Smart Reporting: ChatGPT vs. Radiologists in MSK Knee MRI Analysis

Dr. Vikash Rustagi and Dr. Sakina Naqvi
Apollo Radiology International

RCR GLOBAL AI CONFERENCE 2025

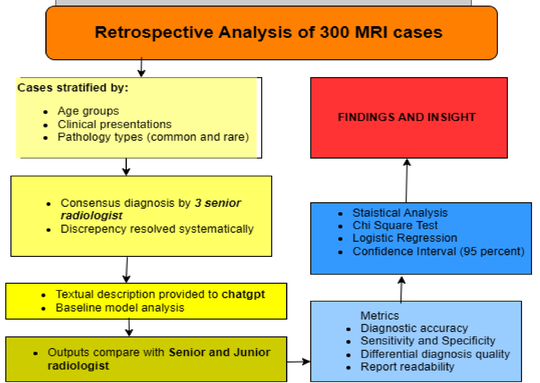
INTRODUCTION

Artificial Intelligence (AI) is reshaping radiology by enhancing efficiency and providing novel tools for clinical decision-making. Within this transformative landscape, natural language processing (NLP) models like ChatGPT have gained attention for their ability to generate detailed, context-specific textual outputs. Unlike traditional AI models in radiology, which primarily focus on image analysis, ChatGPT brings a unique capability: interpreting textual descriptions of imaging findings to provide diagnostic insights and differential diagnoses.

Musculoskeletal (MSK) knee MRI is among the most frequently performed imaging studies due to its diagnostic complexity and high prevalence in clinical practice. Its diverse pathologies—ranging from common injuries like meniscal tears and ACL ruptures to rarer conditions like Hoffa's fat pad impingement—make it an ideal model to evaluate the utility of AI in textual analysis.

This study investigates ChatGPT's ability to analyze textual descriptions of knee MRI findings, comparing its diagnostic accuracy, quality of differential diagnoses, and report readability to that of senior radiologists (>10 years of experience) and junior residents (<3 years of experience). By focusing exclusively on textual inputs, we aim to evaluate ChatGPT's baseline performance, identify its strengths and limitations, and explore its potential as a supportive tool in radiology education and clinical reporting workflows.

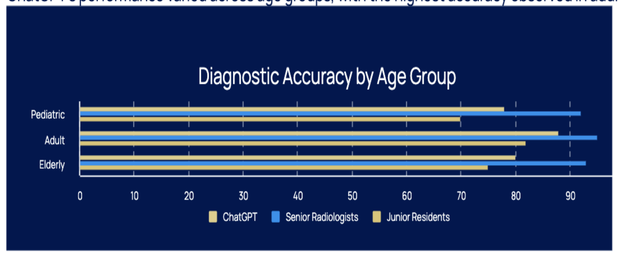
Methodology



Results

Overall Performance

ChatGPT's performance varied across age groups, with the highest accuracy observed in adults.

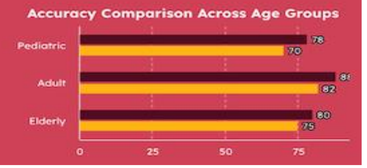


Category	ChatGPT	Senior Radiologists	Junior Residents
Overall Diagnostic Accuracy	84%	93%	76%
Accuracy for Common Pathologies	88%	95%	80%
Accuracy for Rare Pathologies	72%	89%	70%
Differential Diagnoses	3.2	4.1	2.8
Report Readability (Score)	8/10	10/10	7/10

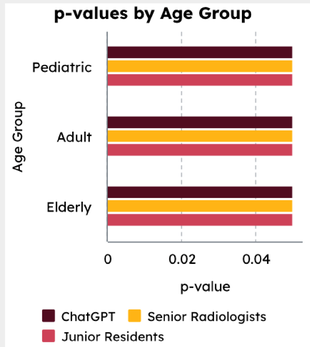
Subgroup Trends

ChatGPT vs. Experts

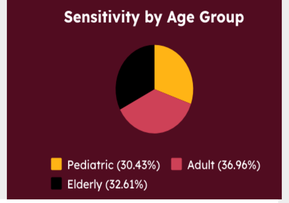
ChatGPT consistently outperformed junior residents, particularly for adult cases.



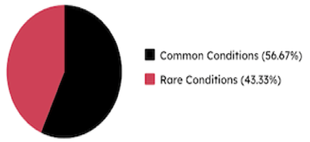
Significant differences in diagnostic accuracy were observed across age groups for all three groups



ChatGPT's sensitivity and specificity were higher for adult cases compared to pediatric and elderly groups.

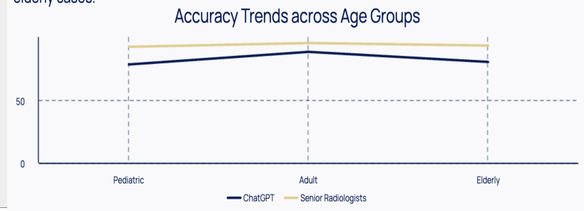


ChatGPT struggled with rare or age-specific conditions, like growth plate injuries in children and degenerative changes in elderly patients.



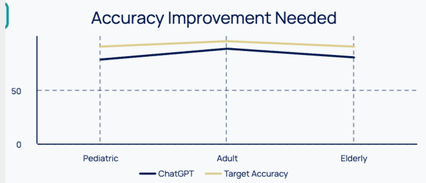
Age-Specific Challenges

ChatGPT showed lower accuracy with rare or age-specific conditions, particularly in pediatric and elderly cases.



Room for Improvement

ChatGPT's performance in pediatric and elderly cases suggests a need for further refinement with specialized datasets.



Implications

- ChatGPT demonstrates strong adaptability in adult cases but requires improvements for pediatric and elderly pathologies.
- Fine-tuning with radiology-specific datasets, especially incorporating pediatric and geriatric pathologies, could enhance performance across all age groups.
- The higher performance in adult cases suggests ChatGPT is better suited to handle common, well-documented conditions prevalent in this demographic.

Conclusion

ChatGPT represents a promising step forward in the application of AI in radiology, particularly as a supportive tool for education, preliminary reporting, and second-opinion diagnostics. However, its limitations in handling rare pathologies and reliance on textual inputs underscore the need for continued refinement. The future of ChatGPT in radiology will depend on its integration into workflows, strategic fine-tuning, and adherence to ethical and regulatory standards. With these improvements, ChatGPT has the potential to become a transformative tool, enhancing radiology practice and training while addressing the increasing demand for efficiency and accuracy in medical diagnostics.