ORIGINAL ARTICLE / ORİJİNAL ARAŞTIRMA

DOI: 10.4274/turkjsurgoncol.galenos.2026.69885
Turk J Surg Oncol 2026;2(3):106-110

Concordance Between AI-generated Recommendations and Multidisciplinary Tumor Board Decisions in Postoperative Breast Cancer Management: A Retrospective Validation Study

Postoperatif Meme Kanseri Yönetiminde Yapay Zeka Tarafından Oluşturulan Öneriler ile Multidisipliner Tümör Konseyi Kararları Arasındaki Uyumluluk: Retrospektif Bir Validasyon Çalışması

Arda Ulaş Mutlu, Emir Çapkınoğlu, Halil Kara, Onur Dülgeroğlu, Nurullah Bakır, Cihan Uras

Acıbadem Mehmet Ali Aydınlar University Faculty of Medicine, Department of General Surgery, İstanbul, Türkiye

Abstract

Objectives: This study aimed to evaluate the concordance between artificial intelligence (AI)-generated treatment recommendations and real-world multidisciplinary tumor board (MDT) decisions in postoperative breast cancer management.

Material and Methods: This retrospective validation study included 30 female patients with breast cancer who were previously discussed in a MDT. Demographic, clinical, pathological, and molecular data were retrospectively collected from institutional records and anonymized before analysis. Structured clinical scenarios were created for each patient and entered into an AI-based decision-support system, developed using ChatGPT version 5.2, in accordance with current ESMO, NCCN, and ASCO guidelines. AI-generated recommendations for systemic therapy, radiotherapy, and genomic testing were compared with corresponding MDT decisions. Concordance was assessed for each treatment domain.

Results: The mean age was 52±13.1 years, and the median interval from pathology reporting to MDT discussion was 7 days (interquartile range, 4-11). Neoadjuvant therapy was administered to 11 patients (36.7%), and 4 (36.6%) achieved a pathological complete response. Breast-conserving surgery was performed in 22 patients (73.3%). Concordance between AI and MDT varied by treatment modality. Agreement was highest for radiotherapy (90.0%), followed by anti-human epidermal growth factor receptor 2 (HER2) therapy (73.3%) and endocrine therapy (70.0%). Lower concordance rates were observed for genomic testing (53.3%) and chemotherapy (33.3%). Chemotherapy showed the highest discordance rate (66.7%).

Conclusion: AI demonstrated high concordance with MDT decisions in treatment domains guided by standardized criteria, particularly radiotherapy, endocrine therapy, and anti-HER2 therapy. Lower agreement was observed for chemotherapy and genomic testing, suggesting that these decisions require more individualized clinical judgment. AI may serve as a useful decision-support tool within multidisciplinary breast cancer care.

Keywords: Artificial intelligence, breast cancer, multidisciplinary tumor board

Address for Correspondence: Arda Ulaş Mutlu, MD, Acıbadem Mehmet Ali Aydınlar University Faculty of Medicine, Department of General Surgery, İstanbul, Türkiye
E-mail: ardamutlu.1999@gmail.com **ORCID ID:** orcid.org/0000-0001-7499-7155
Received: 21.04.2026 **Accepted:** 03.07.2026 **Publication Date:** 11.09.2026

Cite this article as: Mutlu AU, Çapkınoğlu E, Kara H, Dülgeroğlu O, Bakır N, Uras C. Concordance between AI-generated recommendations and multidisciplinary tumor board decisions in postoperative breast cancer management: a retrospective validation study. Turk J Surg Oncol. 2026;2(3):106-110

Öz

Giriş / Amaç: Bu çalışmanın amacı, postoperatif meme kanseri yönetiminde yapay zeka (YZ) tarafından oluşturulan tedavi önerileri ile gerçek yaşam multidisipliner tümör konseyi (MTK) kararları arasındaki uyumu değerlendirmektir.

