Peer Review File

Reviewer A:
This paper is far away from being on the level of an original article. This manuscript to an opinion or commentary. Please think about resubmitting your work as a opinion. But if so, please add clear discussion and else.

Overall Response: We thank the reviewer for their review. We also appreciate the comments for clarification that will help optimize this manuscript for the intended audience, and agree that this work is better suited as an opinion article, however this is not an available article type. We have now decided to submit this article as a Brief Report -"Manuscripts containing pertinent and interesting observations and reports on new observations or studies that do not warrant publication as a full research article will be considered for the Brief Reports."

Reviewer B:
The authors have presented a case study exhibiting the potential of ChatGPT assisting coding for medical machine learning. This is an important topic, but I have a few suggestions which may be useful to incorporate before the manuscript can be published.

Comments:
1. Avoid superfluous judgmental words such as "stunning", "revolution".
2. There is currently little to no discussion about existing applications of ChatGPT (or other LLMs) in medicine. E.g. primary evidence (https://doi.org/10.2196/46599; https://doi.org/10.1001/jamainternmed.2023.1838) and other relevant discussion pieces in NEJM, JAMA, Nature Medicine, Journal of the Royal Society of Medicine.
3. Similarly, there is little to no mention of existing applications of GPT-4 for assisting coders e.g. GitHub Copilot.
4. The structure of the paper is unclear--if this is a case report than this should be stated explicitly. If this is supposed to be original research, there should be more rigorous methods for conducting, analysing, and reporting experiments, and these should be detailed in a methods section. The Editor and journal guidelines are best for specific guidance about format.
5. Some of the figures are difficult to read in the format provided--ensure that they are of sufficient resolution to be read clearly.
6. The challenges and limitations section is extremely brief and non-specific. What are the specific barriers to implementation regarding medical machine learning development, and how might they be overcome?

Response: Thank you for the comment.
1. We have deleted the words "stunning" and "revolution".

2. Agreed, this increased discussion is essential to include. We have added line 70-74 to address this "LLMs are rapidly approaching human-level performance, with ChatGPT successfully completing the Royal College of General Practitioners Applied Knowledge Test with an average score of 60.17%.[3] In another recent study, ChatGPT was shown to be able to respond to patient questions from a social media forum with higher levels of empathy and quality than the responses provided by physicians."

3. Agreed, we have added lines 131-133: "GitHub has recently introduced "Copilot Chat", a built-in ChatGPT-like experience to help coders by providing in-depth explanations and analysis."

4. We have now decided to submit this article as a Brief Report -"Manuscripts containing pertinent and interesting observations and reports on new observations or studies that do not warrant publication as a full research article will be considered for the Brief Reports."

5. All figures are of maximum possible resolution. We are happy to hear further feedback about which figure is difficult to read and can work to improve on this.

6. Added lines 139-143 "ChatGPT must also be used in accordance with local healthcare regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. To ensure this is occurring, patient protected health information must be stored and transmitting securely, while following strict authentication protocols. Compliance with these regulations must also be regularly assessed.".

Reviewer C:
General comments
I appreciate the opportunity to review this article on Bridging Artificial Intelligence in Medicine with GPT Technology. The manuscript is well written and structured, but I have a few suggestions for improvement.

Specific comments
Major comments
1. "ChatGPT version and date": It is important to provide specific details about the version of the ChatGPT model used in the study. This information is critical as different versions may have different capabilities, leading to variance in performance. The date when the answers were generated is also important, as AI models are continuously updated, and performance may vary over time.
2. "Selection of the AI model": Please provide the scientific reasons to choose convolutional neural networks.
3. "ChatGPT's prompt": Please provide the scientific reasons whether this is sufficient to evaluate its scientific method effectively.
4. "Title": Please elaborate the title of the manuscript. It could be more informative and better structured. They suggest that it should accurately represent the content of the article, and if it refers to "medicine" in general, the study should encompass more than one specialty, not only ophthalmology.

5. "Evaluating ChatGPT's answers'': The authors should provide a clear and comprehensive description of the criteria used to evaluate the responses generated by ChatGPT.
6. "Specialized training and reinforcement": Please provide the information about the ChatGPT model received any specific training or reinforcement, particularly in the context of infectious and tropical diseases. This information will help the reader understand if the model's performance is generalizable or specifically suited for this application.
7. "Limitations of ChatGPT in a clinical setting": The authors should address potential limitations of using ChatGPT in a clinical setting, especially regarding compliance with the Health Insurance Portability and Accountability Act (HIPAA).

Overall response: Thank you for the kind review and for your comments to optimize this opinion paper.
1. Agreed, information regarding ChatGPT version and date is essential to be included. This information has now been added. Lines 90-91 "All responses in this paper were generated with GPT-3.5 in early March, 2023 prior to the release of GPT-4"
2. A CNN was chosen as our group has specific expertise within this area. While other models could have been chosen, we feel this choice was ideal.
3. Exact prompts were provided in this paper that we gave ChatGPT to help others reproduce similar results if they replicated these methods. Reproducibility is an essential part of the scientific method.
4. While we understand that our paper does not show examples of GPT technology bridging the AI-clinician gap in every medical specialty (there are hundreds with the inclusion of subspecialties), by following the methods set out in this paper, this should be easily applicable to any area of medicine.
5. We agree that the inclusion of clear criteria would be ideal, however the criteria for evaluating code is too broad to distill into 2 criteria points like "specificity" and "accuracy". Also this paper is targeted at the level that it could be understood by clinicians, to bridge the AI-clinician gap, so further explanations relating to coding intricacies would not be fully appreciated by this audience.
6. No specific training or reinforcement was provided to ChatGPT so this was not mentioned, however this is a very interesting area and we will consider this topic for future papers.
7. Added lines 139-143 "ChatGPT must also be used in accordance with local healthcare regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. To ensure this is occurring, patient protected health information must be stored and transmitting securely, while following strict authentication protocols. Compliance with these regulations must also be regularly assessed.".