Can Artificial Intelligence Correctly Differentiate between Vagal Schwannomas and Sympathetic Chain Schwannomas?

Anthony M. Saad, BA[1], David Herz, BS[1], Nicole Margiotta, BS[1], Sree Chinta, BS[1], Emma Thompson, BA[1], Andrey Filimonov, MD, PharmD[1], Ghayoour S. Mir, DO[1]

[1]Department of Otolaryngology, Rutgers New Jersey Medical School, Newark, NJ

Abstract

Objective: To determine the ability of artificial intelligence (AI) ChatGPT to accurately differentiate between vagal and carotid schwannomas from contrast-enhanced MRI images and CT images.

Study Design: Comparative analysis.

Methods: A variety of peer-reviewed studies from reputable journals in otolaryngology were accessed. Four MRI images, two of vagal schwannomas and two of sympathetic chain schwannomas, and four CT images, two of each diagnosis, were extracted. Each image was independently entered into ChatGPT-4o. The AI model was then prompted to provide the correct diagnosis. Accuracy of responses was graded based on the designations provided in the original sources. Univariate analyses were used to identify statistical associations.

Results: Eight images, four of each diagnosis, were given to ChatGPT. The AI model had an overall accuracy of 62.25%. Notably, ChatGPT diagnosed vagal schwannoma (75.00%) at a greater rate than sympathetic chain schwannoma (50.00%). A similar trend was observed when data were stratified by image modality type, as GPT correctly diagnosed three of the MRI images (75.00%) and two (50.00%) of the CT images. On Fisher's exact test, there was no significant difference between overall accuracy of MRI images compared to CT images ($p > 0.05$).

Conclusions: While ChatGPT is able to recognize schwannomas in CT and MRI, the AI model does not have great accuracy in differentiating between vagal schwannomas and sympathetic chain schwannomas. Further refinement is necessary for reliable clinical application.

