

連載

放射線科では AI Chatは こう使う!

第6回

ChatGPTを用いた研究と 今後の課題

立川裕之

大阪公立大学大学院医学研究科 放射線診断学・IVR学



はじめに

これまでの「放射線科ではAI Chatはこう使う」の企画で紹介されているように、放射線分野の臨床・研究における様々な状況において、ChatGPTなどのLarge Language Models (LLMs)の有用性が知られている。ただし、LLMsを適切に使用するためには、その問題点や限界についても知る必要がある。

我々の施設ではこれまでにChatGPTの有用性を調べる研究を複数行っており、その紹介を交えながらLLMsの利用例とそのピットフォールについて概説する。

画像所見から鑑別診断・ 最終診断を行う研究

1. 「Diagnostic Performance of ChatGPT from Patient History and Imaging Findings on the Diagnosis Please Quizzes」
Radiology, 2023. doi: 10.1148/radiol.231040¹⁾

この研究ではRadiology誌の教育用クイズ「Diagnosis Please」を用いて、GPT-4ベースのChatGPTの診断補助としての潜在能力を評価した。過去に掲載された313例の症例を用い、ChatGPTに患者の病歴と解答で提示される画像所見を入力し、鑑別診断と最終診断の出力を依頼した(図1)。

患者の病歴と画像所見からの鑑別診断の精度は61%、最終診断の精度は54%であった。専門分野によって診断精度に差が

あり、心血管領域では最も高く79%、骨軟部領域では最低の42%の精度を示した。

「Diagnosis Please」では、非常にまれな疾患や非典型的で診断が難しい疾患が多く出題されているが、分野による差はあるものの、ChatGPTは比較的高い診断精度(一般的な放射線科医よりも)を示した。このことから、LLMsに病歴と画像所見を適切に入力することで、強力な診断補助ツールになることが示された。

2. 「Accuracy of ChatGPT generated diagnosis from patient's medical history and imaging findings in neuroradiology cases」
Neuroradiology, 2024. doi: 10.1007/s00234-023-03252-4²⁾

この研究は、American Journal of Neuroradiology誌の教育症例集である「Case of the Week」から神経放射線分野の100例のデータを連続して収集し、患者の病歴と画像所見をGPT-4ベースのChatGPTに入力し、鑑別診断と最終診断を様々な疾患病因にわたって評価した。

脳病変全体の鑑別診断の精度は62%、最終診断の精度は51%であった。中枢神経系の疾患を腫瘍群と非腫瘍群に分けて比較した場合、非腫瘍群で診断精度が高いことが示された。特に、代謝性疾患、変性疾患、および脳血管性疾患では最終診断の精度が高かった。

中枢神経領域は全身の中で最も疾患の種類の多い領域のため、苦手とする放射線科医も多いが、ChatGPTは臨床現場における貴重な診断支援ツールとしての潜在能力を有することが示された。しかし、中枢神経系腫瘍では非腫瘍性疾患に比べて診断精度は