Analysis of Large Language Models for Centralized Processing of Medical Data

Vilayat Valiyev[1], Nuran Abdullayev[2], Cavid Abbasli[3]
[1]Azerbaijan Technical University, Baku, Azerbaijan
[2] Department of Radiology and Neuroradiology, GFO Clinics Troisdorf, Academic Hospital of the Friedrich-Wilhelms-University Bonn, Troisdorf, Germany
[3]Institute of Biomedical Engineering, Azerbaijan Technical University, Baku, Azerbaijan
[1]rector@aztu.edu.az, [2]nuran.abdullayev@gfo-kliniken-troisdorf.de, [3]cavid.abbasli@aztu.edu.az

Abstract. **With the rapid expansion of the medical information ecosystem, systematic processing of large-scale and heterogeneous medical data is essential for optimizing diagnosis, monitoring, and treatment protocols for complex neurological and oncological conditions such as tumor pathologies and ischemic stroke. Accurate evaluation of these diseases relies on the integration of structured and unstructured data sources, including electronic health records, genomic profiles, radiological images, and laboratory results. Transformer-based Large Language Models (LLMs) have significantly advanced natural language processing (NLP), enabling semantic interpretation of clinical texts. These models support the contextual understanding of histopathological descriptions, clinical protocols, contributing to the refinement of diagnostic algorithms in oncology and stroke. This article examines the architectural features, benefits, limitations, and future potential of LLMs in centralized medical data analysis. The aim is to enhance the role of LLMs as accurate analytical tools in clinical decision-making and to foster the development of personalized medical services**.

Keywords: Large Language Models (LLMs), tumor pathologies, ischemic stroke, multimodal medical data, clinical decision support systems

I. INTRODUCTION

Breast cancer is the most common type of cancer among women worldwide, with approximately 2.3 million new cases diagnosed in 2022 [1]. As breast cancer treatments become increasingly complex and the volume of medical literature and clinical guidelines continues to grow, artificial intelligence (AI) tools can assist healthcare professionals in synthesizing information and making informed treatment decisions.

Ischemic stroke is one of the leading causes of death and disability globally. More than 7.6 million people suffer ischemia-related strokes, accounting for approximately 70% of all stroke cases worldwide [2]. Patients often present to the hospital days after the onset of neurological symptoms related to ischemic stroke. Timely identification of stroke onset is critical for therapeutic decision-making, as brain tissue preservation is only possible within a narrow therapeutic window.

In the absence of perfusion imaging, certain magnetic resonance imaging (MRI) findings can help estimate the timing of ischemic strokes. Sequence-specific MRI techniques—particularly apparent diffusion coefficient (ADC) mapping, diffusion-weighted imaging (DWI), and fluid-attenuated inversion recovery (FLAIR) sequences—are useful in estimating the relative age of a stroke.

The use of medical imaging, particularly MRI, continues to rise across healthcare systems and is expected to increase further in the coming years. However, this growing demand for imaging is unlikely to be met with a proportional increase in the number of radiologists. Consequently, radiology departments may face increased pressure, leading to prolonged response times and a higher rate of diagnostic errors. To address these challenges, AI is increasingly being adopted to streamline radiology workflows, especially in time-sensitive domains such as stroke diagnosis [3-7].

Given these challenges and opportunities, large language models (LLMs) integrated into platforms such as **ChatGPT** by OpenAI have demonstrated potential in understanding, processing, and generating clinical recommendations based on medical data across various fields. **GPT-4**, in particular, can generate and comprehend human-like text and has been trained on a broad range of journals and online content. A significant advancement of GPT-4 over its predecessors is its ability to analyze images and PDF files, which is particularly beneficial for processing large and structured medical datasets.

One of GPT-4’s most important improvements is its capacity to interpret and analyze medical images—a feature that is especially relevant in **neuroradiology**, where multiple MRI sequences are routinely used. This raises the question: *Can GPT-4 assist in time-critical and nuanced tasks such as estimating the onset time of a stroke by detecting subtle temporal changes in tissue appearance and structure?*

To enhance GPT-4’s capabilities, it can be trained to recognize visual characteristics of **ischemic stroke lesions** and **mammography findings** at various stages. This is made possible through **Retrieval-Augmented Generation (RAG)**, a method that incorporates external resources (such as online datasets and relevant journal content) into the model's training and reasoning process.

In November 2023, OpenAI introduced **custom GPT-4 models**—specialized versions of GPT-4 fine-tuned for specific tasks. In this study, we analyze two such RAG-enhanced models: **NeuroRadReport** (focused on ischemic stroke imaging) and **MammaBoardGPT** (focused on breast cancer imaging). To improve MammaBoardGPT's performance, we

DOI: 10.25045/SPCDH4.0.2025.02

employed a **few-shot learning** approach, which enables the model to perform tasks effectively after being trained on only a small number of examples.

