continuous advancement of imaging equipment and probes has further enhanced the capacity of molecular imaging to evaluate pathophysiological alternations noninvasively, thereby making the diagnostic capabilities increasingly approach the level of pathological practice. Recently, a novel pattern of pathological practice termed "transpathology," which could comprehensively depict pathophysiological events in vivo from a multiscale perspective, holds the great potential to facilitate the translational processes from the bench to the bedside and drive traditional medicine towards precision medicine [9].

In parallel with the advancement of medical imaging technology, medical image analysis methods have also experienced rapid development, with an increasing focus on quantification and intelligence. In 2012, "radiomics" was proposed as an innovative approach to image analysis, using automated high-throughput extraction of large amounts of quantitative features from standard-of-care medical images [10]. With the assistance of artificial intelligence (AI), radiomics and other medical image analysis approaches could potentially aid more complex decision-making tasks, such as disease prognostication, prediction of response to different treatment modalities, recognition of treatment-related changes, and discovery of imaging representations of phenotypic and genotypic features associated with prognosis [11]. However, the existing AI-based methodologies for medical image analysis encounter various obstacles. The dominant research paradigm heavily depends on a substantial quantity of annotated training samples to construct models tailored to particular tasks, which is heavily reliant on extensive medical imaging datasets [12]. Nevertheless, the scarcity of annotated medical data restricts the model's generalizability, impeding its potential to achieve robust transferability across diverse tasks and diseases [13]. Moreover, a significant proportion of existing medical image intelligence models predominantly rely on image data, with limited incorporation of textual language data. In clinical practice, radiologists often rely on extensive textual information during the process of

✉ Mei Tian
tianmei@fudan.edu.cn

✉ Hong Zhang
hzhang21@zju.edu.cn

[1] Department of Nuclear Medicine and PET Center, The Second Affiliated Hospital of Zhejiang University School of Medicine, 88 Jiefang Road, Hangzhou 310009, Zhejiang, China

[2] Institute of Nuclear Medicine and Molecular Imaging of Zhejiang University, Hangzhou, China

[3] Key Laboratory of Medical Molecular Imaging of Zhejiang Province, Hangzhou, China

[4] Human Phenome Institute, Fudan University, 825 Zhangheng Road, Shanghai 201203, China

[5] College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou, China

[6] Key Laboratory for Biomedical Engineering of Ministry of Education, Zhejiang University, Hangzhou, China

medical image diagnoses, leading to a stark disparity with the model's architecture. This incongruity hampers the model's ability to perform certain image-text tasks, including the automated generation of diagnostic reports for images.

Recently emerged large language models (LLMs) bringing a ray of hope to address the above issues, especially Chat Generative Pre-Trained Transformer (ChatGPT) developed by OpenAI [14]. This model is trained using a large number of textual corpora, acquiring massive knowledge that can be used for various natural language processing tasks, such as language understanding, text generation, and machine translation. It possesses the capability to receive user input and generate coherent natural language responses, thereby accomplishing seamless and articulate conversations. Recent studies indicated that ChatGPT exhibits diverse application scenarios with the domain of medical imaging, including automated reporting, patient communication, addressing specific technical inquiries [15], and educational purposes [16]. However, the limited availability of high-quality medical data in the pre-training dataset of GPT-3.5 has resulted in certain constraints on its accuracy when providing responses to medical inquiries. Furthermore, its incapability to handle image inputs hinders its applicability in the field of medical imaging. Although the updated GPT-4.0 possesses the ability to process image inputs, it still demonstrates relatively restricted proficiency in medical image recognition [17].

The Visual-Linguistic Pre-training (VLP) models exhibit the capacity to acquire transferable visual and linguistic attributes by means of pre-training on extensive multilingual data that encompasses both language and vision [18]. Within the field of medicine, the BiomedCLIP model [19], which is based on the Contrastive Language-Image Pre-training (CLIP) framework [20], has exhibited improved zero-shot predictive abilities, making it well-suited for medical image recognition tasks. Additionally, PubMedCLIP has demonstrated exceptional performance in tasks involving reciprocal retrieval of information between textual and visual modalities [21]. These VLP models have broadened the range of tasks applicable to medical imaging, enabling the seamless integration of textual and visual data. Nevertheless, there is still potential for enhancing the precision of task execution.

