





Chat-GPT on brain tumors: An examination of Artificial Intelligence/Machine Learning's ability to provide diagnoses and treatment plans for example neuro-oncology cases

Giovanni Kozel ^a, Muhammet Enes Gurses ^b  , Neslihan Nisa Gecici ^c, Elif Gökalp ^d, Siyar Bahadır ^e, Martin A. Merenzon ^f, Ashish H. Shah ^b, Ricardo J. Komotar ^b, Michael E. Ivan ^b

^a Warren Alpert Medical School, Brown University, Providence, RI, USA

^b Department of Neurosurgery, Miller School of Medicine, University of Miami, Miami, FL, USA

^c Hacettepe University School of Medicine, Ankara, Turkey

^d Ankara University School of Medicine, Ankara, Turkey

^e Feinstein Institute, New York, USA


^f Department of Neurosurgery, Yale University School of Medicine, New Haven, CT, USA

Received 5 March 2024, Accepted 7 March 2024, Available online 9 March 2024, Version of Record 19 March 2024.

 What do these dates mean?



Show less 

 Share  Cite

<https://doi.org/10.1016/j.clineuro.2024.108238> 

[Get rights and content](#) 

Highlights

- ChatGPT-4 demonstrates improved accuracy in diagnosing brain tumors compared to ChatGPT-3.5, with an overall accuracy of 85% for diagnosis and 75% for treatment plan.
- The average scores on a ten point scale given by twenty independent neurosurgeons for ChatGPT-4's diagnoses and treatment plans were 8.3 and 8.5 out of 10, respectively.

Abstract

Objective

Assess the capabilities of ChatGPT-3.5 and 4 to provide accurate diagnoses, treatment options, and treatment plans for brain tumors in example neuro-oncology cases.

Methods

ChatGPT-3.5 and 4 were provided with twenty example neuro-oncology cases of brain tumors, all selected from medical textbooks. The artificial intelligence programs were asked to give a diagnosis, treatment option, and treatment plan for each of these twenty example cases. Team members first determined in which cases ChatGPT-3.5 and 4 provided the correct diagnosis or treatment plan. Twenty neurosurgeons from the researchers' institution then independently rated the diagnoses, treatment options, and treatment plans provided by both artificial intelligence programs for each of the twenty example cases, on a scale of one to ten, with ten being the highest score. To determine whether the difference between the scores of ChatGPT-3.5 and 4 was statistically significant, a paired t-test was conducted for the average scores given to the programs for each example case.

Results

In the initial analysis of correct responses, ChatGPT-4 had an accuracy of 85% for its diagnoses of example brain tumors and an accuracy of 75% for its provided treatment plans, while ChatGPT-3.5 only had an accuracy of 65% and 10%, respectively. The average scores given by the twenty independent neurosurgeons to ChatGPT-4 for its accuracy of diagnosis, provided treatment options, and provided treatment plan were 8.3, 8.4, and 8.5 out of 10, respectively, while ChatGPT-3.5's average scores for these categories of assessment were 5.9, 5.7, and 5.7. These differences in average score are statistically significant on a paired t-test, with a p-value of less than 0.001 for each difference.

Conclusions

ChatGPT-4 demonstrates great promise as a diagnostic tool for brain tumors in neuro-oncology, as attested to by the program's performance in this study and its assessment by surveyed neurosurgeon reviewers.

Introduction

ChatGPT is a free-to-use artificial intelligence program which uses human language to provide text-based outputs to commands and questions provided by users. Published on November 30, 2022, this chatbot stands as the inheritor of a decade of concentrated research in machine learning, a field which is currently being rigorously tested for applications to neuro-oncology [1]. Emblem et al. have already demonstrated that a support vector machine (SVM) model supplemented with MRI imaging data and a perfusion MRI/rCBV map can predict the survival time of glioblastoma patients more precisely than expert neuroradiologist assessment in the six-month to three-year time frame [2]. Oermann et al. have found that artificial intelligence can perform equally or superiorly to established statistical grading systems for cancer metastasis, such as the Graded Prognostic Assessment, Golden Grading Scale (GPA/GGS), and Logistic Regression (LR) [3].

However, studies such as these have previously been built around large language learning models (LLMs), programs which are individually designed and trained by researchers to utilize certain key variables in order to deliver a desired output. ChatGPT has the potential to widen the application of artificial intelligence in neuro-oncology due to its incredible elasticity as a piece of software. ChatGPT responds to questions without needing any special configuration or additional training from the user. The artificial instead draws upon vast

quantities of data in order to generate an easily-digestible response which is written in the style of human conversation. The question of interest now is to determine whether ChatGPT is as accurate as it is easy to engage with. For the field of neuro-oncology, the function of greatest interest for this program is its ability to diagnose brain tumors. Many researchers hope that the program will not only be able to aid tumor boards for accelerated and automated management suggestions [4], but will also allow for the easier treatment of underserved populations in which there is a shortage of trained neuro-oncologists, as a repetitive and time-consuming task is shifted from physicians to artificial intelligence.

