Brain Cognition Driven Large Model and Small Model Integrated Robot for Dynamic Medical Analysis

Daobin Luo[1,2], Yuru Liu[1,2], Zhanyu Hu[1,2], Baoyu Zhang[1,2], Weishan Zhang[1,2] * Tao Chen[1,2], Shaohua Cao[1,2], Fei-yue Wang[3], Xiaoli Li[4], Zengxiang Li[5]

[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)
[3]Institute of Automation, Chinese Academy of Sciences, China
[4]Agency for Science Technology and Research (A*STAR), Singapore
[5]Nanyang Technological University, Singapore
luodaobin2001@gmail.com, huzhanyu0328@163.com, liuyuru2021@163.com, liuyuange811@gmail.com, zhangws@upc.edu.cn, feiyue.wang@ia.ac.cn, xlli@i2r.a-star.edu.sg, drive-shares-dm-noreply@google.com

Abstract

Medical robots encounter substantial challenges in accommodating continuously changing data and dynamic task scenarios. To address this, in this demo, we design a **M**edical **R**obot driven by Brain Cognition based large model and small model Collaboration (BC-MRobot) for dynamic medical data analysis. BC-MRobot constructs a bidirectional optimization mechanism that achieves collaborative training between large model and small model. Furthermore, BC-MRobot incorporates an intelligent model factory that achieves data-aware model adaptation and customization, which can enhance its capabilities for handling ever changing data features in different analyzing tasks. BC-MRobot currently accommodates 4 data modalities encompassing 27 disease categories, which can be used for interactive medical analysis where different analyzing models can be generated dynamically at run time, with high analysis accuracy. **Video Link:** https://luodaobinn.github.io/#paper/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).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.

On the one hand, federated learning provides us new ideas for data privacy. Scott et al. confirm that the personalized federation method based on hypernetwork can effectively deal with data heterogeneity problems (Scott, Zakerinia, and Lampert 2024). On the other hand, current 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. Furthermore, brain cognition has been igniting artificial intelligence since the discipline started. There exists research that uses brain cognition idea to integrate capabilities of large models and classical small models (Liu et al. 2024) to handle the evolving and changing data features. Research shows that increasing the endogenous potential of the hippocampus to generate new neurons throughout life rejuvenates learning and memory (Danciu et al. 2025), demonstrating that neurogenesis provides crucial neural plasticity for continuous learning. This idea can be applied to the design of medical robot.

Therefore, in this paper a **M**edical **R**obot driven by Brain Cognition based large model and small model Collaboration (BC-MRobot) for dynamic medical data analysis in distributed medical environment. Our contributions are summarized as follows:

- A bidirectional collaborative optimization mechanism between large model and small model in federated learning scenarios 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 medical tasks.

BC-MRobot Overview

BC-MRobot adopts a cloud-edge-device architecture , where the cloud server accommodate large model and hypernetworks. The medical institutions participate in collaborative training from the edge side, and robots execute onsite analysis. Fig 1.(a) demonstrates that the cloud server and medical institutions can 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 designed for data-aware model cus-

*Corresponding author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

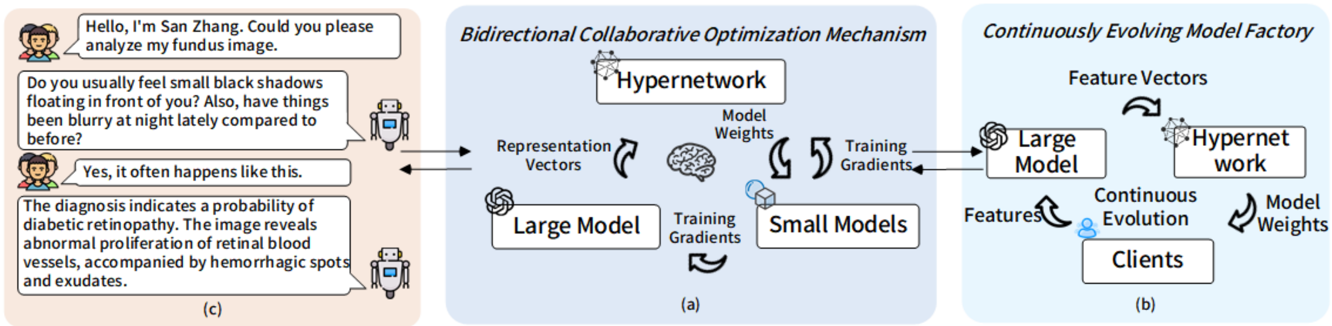

Figure 1: Overview of the design of BC-MRobot working mechanisms.

tomization and continuous model evolution. Fig 1.(c) illustrates how a robot perform medical data analysis.