Introduction

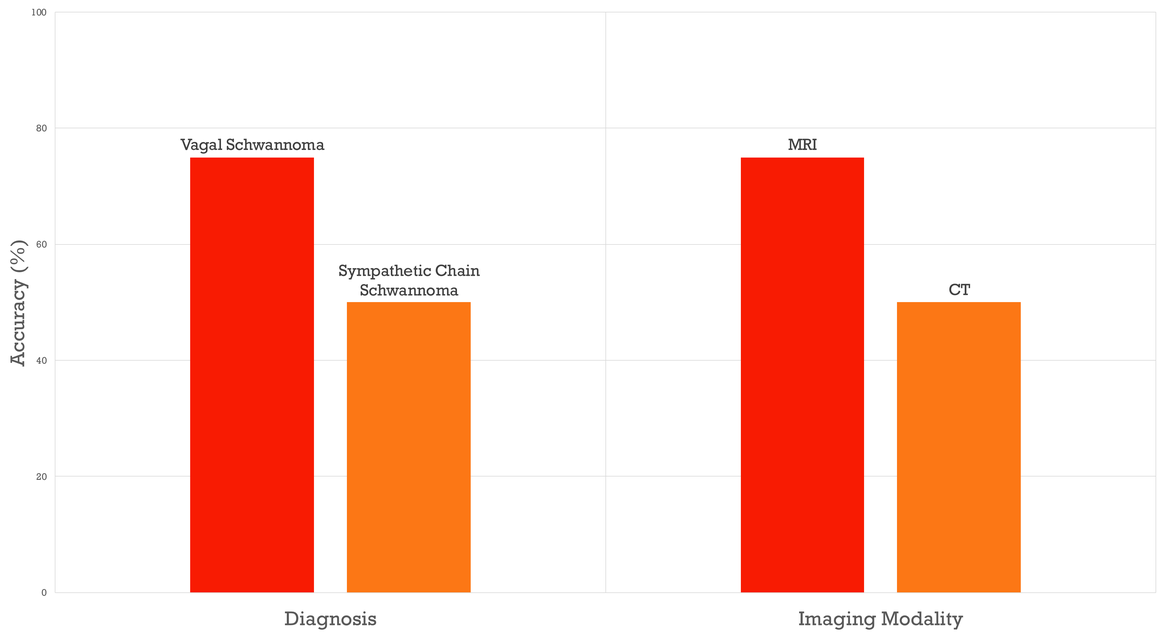
- ChatGPT is an artificial intelligence system developed by OpenAI that uses large-scale machine learning to understand and generate human-like outputs. While best known as a language model, newer versions have the advanced capability to interpret and respond to images, making it a versatile tool for both text- and image-based tasks.[1]
- Many studies show promising applications of OpenAI's GPT-4 (ChatGPT) and other large language models (LLMs) in the field of medicine, with successful performance in diverse tasks including image-based diagnostics, clinical analysis, treatment planning, and open-ended clinical question analyses, even achieving at or near-passing scores on the USMLE Step 1, Step 2CK, and Step 3 examinations without prior training or reinforcement.[1-3]
- In the field of otolaryngology, LLMs such as ChatGPT have been shown to differ in their ability to interpret clinical knowledge and generate comprehensive treatment plans. Among them, ChatGPT is recognized as one of the stronger engines, demonstrating a solid grasp of both the foundational and detailed concepts in the field.[4]
- ChatGPT has achieved variable success in the evaluation of head and neck cancer. In analyzing images and corresponding text of recurrent laryngeal cancers, ChatGPT was able to provide accurate TNM classification and clinical staging according to current guidelines.[5] According to another study, the AI engine overestimated the extent and severity of lymph node metastasis in head and neck cancers.[6]
- Given the heterogenous success of ChatGPT in otolaryngological cancer detection and analysis, further research is warranted to establish the safety and efficacy of potential integration of ChatGPT and other LLMs into current medical practice.
- The objective of this study is to assess the accuracy of ChatGPT-4o in accurately differentiating between vagal and sympathetic chain schwannomas using contrast-enhanced MRI images and CT images from peer-reviewed otolaryngological studies.

Study Design

- Peer-reviewed otolaryngology journals were reviewed to identify eligible imaging cases.
- Image set: 8 total images (4 contrast-enhanced MRI, 4 CT).
  - Contrast-enhanced MRI: 2 vagal schwannomas, 2 sympathetic chain schwannomas
  - CT: 2 vagal schwannomas, 2 sympathetic chain schwannomas
- Each image was independently uploaded into ChatGPT-4o (OpenAI, May 2024 release) for diagnostic interpretation.
- Model outputs were compared with the reference diagnoses reported in the original studies.
- Accuracy was defined as concordance between model responses and source designations.
- Univariate analyses, including Fisher's exact test, were performed to assess statistical associations.

Results

Figure 1:
ChatGPT Accuracy in Recognizing Vagal vs. Sympathetic Chain Schwannomas from MRI and CT images



Discussion

- In this image set, ChatGPT demonstrated moderate overall accuracy (62.25%) in successfully identifying vagal and sympathetic chain schwannomas. Performance was stronger for vagal schwannomas (75.00%) when compared to sympathetic chain schwannomas, suggesting possible diagnostic bias. This outcome aligns with the heterogenous LLM performance seen across otolaryngological diagnoses. For example, correct diagnostic accuracy ranged from 38% to 86% in one study of otolaryngological pathologies stratified by subspecialty.[7]
- By modality, contrast-enhanced MRI (75%) outperformed CT (50%), though this difference was not statistically significant (> 0.05). These findings underscore both the potential utility of AI in otolaryngological radiologic interpretation, as well as its current limitations.
- One meta-analysis showed that in 91 studies of the utility of artificial intelligence in otolaryngology, generative AI performed with high accuracy in education (83.0%) and imaging tasks (84.9%), but underperformed in clinical decision support (67.1%).2 Another study demonstrated ChatGPT's strong performance on open-ended otolaryngology questions (75% accuracy across 3 trials) while also revealing frequent issues with self-contradiction and hallucinations.[3]
- Overall, the variability in accuracy suggests that ChatGPT's diagnostic reasoning is not yet optimized for nuanced imaging distinctions, reinforcing the need for refinement before reliable clinical application.