II.MATERIALS AND METHODS

This study presents two specialized general-purpose large language models (LLMs): **NeuroRadReport** and **MammaBoardGPT**. These models are built on the GPT-4 framework and are capable of integrating both visual inputs and textual data. The **NeuroRadReport** model is designed for analyzing and reporting brain MRI scans for the purpose of ischemic stroke diagnosis, while the **MammaBoardGPT** model employs Retrieval-Augmented Generation (RAG) technology to support decision-making in the diagnosis and treatment of breast cancer.

NeuroRadReport was trained to identify signal characteristics within MRI sequences relevant to estimating stroke onset time. These sequences include **diffusion-weighted imaging (DWI)**, **apparent diffusion coefficient (ADC) maps**, and **fluid-attenuated inversion recovery (FLAIR)** images. The diagnostic capabilities of the NeuroRadReport tool were evaluated in a retrospective observational study using two validation datasets:

1.**Validation Dataset 1** – Consisted of 20 fully anonymized retrospective cases from GFO Kliniken Troisdorf, Germany, involving patients admitted with suspected stroke between January 1, 2023, and October 20, 2024.

2.**Validation Dataset 2** – Included 100 fully anonymized retrospective stroke-confirmed cases from the University Hospital of the Technical University of Munich. These cases are part of the **ISLES 2022 challenge** core dataset and have been previously published [8,9]. The ISLES challenge aimed to develop and benchmark algorithmic methods capable of automatic, accurate, and robust segmentation of stroke lesions.

This paper focuses specifically on analyzing stroke onset estimation using the GPT-4-based model.

Patients included in the analysis represented both sexes, various age groups, and diverse ethnic backgrounds. All cases were confirmed as ischemic stroke by a certified neuroradiologist.

Exclusion criteria were as follows:

1. Incomplete MRI datasets (e.g., missing DWI, FLAIR, or ADC sequences);
2. Low-quality MR images (e.g., motion artifacts in lesion slices);
3. Comorbid conditions that could interfere with stroke onset estimation (e.g., brain tumors, infections, hemorrhages).

The study was ethically approved by the Ärztekammer Nordrhein Ethics Committee (Approval No. 235-2024). Due to its retrospective design, the requirement for informed consent was waived. No personally identifiable information was introduced into the AI model.

The **NeuroRadReport** tool was queried with the following prompt: *“Identify the sequences used in the provided brain MRI scans and describe the findings in these images. Based on your analysis, where is the lesion located in the brain, and if stroke is suspected, what is the estimated time of onset (hyperacute, acute, subacute, or chronic)?”*

NeuroRadReport automatically recognized the MRI sequences (e.g., T1, T2, FLAIR, DWI) and described the lesion findings. Based on the imaging data, it identified the anatomical location of the lesion within the brain and, if stroke was suspected, estimated the stroke onset phase (hyperacute, acute, subacute, or chronic).

Before evaluating the decision-support potential of **MammaBoardGPT**, the model was trained using **Retrieval-Augmented Generation (RAG)** technology. For this purpose, the model was provided with PDF documents of the following official treatment protocols:

- **ESMO 2024** Clinical Practice Guidelines for Early-Stage Breast Cancer[9]
- **ESMO 2021** Guidelines for the Management of Metastatic Breast Cancer[10]
- **NICE 2017**: Guidelines for the Diagnosis and Treatment of Advanced Breast Cancer
- **Germany 2017**: Evidence-Based National Guidelines for Early Detection, Diagnosis, Treatment, and Follow-Up of Breast Cancer
- **St. Gallen 2023**: Key Recommendations on Radiotherapy from the 18th International Breast Cancer Conference
- Five additional peer-reviewed scientific articles [11]

For initial training, five clinical cases from **GFO Kliniken** were provided, each containing a fully documented tumor board decision. These cases were used to facilitate **few-shot learning**, allowing the model to learn how to generate treatment recommendations based on structured expert decisions.

Subsequently, the model was tested using an additional **25 clinical cases**, for which it generated its own treatment suggestions based on the provided guidelines. Importantly, the tumor board decisions for these 25 test cases were **not** disclosed to the model during inference.

The **GPT model** was configured to deliver treatment recommendations based solely on the clinical and diagnostic information presented for each patient. At no point was the model given access to any personally identifiable information. All inputs were fully anonymized, ensuring data privacy and compliance with ethical standards.The overall workflow of the study is presented in Fig.1.

