Large Model and Small Model Integrated Robot for Dynamic Medical Analysis

Daobin Luo[1,2], **Zhanyu Hu**[1,2], **Yuru Liu**[1,2], **Yuange Liu**[1,2], **Weishan Zhang**[1,2] *

[1]College of Computer Science and Technology, China University of Petroleum (East China)
[2]Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, China University of Petroleum (East China)
luodaobin2001@gmail.com, huzhanyu0328@163.com, liuyuru2021@163.com, liuyuange811@gmail.com,
zhangws@upc.edu.cn

Abstract

Medical robots encounter substantial challenges in accommodating continuously changing data distributions and dynamic task scenarios. In this demo, we present BiOpt-MRobot, a large model and small model integrated robot for dynamic medical analysis capabilities. Based on a cloud-edge-device collaborative architecture, BiOpt-MRobot innovatively constructs a bidirectional optimization mechanism that achieves collaborative training between large model and small model. Furthermore, BiOpt-MRobot incorporates an intelligent model factory that implements data-aware model customization and adaptive continuous model evolution, thereby enhancing the reliability and accuracy of medical analysis outcomes. The system currently accommodates 4 data modalities encompassing 27 disease categories. In practical application, BiOpt-MRobot utilizes terminal medical robots as carriers to perform natural language interactions with users and implements multimodal intelligent medical analysis functions. **Video Link:** XXXX.

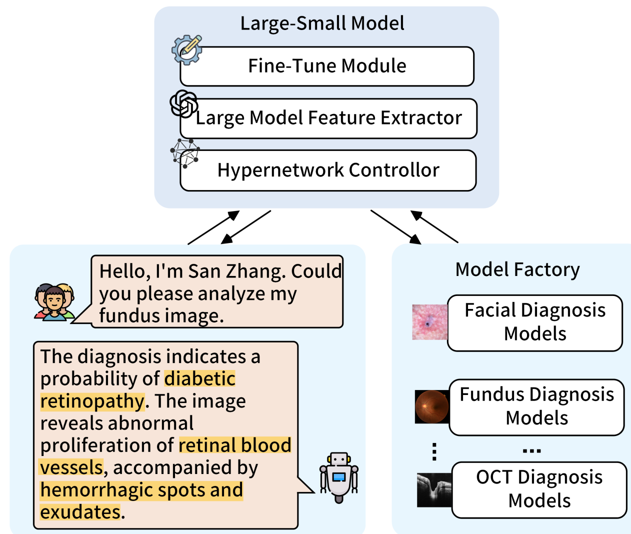

Figure 1: Overview of BiOpt-MRobot.

Introduction

Medical robot for healthcare and eldercare services are attracting increasing attentions (Jones and Dolsten 2024; Luo, Su, and Zheng 2021) (Lueg and Jungo 2021; Duan et al. 2021). Asada et al. developed a robot-assisted diabetes care consultation system (Asada et al. 2025). Song et al. proposed an AI-based emotional detection robot for real-time monitoring of elderly mental health (Song 2024). Palani et al. designed a BiLSTM-GRU-based physiological tremor classification robot (Palani et al. 2025). Min et al. further developed diagnostic assistance robots leveraging large vision models (Min, Lai, and Ren 2025).

However, current medical robots face some critical limitations: (1) Centralized training may lead to data leakage risks. (2) Significant variations in medical data make robot difficult to adapt to heterogeneous data. (3) Low intelligence to adapt to dynamic scenarios and evolving data distributions, with low capabilities for continuous learning.

Large models including domain specific larges models are showing remarkable capabilities in reasoning, knowledge representation, and decision-making, which provides us new opportunities to design a more intelligent robot that can integrating the capabilities of large models and classical small models (Liu et al. 2024). We design a **M**edical **R**obot driven by Collaborative Large model and Small Model **Bi**directional **Opt**imization (BiOpt-MRobot), providing stable and reliable intelligent consultation services for distributed medical environments. Our contributions are summarized as follows:

- A bidirectional collaborative optimization mechanism betweem large model and small model is designed, which forms a closed loop for bidirectional knowledge transferring.
- We design an intelligent model factory to quickly customize personalized diagnostic models for specific medical tasks, which enables continuous improvement of intelligence of robots, and can enhance its capabilities of adaptation to evolving and ever-changes medical tasks.

BiOpt-MRobot Overview

BiOpt-MRobot adopts a cloud-edge-device collaborative architecture as its foundation, where the cloud server accom-

*Corresponding author

modate large model and hypernetwork, edge medical institutions participate in collaborative training, and terminal robot devices execute consultation analysis. Fig 1.(a) demonstrates that the cloud server and edge institutions conduct federated training through a bidirectional collaborative optimization mechanism between large model and small model. The intelligent model factory shown in Fig 1.(b) is employed for data-aware model customization and continuous model evolution. Fig 1.(c) illustrates how terminal robot devices perform intelligent consultation services.

Bidirectional Collaborative Optimization Mechanism between Large Model and Small Model BiOpt-MRobot constructs a bidirectional collaborative optimization mechanism between large model and small model. The mechanism include three key components: a large model feature extractor, a hypernetwork model generator, and edge small models. The large model feature extractor comprises a large model with task-specific head removed and integrated LoRA adaptive module.