To this end, ChatGPT was given the clinical status, surgical outcome, textual imaging information, and immune-pathology of ten real patients with brain gliomas by Haemmerli et al., who then asked the program to choose a course of adjuvant treatment and regimen for each patient while also considering their respective functional status. Expert physicians who reviewed ChatGPT's responses concluded that the program performed poorly in classifying types of gliomas in its diagnoses but was successful in providing good adjuvant treatment recommendations for patients [4]. Since an accurate diagnosis must always precede proper treatment, it is the goal of this study to expand on Haemmerli et al.'s investigation by testing the capabilities of ChatGPT in diagnosing example brain tumors and providing treatment plans and treatment options for these tumors. To that end, it was found that ChatGPT-4 demonstrated potential as an assistive diagnostic tool for neuro-oncologists in the development of diagnoses, treatment options, and treatment plans for patients.

Section snippets

Case selection and ChatGPT dialogue

Study design is shown in Fig. 1. Twenty unique neuro-oncology example cases for seven malignant and thirteen benign brain tumors were selected from medical literature [5], [6] by team members and are detailed in Table 1. It was decided by team members that example cases, rather than actual patient data, would be used for two reasons: first, example cases have a predetermined "correct" answer for each of four command questions given to ChatGPT. Therefore, the performance of the artificial...

Results

As detailed in Tables 2 and 3, on initial review, ChatGPT-3.5 correctly diagnosed thirteen of the twenty tumors, four of the seven malignant tumors and nine of the thirteen benign tumors, and provided the correct treatment plan for two of the twenty tumors, both of which were benign tumors. ChatGPT-4 performed better than the previous version of the program, correctly diagnosing seventeen of the twenty tumors, five of the seven malignant tumors and twelve of the thirteen tumors, and providing...

The abilities of artificial intelligence

Currently, user-friendly artificial intelligence programs in their most advanced form, as represented by ChatGPT 4, can correctly diagnose brain tumors in 85% of example cases, with a higher accuracy for benign tumors (92.3%) than malignant tumors (71.4%), and can correctly provide a treatment plan in 75% of cases, with a higher accuracy for benign tumors (84.6%) than malignant tumors (57.1%). In a broad review by neurosurgeons, this latest form of the program is also highly rated, scoring an...

Conclusion

ChatGPT has demonstrated a potential for clinical applications to the provision of diagnoses, treatment options, and treatment plans for brain tumors within the field of neuro-oncology. However, there are several issues to this program's integration within clinical practice, beginning with the shortcomings and errors of the technology itself. Incorrect predictions must first be addressed, especially among malignant tumors, followed by a clear establishment of the role of artificial intelligence ...

Statement of Ethics

According to the research policy of the institution where this research was conducted, ethics committee approval is not required for this type of study. The participants and any identifiable individuals consented to the publication of his/her image. Copies of conversations held with the artificial intelligence programs consulted may be provided upon request....

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors....

CRedit authorship contribution statement

Michael Ivan MD: Project administration, Supervision. **Ricardo Komotar:** Project administration, Supervision. **Ashish Shah:** Project administration, Supervision. **Martin Merenzon:** Project administration, Supervision. **Siyar Bahadir:** Data curation, Formal analysis, Investigation, Software, Validation, Writing – review & editing. **Elif Gökalp:** Project administration, Supervision. **Neslihan Gecici:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration,...

Declaration of Competing Interest

The authors report no personal, financial, or institutional interest in the materials or devices described in this article....

[Recommended articles](#)

references (13)

M.M. Bonsanto *et al.*

Artificial intelligence in neurosurgery

Surgeon (2020)

K.E. Emblem

Machine learning in preoperative glioma MRI: survival associations by perfusion-based support vector machine outperforms traditional MRI

J. Magn. Reson. Imaging (2014)

E.K. Oermann

Predicting survival in patients with brain metastases treated with radiosurgery using artificial neural networks

Neurosurgery (2013)

J. Haemmerli

ChatGPT in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board?

BMJ Health Care Inform. (2023)

J.S. Citow *et al.*

Neurosurgery Oral Board Review(2019)

R. Nader *et al.*

Neurosurgery Case Review: Questions and Answers(2020)

There are more references available in the full text version of this article.

Cited by (2)

Chat-GPT on brain tumors: Comment

2024, Clinical Neurology and Neurosurgery

Letter: Chat-GPT on brain tumors: An examination of Artificial Intelligence/Machine Learning's ability to provide diagnoses and treatment plans for example neuro-oncology cases

2024, Clinical Neurology and Neurosurgery

[View full text](#)

© 2024 Elsevier B.V. All rights reserved.



All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