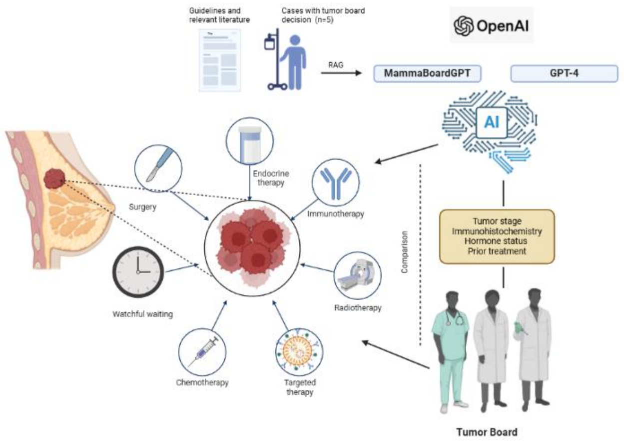

Figure 1. Both the fine-tuned MammaBoardGPT and the baseline (non-fine-tuned) GPT-4 model, as well as the GFO Kliniken Troisdorf tumor board, were evaluated in their ability to recommend the most optimal treatment plan for a cohort of consecutive breast cancer patients.

III.CONCORDANCE BETWEEN MAMMABOARDGPT AND TUMOR BOARD DECISIONS

The level of agreement between MammaBoardGPT, baseline GPT-4, and the tumor board was assessed before and after the application of recursive prompting and is presented in **Figure 2**. The comparison of MammaBoardGPT’s recommendations with the tumor board’s decisions for 25 anonymized patients revealed the following results:

- Full Concordance: MammaBoardGPT’s treatment recommendations were completely aligned with those of the tumor board. This occurred in 60% of cases (n = 15/25) before recursive prompting and 84% (n = 21/25) after recursive prompting.
- Partial Concordance: MammaBoardGPT’s recommendations were generally consistent with the tumor board’s decisions but differed in specific details such as drug selection, dosage, or adjunct therapies. This was observed in 32% of cases (n = 8/25) before prompting and 16% (n = 4/25) after prompting.
- Complete Discordance: In these cases, MammaBoardGPT suggested treatment plans that significantly diverged from those of the tumor board, often due to differences in guideline interpretation or limitations in the model's clinical contextualization. This occurred in 8% of cases (n = 2/25) prior to recursive prompting and 0% (n = 0/25) afterwards.

Although the application of recursive prompting improved alignment, no statistically significant difference was observed in concordance levels before and after prompting (P = 0.11).

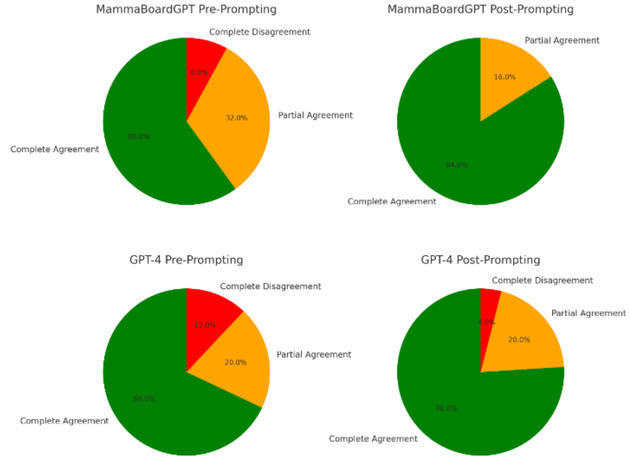

Figure 2. *Comparison of agreement levels between MammaBoardGPT, GPT-4, and the Tumor Board before and after the application of recursive prompting.*

Example Case 1

The first case is of an 87-year-old female diagnosed with invasive ductal carcinoma. She underwent a core biopsy of the right breast on 16 July 2024, followed by a core biopsy of the right axilla on 16 August 2024. On 20 August 2024, she underwent breast-conserving surgery without prior localization, under local anesthesia. Her pathology-report indicated she has pT2, pN+, M0 disease, with R0 resection status and no lymphovascular invasion (L0/V0). The tumor was high grade (G3), with estrogen receptor (ER) positivity at 20%, progesterone receptor (PR) negativity, HER2-negative status, and a Ki-67 proliferation index of 45%. The Tumor Board recommended endocrine therapy with an aromatase inhibitor (AI) and adjuvant radiotherapy. No chemotherapy was decided, as the patient has a significantly reduced general condition.

The ESMO guidelines highlight that AI’s provide additional benefits for patients with high risk breast cancer in comparison to tamoxifen, which applies to this patient (pN+, high Proliferation Index), as they further reduce distant recurrence risk compared to tamoxifen alone.

MammaBoardGPT suggested adjuvant radiotherapy and chemotherapy, as per ESMO guidelines chemotherapy is prioritized over endocrine therapy. Upon further recursive prompting and mentioning this patient has a significantly reduced general condition MammaBoardGPT suggested adjuvant radiotherapy and endocrine therapy with an aromatase inhibitor, without chemotherapy recommendation.

