

Dear Editor and Reviewers:

Thank you for your letter and the reviewers' comments on our paper entitled " Revolutionizing gastroenterology and hepatology with artificial intelligence: From precision diagnosis to equitable healthcare through interdisciplinary practice (Manuscript NO. 108021). These valuable comments helped us revise and improve our paper and provided important guidance for our research. We have carefully studied these comments and made corrections for your approval. All the revisions made in response to the comments from the editor and reviewers have been prominently marked in yellow highlight in the revised manuscript. The reviewers' comments are listed below in italicized and blue font, and we have numbered the specific questions. Our responses are given in normal font. Our point-by-point responses to reviewers' comments/questions are given below.

Suggestions from Reviewer 1 and responses from the authors

***Point 1:** While the manuscript presents an extensive review of AI applications in gastroenterology and hepatology, it lacks a clear discussion on the originality of the AI applications introduced. The authors should emphasize specific innovations or unique contributions of AI in this field that distinguish their work from existing literature. For instance, the manuscript could highlight breakthroughs in multimodal data integration, explainable AI systems, or novel ethical frameworks for AI adoption. This would not only strengthen the manuscript's originality but also provide readers with a clearer understanding of how AI is advancing the field beyond conventional methods.*

Response 1: We would like to express our profound gratitude for your invaluable suggestions. As previously mentioned, the manuscript requires enhancement of the discussion of the originality of AI applications. This is a crucial component in improving the quality of the paper. We concur with your proposal that identifying the distinctive innovations of AI technology in the fields of gastroenterology and hepatology is paramount for distinguishing the extant literature and showcasing to readers the state-of-the-art advancements in the discipline.

To address this issue, an explicit statement of original contribution has been incorporated in the introduction: "We expect that this review will provide the latest research progress and unique contributions on AI in gastroenterology and hepatology, break through the original limitations through emerging AI tools or algorithms such as multimodal data integration, interpretable AI systems, or new ethical frameworks adopted by AI, and provide a comprehensive overview for clinical researchers and AI enthusiasts, as well as insights and ideas to further ground the clinical translation of AI."

Thank you again for your professional input, which gave us the opportunity to further optimize the academic depth of the manuscript!

Point 2: *Although the manuscript mentions challenges related to data scarcity and algorithm limitations, it does not provide sufficient detail on how these issues were addressed in the reviewed studies. To address this, the authors should include specific examples or case studies illustrating how researchers or clinicians have mitigated these challenges. For example, they could discuss how federated learning or transfer learning has been applied in specific studies to overcome limited datasets, or how attention mechanisms and heatmaps have been used to enhance the interpretability of AI models. Providing concrete examples would not only improve the scientific rigor of the manuscript but also offer practical insights for readers interested in implementing AI in clinical settings.*

Response 2: Thank you for your insightful and constructive suggestions. We fully acknowledge the importance of incorporating specific cases to enhance both the scientific rigor and practical applicability of the paper. In accordance with your recommendations, we have implemented the following targeted revisions to the original draft:

We have expanded the discussion section in the original text: "In one of the largest federated machine learning (FL) studies to date" and provided a more detailed description: "In one of the largest federated machine learning (FL) studies to date, researchers integrated data from 6,314 patients from 71 healthcare sites across 6 continents to construct an automated tumor boundary detection model. The framework effectively integrates heterogeneous datasets from multiple sources without sharing the original data by weighting and averaging the encrypted model parameters across sites via a central aggregation server. Compared with the model trained only on publicly available datasets, the federated learning model achieves significant breakthroughs in surgical target area detection (33% enhancement) and overall tumor extent identification (23% enhancement), which directly corroborates the critical role of distributed data aggregation in breaking through the data size limitations of a single center. In the study, a pre-training model initialization strategy was adopted to accelerate convergence, and data diversity was gradually improved through phased training (from the initial model at 16 sites to the final consensus model at 71 sites), successfully constructing a cross-regional privacy-preserving collaborative network in rare disease scenarios, which provides a scalable technological paradigm for solving the problem of scarcity of diagnostic resources in low-resource regions. "