During federated training, the large model feature extractor performs deep feature embedding on data to extract representation vectors. Hypernetwork learns the mapping relationship between representation vectors and model parameters to dynamically generate small model weights adapted to specific data characteristics, enabling knowledge transfer from large to small models. Based on training gradients of small model, the system updates LoRA module parameters and hypernetwork weights through chain rule derivation, forming a bidirectional optimization loop between large and small models.

Adaptive Continuously Evolving Intelligent Model Factory Given the continuously data changing characteristics in medical scenarios, BiOpt-MRobot incorporates an adaptive continuously evolving intelligent model factory, which utilizes the large model feature extractor and hypernetwork to provide dynamic model customization services. When data distribution shifts, the large model feature extractor re-extracts features from client data, while hypernetwork generates customized small model weights instantly based on feature vectors. This mechanism allows newly joined medical institutions to directly customize the model for use without waiting for model training. Furthermore, the model factory periodically collects usage feedback and incremental data from medical institutions to fine-tune the large model feature extractor and hypernetwork parameters, achieving continuous model evolution.

Demonstration Logistics

The demonstration comprises two primary components. The first component presents the BiOpt-MRobot platform, demonstrating federated collaborative training and continuous model evolution capabilities. The second component demonstrates intelligent medical consultation interactions between the medical robot and users, implementing multimodal medical analysis based on user medical images and consultation processes.

Platform Introduction The BiOpt-MRobot platform consists of a cloud platform and a client platform. The cloud platform manages the configuration, initialization, and visualization of federated collaborative training, achieving continuous model evolution through periodic update mechanisms. It integrates a general intelligent consultation module that supports multimodal diagnostic analysis. The cloud platform also provides deployment and operational management for large models and hypernetworks. The client platform serves edge medical institutions. It features a data-aware personalized intelligent consultation module, which customizes models based on institutional data distribution. When local data distribution shifts occur, the client platform adaptively reconstructs personalized diagnostic models and implements autonomous edge deployment through the robot model management module.

Application of Medical Robots In practical applications, BiOpt-MRobot utilizes medical robots as carriers to conduct intelligent consultation interactions with users. The robot identifies users who wave in the crowd, automatically navigates to their location, and initiates conversation invitations. Users can match medical images in the institutional database through identity information description or capture examination areas directly using the robot's onboard camera. The medical image matching method supports acquisition of fundus images, retinal OCT images, and dermoscopy images, while the onboard camera capture method enables acquisition of users' ocular appearance images. After obtaining images, the robot performs preliminary diagnosis using local small models, uploads the images and preliminary diagnostic results to the cloud-based large model for consultation. Finally, the system performs multimodal fusion analysis based on images, small model diagnostic results, and consultation processes to provide further diagnostic conclusions and relevant medical suggestions.

Conclusion

BiOpt-MRobot demonstrates the practical application value of large-small model bidirectional optimization collaborative framework in real medical scenarios. Through the cloud-edge-device collaborative architecture design, the system achieves personalized model customization and continuous evolution, providing intelligent consultation solutions for medical institutions. The demonstration showcases the promising prospects of federated learning and model collaborative optimization technologies in the medical robot domain.

References

Asada, Y.; Horiguchi, T.; Yagi, K.; Komatsu, M.; Yamashita, A.; Ueta, R.; Yamaaki, N.; Shikida, M.; Nishio, S.; and Inagaki, M. 2025. Robot-Assisted Approach to Diabetes Care Consultations: Enhancing Patient Engagement and Identifying Therapeutic Issues. *Medicina*, 61(2): 352.

Duan, S.; Liu, L.; Chen, Y.; Yang, L.; Zhang, Y.; Wang, S.; Hao, L.; and Zhang, L. 2021. A 5G-powered robot-assisted teleultrasound diagnostic system in an intensive care unit. *Critical Care*, 25(1): 134.

Jones, C. H.; and Dolsten, M. 2024. Healthcare on the brink: navigating the challenges of an aging society in the United States. *npj Aging*, 10(1): 22.

Liu, Y.; Zhang, W.; Bao, Z.; Chai, X.; Gu, M.; Jiang, W.; Zhang, Z.; Tian, Y.; and Wang, F.-Y. 2024. Brain-like cognition-driven model factory for IIoT fault diagnosis by combining LLMs with small models. *IEEE Internet of Things Journal*.

Lueg, C.; and Jungo, V. 2021. Mobile remote presence robots for medical consultation and social connectedness. In *Public Health and Informatics*, 999–1003. IOS Press.

Luo, Y.; Su, B.; and Zheng, X. 2021. Trends and challenges for population and health during population aging—China, 2015–2050. *China CDC weekly*, 3(28): 593.

Min, Z.; Lai, J.; and Ren, H. 2025. Innovating robot-assisted surgery through large vision models. *Nature Reviews Electrical Engineering*, 1–14.

Palani, P.; Panigrahi, S.; Renganathan, G.; Kurita, Y.; and Thondiyath, A. 2025. Empirical classification of fatigue-induced physiological tremor in robot-assisted manipulation tasks using BiLSTM-GRU network. *Frontiers in Rehabilitation Sciences*, 6: 1474203.

Song, S. 2024. Emotion detection of elderly people in nursing homes based on AI robot vision. *Soft Computing*, 28(23): 13989–14002.