Conclusion

- ChatGPT can recognize schwannoma patterns on CT (50.00%) and MRI (75.00%) in curated examples; however, the model does not reliably differentiate vagal vs. sympathetic chain schwannomas with high accuracy. In this image set, ChatGPT displayed 62.25% accuracy, indicating the need for improved performance before implementation into clinical medicine.
- Current evidence suggests that LLMs like ChatGPT are not reliable enough for autonomous diagnostic use in nuanced radiologic differentials. However, they show promise in supplementing current medical practice as adjunct tools, with potential applications in preliminary triage, education, and workflow support. Continued research, AI model refinement, and policy development are necessary for safe and effective implementation into clinical medicine.8
- Until reproducible, high accuracy is achieved, diligent human oversight remains essential for clinical application of LLMs such as ChatGPT.
- Future studies should assess ChatGPT's performance across broader radiological pathologies to better understand its generalizability to everyday imaging practice.

Contact

Ghayoour S. Mir, DO
Department of Otolaryngology - Head & Neck Surgery
Rutgers New Jersey Medical School
mirgs@njms.rutgers.edu

References

1. Kung, Tiffany et al. "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models." PLOS Digit Health. 2023 Feb 9;2(2):e0000198. doi: 10.1371/journal.pdig.0000198. PMID: 36812645; PMCID: PMC9931230.
2. Hack, Sholem et al. "Performance of generative AI across ENT tasks: A systematic review and meta-analysis." Auris Nasus Larynx. 2025 Sep 4;52(5):585-596. doi: 10.1016/j.anl.2025.08.010. Epub ahead of print. PMID: 40912131.
3. Long, Cai et al. "A Novel Evaluation Model for Assessing ChatGPT on Otolaryngology–Head and Neck Surgery Certification Examinations: Performance Study." JMIR Medical Education, vol. 10, 2024, ISSN 2369-3762. ScienceDirect Online Library, https://doi.org/10.2196/49970.
4. Merlino, Dante et al. "Comparative Assessment of Otolaryngology Knowledge Among Large Language Models." Laryngoscope. 2025 Feb;135(2):629-634. doi: 10.1002/lary.31781. Epub 2024 Sep 21. PMID: 39305216.
5. Yan, Danfang et al. "Evaluating ChatGPT's recommendations for systematic treatment decisions in recurrent or metastatic head and neck squamous cell carcinoma: Perspectives from experts and junior doctors." Int J Cancer. 2025 Nov 1;157(9):1888-1897. doi: 10.1002/ijc.70001. Epub 2025 Jul 19. PMID: 40682765; PMCID: PMC12407031.
6. Schmidl, Benedikt et al. "Assessing the value of artificial intelligence-based image analysis for pre-operative surgical planning of neck dissections and iENE detection in head and neck cancer patients." Discover Oncology. 2025 May 30;16(1):956. doi: 10.1007/s12672-025-02798-4. PMID: 40445459; PMCID: PMC12125457.
7. Lechien, Jérôme et al. "Performance and Consistency of ChatGPT-4 Versus Otolaryngologists: A Clinical Case Series. Otolaryngology - Head and Neck Surgery. 2024 Jun;170(6):1519-1526. doi: 10.1002/ohn.759. Epub 2024 Apr 9. PMID: 38591726.
8. Toh, Yvana. "ChatGPT and the future of radiology reporting." European Journal of Radiology Artificial Intelligence, vol. 2, 2025, 100007, ISSN 3050-5771. ScienceDirect Online Library, https://doi.org/10.1016/j.ejrai.2025.100007.