Furthermore, we have also cited additional literature to supplement some examples regarding the application of federated learning in disease diagnosis: "In a multicenter study of prostate MRI segmentation, three institutions (University of California, Los Angeles, State University of New York Upstate Medical Center, and National Cancer Institute) integrated their respective private datasets (100 cases of T2-weighted MRI images each) through federated learning, and aggregated the model weights by weighted averaging without sharing the original data, so that the federated model on the external dataset (343 cases) demonstrated significantly

higher performance than single-agency models^[186]. Furthermore, in breast cancer diagnosis, federated learning demonstrates high diagnostic efficacy by integrating data from multiple sources and protecting privacy, enabling model accuracy of 98.9% and 95.3%^[187, 188].”

[186] Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, Xu D, Flores MG, Raman AG, Kulkarni R, Wood BJ, Choyke PL, Priester AM, Marks LS, Raman SS, Enzmann D, Turkbey B, Speier W, Arnold CW. Federated learning improves site performance in multicenter deep learning without data sharing. *Journal of the American Medical Informatics Association* 2021; 28(6): 1259-1264 [DOI: 10.1093/jamia/ocaa341]

[187] Sharma J, Kumar D, Verma R. Secure and Collaborative Breast Cancer Detection Using Federated Learning. 2024 2nd World Conference on Communication & Computing (WCONF), 2024: 1-4

[188] AlSalman H, Al-Rakhami MS, Alfakih T, Hassan MM. Federated Learning Approach for Breast Cancer Detection Based on DCNN. *IEEE Access* 2024; 12: 40114-40138 [DOI: 10.1109/access.2024.3374650]

In response to the fact that no study has yet applied the concept of federated learning to the diagnosis and treatment of gastrointestinal diseases, we propose our own vision: “However, no study has applied federated learning to the diagnosis and treatment of gastrointestinal diseases. For this reason, as shown in Figure 3, we design a federated learning-based collaborative framework, which relies on a central aggregation server and optimizes the encryption parameter ($\Delta\theta_i$) by using the weighted average method, in order to integrate multi-site heterogeneous datasets, and then generates a more generalized consensus model $\theta(k+1)$, which can both ensure privacy and security of gastrointestinal data and provide a scalable technology paradigm for cross-center AI collaboration. process ensures the privacy and security of gastrointestinal data while providing a scalable technical paradigm for cross-center AI collaboration.”

In order to better explain the issue of model interpretability, we cite additional literature for a more detailed case study of Gradient-weighted Class Activation Mapping (Grad-CAM) mentioned in the paper, and we hope that this will help readers better understand the complex modeling decision-making process: “Grad-CAM generates a heat map to visualize the basis of model decision-making by quantifying the gradient of the target category probability relative to the final convolutional layer feature map. Its implementation is divided into three steps: first, calculate the gradient of the target category scores with respect to a specific convolutional layer feature map to obtain the importance weights of each channel; second, perform channel-weighted summation of the feature map to generate a coarse-grained heatmap; and finally, upsample the heatmap to the input image size by bilinear interpolation to highlight the key regions. The method does not require modification of the model structure and is suitable for all types of convolutional neural networks. In biomedical research, Grad-CAM can localize molecular pathways driving classification (e.g., the KRAS signaling pathway in HPV status prediction) or regions of differential brain function (e.g., the default mode network of schizophrenic patients) to provide biologically plausible validation of model decisions. Its advantage lies in combining gradient information with spatial features to generate intuitive and interpretable heat maps while maintaining high resolution and class specificity, making it an important tool for connecting deep learning “black boxes” with domain knowledge^[177, 178].”

[177] Ghassemi N, Shoeibi A, Khodatars M, Heras J, Rahimi A, Zare A, Zhang YD, Pachori RB, Gorriz JM. Automatic diagnosis of COVID-19 from CT images using CycleGAN and transfer learning. *Appl Soft Comput* 2023; 144: 110511 [PMID: 37346824 PMCID: PMC10263244 DOI: 10.1016/j.asoc.2023.110511]

[178] Lin QH, Niu YW, Sui J, Zhao WD, Zhuo C, Calhoun VD. SSPNet: An interpretable 3D-CNN for classification of schizophrenia using phase maps of resting-state complex-valued fMRI data. *Med Image Anal* 2022; 79: 102430 [PMID: 35397470 DOI: 10.1016/j.media.2022.102430]

These concrete examples collectively address the scientific rigor concerns while providing actionable insights for clinical AI implementation. We thank the reviewer for this valuable suggestion to improve the manuscript's translational relevance.