Bidirectional Collaborative Optimization Mechanism between Large Model and Small Model BC-MRobot leverages the brain's self-reflection mechanism to construct a bidirectional optimization method between large and small models, where a high dimensional feature extractor is designed to extract high-level representations, a hypernetwork is designed to generate model weights and architectures, and small models are applied to specific tasks. 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 mapping relationships between representation vectors and model parameters to dynamically generate small model weights adapted to specific data characteristics, enabling knowledge transfer from large model to small model. Based on training gradients of small model, BC-MRobot updates the LoRA module parameters and hypernetwork weights by applying backpropagation, forming a bidirectional optimization loop between large model and small model.

Adaptively Evolving Model Factory BC-MRobot incorporates an adaptively and continuously evolving intelligent model factory, which utilizes the large model feature extractor and hypernetwork to provide dynamic model customization services. When data distribution changes, the large model feature extractor re-extracts features from client data, while the hypernetwork generates customized small model weights instantly based on these feature vectors. This allows newly joined medical institutions to directly customize small models without the need for re-training. Furthermore, the model factory periodically collects usage feedback and incremental data from medical institutions to train the large model feature extractor and hypernetwork parameters, achieving continuous model evolution.

Demonstration Logistics

We will show how BC-MRobot performs federated collaborative training and continuous model evolution capabilities when facing changing data and tasks, as well as its application in medical data analysis. We simulate the federated training process on an Ubuntu server equipped with dual RTX 5090 GPUs. A Kuavo 4PRO MaxB Robot serves as the terminal robot for application demonstration. In this demo, Lingshu-7B is used as the multimodal medical diagnosis large model, and CNN models are employed as the diagnostic small models. Different number of layers and parameters will be generated dynamically by the model factory.

BC-MRobot Platform The BC-MRobot platform has the following functionalities:

- Federated collaborative training management encompasses the configuration, initiation, and visualization of training tasks, reflecting the bidirectional optimization process between large models and small models. The model factory achieves continuous model evolution through periodic federated collaborative training.
- Dynamic model customization services provide personalized model customization tailored to client-specific data distributions. The personalized model can achieve higher accuracy compared to a generic model.

BC-MRobot for medical data analysis BC-MRobot identifies a user who waves to it, and automatically navigates to his location, and starts conversations. The user's medical images are matched through his identity. It can capture examination areas directly using its onboard cameras. BC-MRobot supports analyzing of images of fundus, retinal OCT , dermoscopy, ocular appearance, etc. After obtaining images, the robot performs preliminary diagnosis using its local small model, and uploads the images and preliminary diagnostic results to the cloud-based large model for higher accurate analysis. Finally, the system performs multimodal fusion analysis based on images, small model diagnostic results, and consultation processes to provide comprehensive diagnostic, and give further healthcare suggestions.

Conclusion

BC-MRobot can adapt to data feature changes. A bidirectional collaborative optimization mechasim between large model and small model is designed. A model factory is proposed to enable personalized model customization and continuous evolution. This demonstration showcases promising prospects of brain cognition based large and small model collaborative optimization for medical robots.

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.

Danciu, D.-P.; Klawe, F. Z.; Kazarnikov, A.; Femmer, L.; Kostina, E.; Martin-Villalba, A.; and Marciniak-Czochra, A. 2025. Unraveling regulatory feedback mechanisms in adult neurogenesis through mathematical modelling. *npj Systems Biology and Applications*, 11(1): 82.

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.

Scott, J.; Zakerinia, H.; and Lampert, C. H. 2024. PeFLL: Personalized Federated Learning by Learning to Learn. In *ICLR*.