Gereç ve Yöntem: Bu retrospektif validasyon çalışmasına, daha önce MTK'da değerlendirilmiş 30 kadın meme kanseri hastası dahil edildi. Demografik, klinik, patolojik ve moleküler veriler kurumsal kayıtlardan retrospektif olarak toplandı ve analiz öncesinde anonimleştirildi. Her hasta için yapılandırılmış klinik senaryolar oluşturuldu ve bu senaryolar, güncel ESMO, NCCN ve ASCO kılavuzlarına uygun şekilde yapılandırılmış, ChatGPT sürüm 5.2 tabanlı YZ destekli karar destek sistemine girildi. Sistemik tedavi, radyoterapi ve genomik testlere ilişkin YZ önerileri, ilgili MTK kararları ile karşılaştırıldı. Her tedavi alanı için uyum değerlendirildi.

Bulgular: Ortalama yaş 52±13,1 yıl olup, patoloji raporundan MTK değerlendirmesine kadar geçen medyan süre 7 gün (IQR, 4–11) idi. On bir hastaya (%36,7) neoadjuvan tedavi uygulanmıştı ve bunların 4'ünde (%36,6) patolojik tam yanıt elde edildi. Meme koruyucu cerrahi 22 hastada (%73,3) uygulanmıştı. YZ ile MTK arasındaki uyum tedavi alanına göre değişkenlik gösterdi. En yüksek uyum radyoterapi için saptandı (%90). Bunu anti-insan epidermal büyüme faktörü reseptörü (HER2) tedavi (%73,3) ve endokrin tedavi (%70) izledi. Daha düşük uyum genomik test (%53,3) ve kemoterapi (%33,3) alanlarında görüldü. En yüksek uyumsuzluk oranı kemoterapi kararlarında saptandı (%66,7).

Tartışma / Sonuç: YZ, özellikle radyoterapi, endokrin tedavi ve anti-HER2 tedavi gibi standart ölçütlerle yönlendirilen tedavi alanlarında MTK kararları ile yüksek düzeyde uyum göstermiştir. Kemoterapi ve genomik test alanlarında ise daha düşük uyum saptanmış olup, bu durum söz konusu kararların daha fazla bireyselleştirilmiş klinik değerlendirme gerektirdiğini düşündürmektedir. YZ, multidisipliner meme kanseri yönetiminde yararlı bir karar destek aracı olarak kullanılabilir.

Anahtar Kelimeler: Yapay zeka, meme kanseri, multidisipliner tümör konseyi

Introduction

Breast cancer is the most common cancer among women (1). Over the past decades, advances in screening, surgical techniques, and oncological therapies have significantly improved survival outcomes. However, these advances have also made treatment strategies more complex, requiring a more individualized and multidisciplinary approach to patient care. This multidisciplinary approach requires a team built on the collaboration of breast surgeons, medical oncologists, radiation oncologists, pathologists, and radiologists (2).

Decision-making in cancer care relies on both structured frameworks from international guidelines and individualized patient evaluation. Multidisciplinary tumor boards (MDTs) are central to combining these elements for clinicians and patients (3,4). While MDTs are necessary, their meetings also have practical limitations. Because coordinating multiple specialists means meetings are typically held at set intervals, this can limit flexibility and possibly cause delays. Furthermore, the increasing patient load and the complexity of treatment algorithms add to the burden on multidisciplinary teams (4).

The use of artificial intelligence (AI) and, more recently large language models (LLMs), has rapidly expanded in healthcare (5). These systems can interpret clinical data and guideline-based recommendations. They can also generate structured treatment suggestions when given appropriate inputs (6). In oncology, decision-making often relies on integrating multiple variables. AI-based tools can support clinicians with consistent, guideline-informed recommendations (7).

Given these developments, new AI and LLM technologies can be integrated into daily practice to support clinicians' decision-making. In this study, we aim to evaluate the concordance between an AI-based decision-support system and real-world MDTs in postoperative breast cancer treatment recommendations.