Point 3: *While the manuscript discusses the potential of AI to improve healthcare equity, it does not adequately address the practical barriers to implementation in low-resource settings. The authors should expand this section to explore how AI tools can be adapted for use in resource-limited environments, such as through lightweight models, edge computing, or federated learning frameworks. Citing examples of AI applications that have been successfully deployed in such settings would further strengthen the manuscript's relevance and practicality.*

Response 3: We sincerely appreciate the reviewer's insightful feedback. As suggested, we have substantially expanded the discussion on overcoming practical barriers in low-resource settings and incorporated concrete examples. The revisions include:

To address the arithmetic limitations of low-resource environments, we further refine the technical features of the presented PSE-CNN-PCA-DELM lightweight framework based on the original article as follows: "An interpretable three-stage lightweight deep learning framework (PSE-CNN-PCA-DELM) achieves 97.24% classification accuracy and 99.38% ROC-AUC metrics in a multi-level classification task of gastrointestinal diseases by integrating XAI (interpretable AI) technologies, such as Grad-CAM visualization and heat map analysis. With a model volume of only 14.88MB and extremely fast inference performance of 59ms, the framework breaks through the dependence of traditional AI models on high-performance arithmetic, and can be efficiently deployed in edge devices such as portable ultrasound machines and low arithmetic terminals, which is especially suitable for primary care scenarios and remote areas with a lack of resources, and can help narrow the geographic disparity of healthcare resources through the sinking of transparent and trusted AI-assisted diagnostic capabilities, contributing to promote AI-driven healthcare fairness provides a practical technical solution. "

Furthermore, we have rewritten the last paragraph of the discussion to discuss in more detail how to apply AI tools to resource-limited environments, proposing a technology integration strategy of "lightweighting + localization + privacy protection", as follows: With the clinical validation and scale deployment of AI-assisted decision-making tools (e.g., the UC-SCALE standardized scoring system), complex disease screening technologies are being embedded into different tiers of the healthcare system in an algorithmically standardized manner to promote equal coverage of diagnostic capabilities from the bottom of the technology.

In resource-constrained scenarios, the technology integration strategy of “lightweight + localization + privacy protection” is especially critical: lightweight neural network combined with edge computing architecture realizes offline inference, ensuring stable operation on terminal devices in underdeveloped regions; the federated learning framework supports the construction of cross-institutional diagnosis and treatment knowledge, avoiding the risk of leaking the original data; and the rare case data synthesized by the generative adversarial network complements the regional database. Together, the three can build an “end - side - cloud” collaborative computing architecture to provide sustainable technology empowerment for primary care. These technological advances not only improve the efficiency of diagnosis and treatment, but also reconstruct the medical ecology in terms of deep logic: through algorithmic optimization of resource allocation model, innovation of medical knowledge discovery mechanism, and promotion of adaptive evolution of ethical framework, AI is upgrading from an auxiliary tool to an infrastructure for health equity. In the future, with the improvement of the technology-ethics synergy, machine learning is expected to become the core driving force to break through the plight of global health inequality, and promote the meta-goal of “health equity” from theoretical conception to practical realization.

Once again, thank you for your expert advice; these revisions significantly enhance the real-world relevance and application of the paper.

Suggestions from Reviewer 2 and responses from the authors

It is a great honor to receive the high recognition and positive comments from the reviewers on our paper! If any information needs to be supplemented or adjustments need to be made in the subsequent process, we will fully cooperate.

We put forth our best efforts to improve the manuscript and made some additional changes. These modifications did not change the content or framework of the paper and have not been explicitly listed here, but they are highlighted in yellow in the revised manuscript. We greatly appreciate the Editors/Reviewers’ work and hope that these corrections will meet with your approval.

Once again, thank you very much for your comments and suggestions.

Sincerely,
Fang Wang,
College of Basic Medical Sciences
Jilin University