Herein, we propose the concept of medical image GPT (MI-GPT), a pre-training foundation model that predominantly utilizes medical imaging as a primary data source, while also integrating multi-omics data and electronic health records, which might be the future direction of foundation model for application in the medical imaging field in clinical practice (Fig. 1). The data formats used for MI-GPT can be derived from either pure image data, pure text data, or a combination of both image and text information.

To enhance the interpretability and generalizability of MI-GPT models in clinical practice, it is crucial to foster inter-institutional and multi-disciplinary research collaborations by training models on extensive datasets obtained from various medical centers, scanners, and protocols, with a focus on disease detection, segmentation, and classification tasks in specific application scenarios. Furthermore, through the integration of diverse data types (e.g., text, images, and videos) along with multidimensional data (e.g., genomics, proteomics, transcriptomics, and phenomics), the future

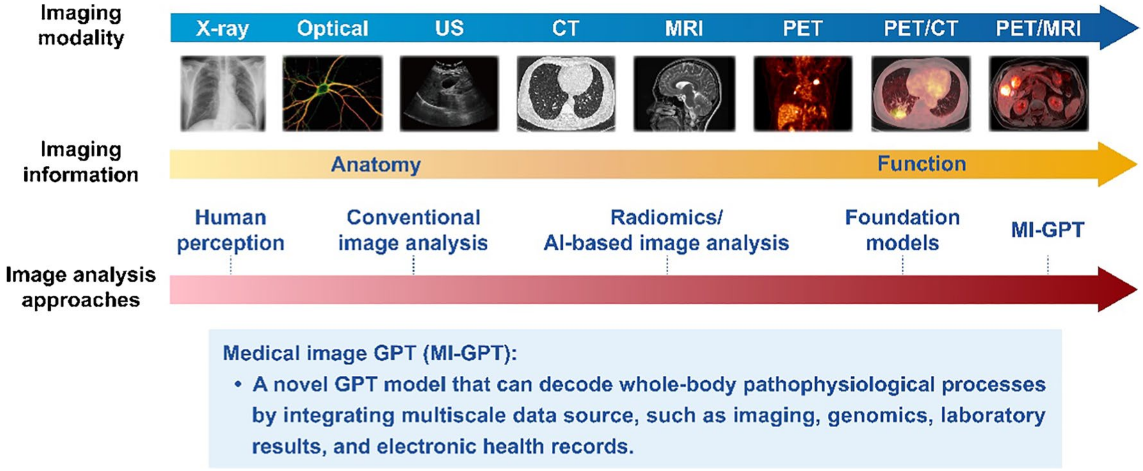
Imaging modality
X-ray
Optical
US
CT
MRI
PET
PET/CT
PET/MRI
Imaging information
Anatomy
Function
Human perception
Conventional image analysis
Radiomics/ AI-based image analysis
Foundation models
MI-GPT
Image analysis approaches
Medical image GPT (MI-GPT):
• A novel GPT model that can decode whole-body pathophysiological processes by integrating multiscale data source, such as imaging, genomics, laboratory results, and electronic health records.

Fig. 1 The development of medical imaging modalities and image analysis approaches. With the continuous advancements in the fields of biological and instrumental sciences, medical imaging technologies have progressed from unimodal structural imaging towards multimodal structural–functional imaging. Simultaneously, there is a growing inclination towards the intelligent automation of image analysis methodologies, shifting from subjective evaluations to more accurate quantitative assessments. Considering the continuous progress in foundational models within contemporary medical research, we believe that the future integration of medical foundational models customized for specific pathophysiological conditions, such as medical image Generative Pre-Trained Transformer (MI-GPT), will substantially drive the advancement of precision medicine

