PS3-04-14: Artificial Intelligence (AI) as a Decision Support Tool for Practicing Oncologists: Breast Cancer Cases

M. Jahanzeb, K. Haines, M. Kiyani, T. Buchholz, R. Butler, W. J. Gradishar, S. A. Hurvitz, T. A. King, R. L. Mahtani, T. Mamounas, H. McArthur, M. Morrow, A. P. O'Dea, J. O'Shaughnessy, M. D. Pegram, H. S. Rugo, C. Shah, F. Vicini, and N. Wolmark

SAN ANTONIO BREAST CANCER SYMPOSIUM®
UT Health San Antonio Mays Cancer Center
AACR American Association for Cancer Research

Abstract

Careful human oversight remains essential to ensuring safe and personalized cancer care. We have used leading artificial intelligence (AI) models to show a high degree of concordance with recommendations from a multidisciplinary review (MDR) consisting of specialists with expertise in breast cancer.

Introduction

Second opinions and MDRs can inform and impact the management of cancer patients, yet these resources are not universally accessible. We previously presented over 400 anonymized complex cancer scenarios across 5 major tumor types via video conference to MDR panel experts and tracked their treatment recommendations. These experts were specialists in radiology, medical, surgical, and radiation oncology. In this study, we focus on breast cancer cases to validate how recommendations from 3 leading AI models compare to those of MDR panel experts.

Methods and Materials

We selected 50 complex breast cancer scenarios previously adjudicated by MDR panels from a larger database of over 400 anonymized cancer cases collected between 2020 and 2021. These were analyzed by 3 different foundational AI models (OpenAI's ChatGPT 4.5, Anthropic's Claude Opus 4, and Google's Gemini Ultra) using PrecisCa's proprietary method for prompts. The AI recommendations from each system were scored on a scale of 1-5 (5 being the highest) for completeness, reasoning, clarity, menu of options, recency, and relevance, compared to the MDR panel recommendations. The maximum possible competence score was 30 points per scenario and an aggregate score of 1500. Final AI recommendations were also compared with current National Comprehensive Cancer Network (NCCN) guidelines for serious omissions. Comparison in reverse (additional AI options that the experts missed) was not performed due to interim changes in treatment recommendations over the past 4 years.

Results

Patient characteristics and aggregate concordance and competence scores are listed in **Table 1** below.

While there was some variation in the concordance and competence rates of the 3 AI models, there was excellent concordance with the expert opinion for all of them. Discordant cases were reviewed and primarily involved minor differences that would not have altered patient management significantly. Within the 6 categories listed above, AI systems broadly excelled in relevance and less so in completeness.

Conclusions

We demonstrate a high degree of concordance between 3 leading AI models and expert MDR panels for breast cancer clinical scenarios. These findings suggest that AI tools have matured sufficiently to serve as valuable decision support aids in oncology practice, particularly where expert review may be limited or unavailable.

Number of Breast Cancer Cases	50
Median Age (Range)	55 (27-83)
Stage	
I	7 (14%)
II	11 (22%)
III	8 (16%)
IV	24 (48%)
Histology	
Ductal Carcinoma	46 (92%)
Lobular Carcinoma	4 (8%)
Biomarker Status	
HR+/HER2-	25 (50%)
HR-/HER2-	11 (22%)
HR+/HER2+	11 (22%)
HR-/HER2+	3 (6%)
Aggregate/Median Competence Score (Range)	
ChatGPT 4.5	1311 / 26.5 (23-30)
Claude Opus 4	1408 / 27 (24-30)
Gemini Ultra	1423 / 28 (26-30)

Table 1: Patient Characteristics and Aggregate Competency Score

Contact
Mohammad Jahanzeb
Medical Oncology, Florida Atlantic University Schmidt College of Medicine, Boca Raton, FL
Email: mjahanzeb@precisca.com
Phone: (901) 483-0486

This presentation is the intellectual property of the author/presenter. Contact them at mjahanzeb@precisca.com for permission to reprint and/or distribute.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com