Example case 2

Another case in the research involved a 90-year-old female who presented with an ipsilateral loco-regional recurrence six years after breast-conserving surgery (R0). Her initial diagnosis was pT2 N0 M0 invasive lobular carcinoma, for which she received adjuvant radiotherapy and Tamoxifen treatment (currently ER-positive, PR-negative, and HER2-negative).

Guidelines from ESMO, NICE, and the 18th St. Gallen International Breast Cancer Consensus Conference do not recommend re-irradiation for loco-regional recurrence following initial breast-conserving surgery.

In this case, MammaBoardGPT recommended continuation of Tamoxifen treatment after radical mastectomy without radiotherapy, which is considered a reasonable alternative given the patient's advanced age and the risk of late radiation effects (e.g., damage to cardiac tissue).

The assessment of toxicity risk related to re-irradiation by an experienced radiotherapist, along with patient preferences, should guide the clinical decision-making process in such cases.

IV.CONCLUSION

It emphasizes that personalized GPT models like NeuroRadReport may contribute significantly to the diagnosis of ischemic stroke. NeuroRadReport has demonstrated strong performance in the identification of sequences and satisfactory capabilities in lesion localization, suggesting that, with further development, it could serve as a valuable support tool for radiologists—especially in settings where time and resources are limited. NeuroRadReport represents a promising step toward enhancing the radiological diagnosis of stroke and, when thoughtfully implemented in clinical practice, highlights the potential of artificial intelligence to improve patient outcomes.

Similarly, MammaBoardGPT, equipped with RAG capabilities and few-shot learning, as well as GPT-4, has shown encouraging potential as a decision-support tool during tumor board discussions for breast cancer treatment. The alignment of the model’s recommendations with current clinical guidelines and tumor board decisions indicates the possibility for AI tools to enhance clinical decision-making processes in the future. However, real-time decision-making capabilities and integration into clinical practice—particularly in complex cases—require further extensive investigation.

Both studies demonstrate that customized AI models—whether NeuroRadReport in stroke diagnosis or MammaBoardGPT in oncologic decision-making—may evolve into valuable support tools in clinical settings.

These tools can assist medical teams in the identification of diagnostic sequences, analysis of lesions, and formulation of treatment plans. Nonetheless, effective implementation requires integration into the clinical context and testing under real-world conditions. Proper integration of artificial intelligence has the potential to significantly improve the quality of healthcare services and patient outcomes in the future.

REFERENCES

[1] World Health Organization. (2024, February 1). **Global cancer burden growing amidst mounting need for services**. Retrieved from https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

[2] World Stroke Organization. (2022). *Global stroke fact sheet*. Retrieved from https://www.world-stroke.org/assets/downloads/WSO_Global_Stroke_Fact_Sheet.pdf

[3] Kwee TC, Kwee RM. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. Insights Imaging. 2021;12(1):88. doi:10.1186/s13244-021-01031-4. PMID: 34185175; PMCID: PMC8241957.

[4] Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. 2017;8(1):171-182. doi:10.1007/s13244-016-0534-1. PMID: 27928712; PMCID: PMC5265198.

[5] Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. AJR Am J Roentgenol. 2017;208(4):739-749. doi:10.2214/AJR.16.16963. PMID: 28026210.

[6] Gupta R, Krishnam S, Schaefer P, Lev M, Gonzalez R. An East Coast perspective on artificial intelligence and machine learning. Neuroimaging Clin N Am. 2020;30(4):467-478. doi:10.1016/j.nic.2020.08.002.

[7] Zhu G, Jiang B, Chen H, et al. Artificial intelligence and stroke imaging: A West Coast perspective. Neuroimaging Clin N Am. 2020;30(4):479-492. doi:10.1016/j.nic.2020.07.001. PMID: 33038998.

[8] Hernandez Petzsche MR, de la Rosa E, Hanning U, et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Sci Data. 2022;9:762. https://doi.org/10.1038/s41597-022-01875-5.

[9] Ge Loibl, S., André, F., Bachelot, T., et al. (2023). Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology, 34*(4). https://doi.org/10.1016/j.annonc.2023.04.011

[10] Gennari, A., André, F., Barrios, C. H., et al. (2021). ESMO Clinical Practice Guideline for the diagnosis, staging, and treatment of patients with metastatic breast cancer. *Annals of Oncology, 32*(12). https://doi.org/10.1016/j.annonc.2021.10.013

[11] German Guideline Program in Oncology. (2021). Evidence-based guideline for the early detection, diagnosis, treatment, and follow-up of breast cancer (Version 4.4). German Guideline Program in Oncology. AWMF-Registernummer.