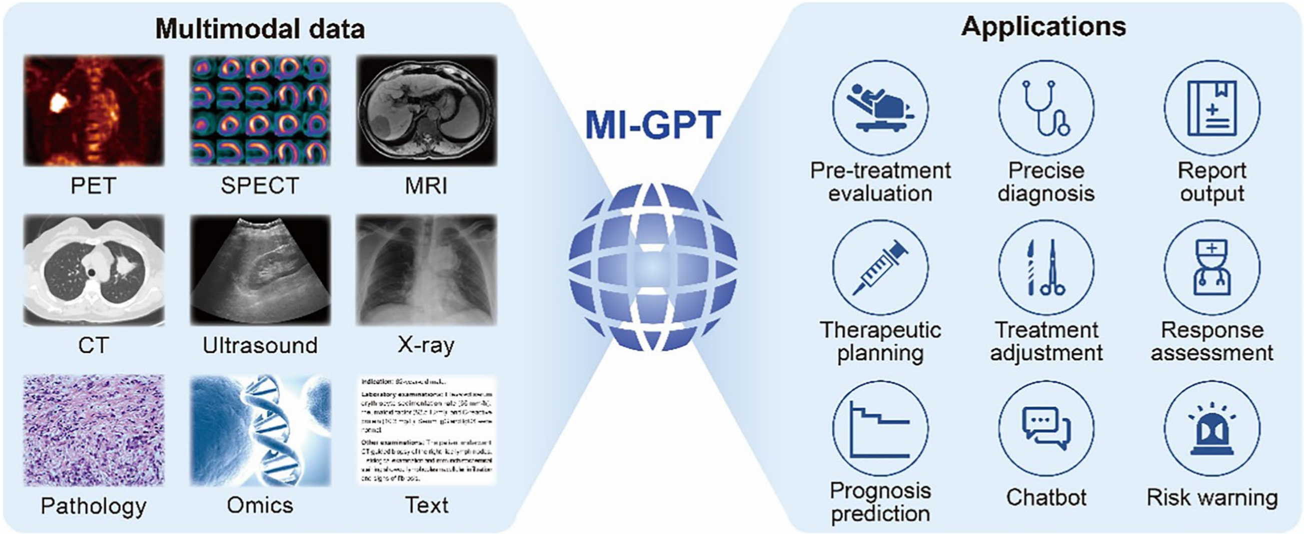

Fig. 2 MI-GPT in clinical practice. By integrating multimodal data including imaging, omics, and electronic health records, MI-GPT holds potential for the advancement of clinical applications that cater to diverse user bases and disciplines, thereby facilitating the potential for achieving more precise disease diagnoses and formulating individualized therapeutic decision-making

multi-modality MI-GPT models hold enormous promise for acquiring more comprehensive understanding of patients' condition, thereby facilitating the potential for achieving more precise disease diagnoses and formulating individualized therapeutic strategies [22, 23].

The progression of MI-GPT models holds potential for the advancement of clinical applications that cater to diverse user bases and disciplines (Fig. 2). One prominent application is that they can aid radiologists in their workflow by automating the generation of structured radiology reports and describing abnormalities and findings, while also taking into account the patient's history. Clinicians can receive additional support from MI-GPT through the combination of text reports and interactive visualizations, which may include the highlighting of the corresponding region for each phrase. Additionally, MI-GPT can assist clinicians by integrating image, language, and audio modalities, enabling real-time decision-making in clinical practice (e.g., pre-treatment comprehensive evaluation, adjustment of surgical alternatives during surgery, monitoring in vivo drug delivery and therapeutic response), leading to more efficient and effective patient management and healthcare. Furthermore, the MI-GPT is expected to predict the risk of a certain disease in the future based on the patient's previous and current conditions. Through extracting meaningful information from a patient's time series data (e.g., imaging, vital laboratory parameters, and clinical notes), the MI-GPT possess the ability to provide a comprehensive summary of the patient's current clinical state, while also projecting potential future states and offering treatment recommendations. We believe that MI-GPT can also be utilized as a chatbot to leverage multimodal data and construct a holistic understanding of a patient's condition. It possesses the capability to decipher diverse data formats and engage in interactive conversations with patients to provide detailed medical advice and explanations, which will be crucial for the comfortable and precise medicine in the future.

Author contribution All authors contributed to the manuscript writing. All authors read and approved the final manuscript.

Funding This editorial was supported by the National Key Research and Development Program of China (2021YFE0108300, 2021YFA1101700, 2022YFE0118000, 2022YFC2009900), National Natural Science Foundation of China (82030049, 32027802), and Fundamental Research Funds for the Central Universities.

