

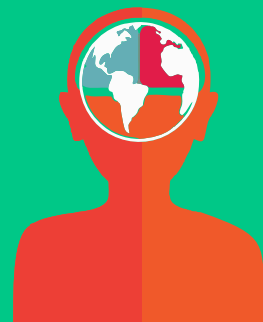
# ONCODAILY MEDICAL JOURNAL

*abstract*

## **Leveraging Large Language Models for Structured Data Extraction from Pediatric Brain Tumor Baseline MRI Reports: A Feasibility Study**

**Shahroze Ahmed, Kashif Siddique**

DOI: 10.69690/ODMJ-018-0915-5538



6th Pakistan Pediatric Neuro-Oncology Symposium, Pakistan, 2025



# Leveraging Large Language Models for Structured Data Extraction from Pediatric Brain Tumor Baseline MRI Reports: A Feasibility Study

**Authors:** Shahroze Ahmed, Kashif Siddique

**Affiliation:** Department of Radiology, Shaukat Khanum Memorial Cancer Hospital and Research Center, Lahore, Pakistan

**DOI:** [10.69690/ODMJ-018-0915-5538](https://doi.org/10.69690/ODMJ-018-0915-5538)

**Introduction:** Pediatric neuroradiology reports contain rich diagnostic details but are often unstructured. Manual data extraction for research and multidisciplinary team (MDT) discussions is time-consuming and error-prone. Large language models (LLMs) may automate this process by structuring reports into standardized, research-ready datasets.

**Methodology:** A dataset of 100 baseline pediatric brain MRI reports was analyzed. Two expert radiologists provided gold-standard annotations. Three LLMs (ChatGPT-5, Claude 3.5 Sonnet, Gemini 1.5) were tested for extraction of seven key clinical features: tumor location, tumor size, enhancement pattern, hemorrhage, leptomeningeal spread, hydrocephalus, and midline shift/herniation. Model performance was evaluated using precision, recall, and F1 score.

**Results:** All LLMs achieved >90% accuracy across features.

ChatGPT performed best (F1: 95.3%), followed by Gemini (94.4%) and Claude (94.2%). Feature-wise analysis showed highest accuracy for leptomeningeal spread (>97% F1) and relatively lower performance for midline shift/herniation and enhancement pattern (<94% F1), likely due to variable reporting styles. Performance variance across models was minimal (1.1%).

**Conclusion:** LLMs demonstrate feasibility for structured data extraction from pediatric MRI brain reports with high accuracy, comparable to expert annotations. Their use can reduce manual workload, standardize reporting language, improve MDT efficiency, and support research registries. Future work includes fine-tuning models, expanding datasets, and integration into clinical workflows.

**Conflict of Interest:** None

**Funding:** None

**Disclosure statement:** None

**License:** This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

© Kashif Siddique, 2025. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.