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

1. Assmus A. Early history of X rays. Beam Line. 1995;25:10–24.
2. Richmond C. Sir Godfrey Hounsfield. British Medical Journal Publishing Group; 2004.

3. Mansfield P, Maudsley AA. Medical imaging by NMR. Br J Radiol. 1977;50:188–94. https://doi.org/10.1259/0007-1285-50-591-188.
4. Nutt R. The history of positron emission tomography. Mol Imaging Biol. 2002;4:11–26. https://doi.org/10.1016/s1095-0397(00)00051-0.
5. Zhang K, Sun Y, Wu S, Zhou M, Zhang X, Zhou R, et al. Systematic imaging in medicine: a comprehensive review. Eur J Nucl Med Mol Imaging. 2021;48:1736–58. https://doi.org/10.1007/s00259-020-05107-z.
6. Wells RG. Instrumentation in molecular imaging. J Nucl Cardiol. 2016;23:1343–7. https://doi.org/10.1007/s12350-016-0498-z.
7. Zhang X, Jiang H, Wu S, Wang J, Zhou R, He X, et al. Positron emission tomography molecular imaging for phenotyping and management of lymphoma. Phenomics. 2022;2:102–18. https://doi.org/10.1007/s43657-021-00042-x.
8. Tian M, Zuo C, Civelek AC, Carrio I, Watanabe Y, Kang KW, et al. International nuclear medicine consensus on the clinical use of amyloid positron emission tomography in Alzheimer's disease. Phenomics. 2023;3:375–89. https://doi.org/10.1007/s43657-022-00068-9.
9. Tian M, He X, Jin C, He X, Wu S, Zhou R, et al. Transpathology: molecular imaging-based pathology. Eur J Nucl Med Mol Imaging. 2021;48:2338–50. https://doi.org/10.1007/s00259-021-05234-1.
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6. https://doi.org/10.1016/j.ejca.2011.11.036.
11. Visvikis D, Lambin P, Beuschau Mauridsen K, Hustinx R, Lassmann M, Rischpler C, et al. Application of artificial intelligence in nuclear medicine and molecular imaging: a review of current status and future perspectives for clinical translation. Eur J Nucl Med Mol Imaging. 2022;49:4452–63. https://doi.org/10.1007/s00259-022-05891-w.
12. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, et al. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. Front Oncol. 2022;12: 773840. https://doi.org/10.3389/fonc.2022.773840.
13. Ibrahim A, Primakov S, Beuque M, Woodruff HC, Halilaj I, Wu G, et al. Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework. Methods. 2021;188:20–9. https://doi.org/10.1016/j.ymeth.2020.05.022.
14. Introducing ChatGPT. 2023. Accessed March 15, 2023. https://openai.com/blog/chatgpt
15. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology. 2023;307: e230163. https://doi.org/10.1148/radiol.230163.
16. Baidoo-Anu D, Owusu Ansah L. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. 2023.
17. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4: a new era of artificial intelligence in medicine. Ir J Med Sci. 2023. https://doi.org/10.1007/s11845-023-03377-8.
18. Gan Z, Li L, Li C, Wang L, Liu Z, Gao J. Vision-language pretraining: basics, recent advances, and future trends. Foundations and Trends® in Computer Graphics and Vision. 2022;14:163–352. doi: https://doi.org/10.1561/0600000105.
19. Wang Z, Wu Z, Agarwal D, Sun J. MedCLIP: contrastive learning from unpaired medical images and text. ArXiv. 2022;abs/2210.10163.
20. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: Marina M, Tong Z, editors. Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research: PMLR; 2021. p. 8748–63.
21. Eslami S, de Melo G, Meinel C. Does clip benefit visual question answering in the medical domain as much as it does in the general domain? ArXiv. 2021:abs/2112.13906.
22. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616:259–65. https://doi.org/10.1038/s41586-023-05881-4.
23. Zhang S, Xu Y, Usuyama N, Bagga JK, Tinn R, Preston S, et al. Large-scale domain-specific pretraining for biomedical vision-language processing. ArXiv. 2023;abs/2303.00915.