Materials and Methods

This retrospective validation study evaluated the concordance between an AI-based decision support system and real-world MDT decisions in postoperative breast cancer management.

The AI-based decision support system was developed using ChatGPT, version 5.2. It followed current international clinical guidelines, including those of ESMO, NCCN, and ASCO. The tool is guideline-informed, not an autonomous decision-making system. For each patient, it generated structured recommendations for systemic therapy, radiotherapy, and genomic testing. The system operated in a locked configuration, with no changes to the model version, architecture, or prompt strategy during the study.

Patients evaluated by a multidisciplinary breast cancer tumor board were identified retrospectively. Demographic, clinical, pathological, and molecular data were extracted from institutional records. All data were anonymized before analysis. For each patient, structured clinical scenarios were created. These included age, tumor stage, family history, BRCA mutation, histopathological characteristics, hormone receptor status, human epidermal growth factor receptor 2 (HER2) status, Ki-67 index, molecular subtype, breast-conserving surgery status, axillary management, and postoperative tumor-node-metastasis

staging. No personally identifiable information was provided to the AI system.

Anonymized patient scenarios were entered into the AI-based decision support system. The generated recommendations were recorded under predefined categories: systemic therapy strategy, radiotherapy indication, and genomic testing recommendation. Outputs were then compared with the corresponding real-world MDT decisions for the same patients. The level of agreement was classified as concordant or discordant for each parameter.

The AI system was not used in real-time clinical decision-making and did not influence patient management. All AI-generated outputs were evaluated solely for validation purposes in a shadow mode setting.

Statistical Analysis

All statistical analyses were performed using Jamovi software, version 2.6. Categorical variables were presented as frequencies and percentages. Continuous variables were summarized as mean ± standard deviation or median with interquartile range (IQR), depending on data distribution.

The primary endpoint of the study was the concordance between AI-generated recommendations and MDT decisions. Concordance rates were calculated as the proportion of cases in which AI recommendations and MDT decisions agreed.

All data were anonymized and handled in accordance with data protection regulations. The study followed ethical standards for retrospective observational research. The Acıbadem Mehmet Ali Aydınlar University Institutional Review Board approved the study (approval number: 2026/02, date: 22.02.2026).

Results

A total of 30 female patients with breast cancer were included in the study. The mean age of the cohort was 52±13.1 years, and the median interval from pathology reporting to MDT discussion was 7 days (IQR, 4-11). Nineteen patients (63.3%) were postmenopausal, while 11 (36.7%) were premenopausal (Table 1).

Most tumors were invasive carcinomas (n=27, 90.0%). Three patients (10.0%) had *in situ* disease. Among invasive tumors, cT2 was the largest subgroup (59.3%). Seventeen patients (63%) had node-negative disease, and 10 (37%) patients had nodal involvement. All tumors were ductal carcinomas. Thirteen patients (44.8%) had grade 2 tumors; 16 (55.2%) had grade 3 tumors. Estrogen receptor positivity was observed in 20 patients (66.7%), and progesterone receptor positivity was observed in 14 patients (46.7%). HER2 status was negative in 16 patients (57.1%), positive in 6 (21.4%), and inconclusive in 6 (21.4%). The median Ki-67 index was 17%.

Table 1. Patient characteristics

	Variable	n (%)
Age		52±13.1
Pathology-MDT day		7 (4-11)
Menopause	Postmenopausal	19 (63.3)
	Premenopausal	11 (36.7)
Tumor type	Invasive carcinoma	27 (90.0)
	In situ	3 (10.0)
T stage	T0	2 (7.4)
	T1	1 (3.7)
	T1a	2 (7.4)
	T1b	2 (7.4)
	T1c	4 (14.8)
	T2	16 (59.3)
N stage	N0	17 (63.0)
	N+	10 (37)
Histology	Ductal carcinoma	30 (100)
Tumor grade	Grade 2	13 (44.8)
	Grade 3	16 (55.2)
ER	Positive	20 (66.7)
	Negative	10 (33.3)
PR	Positive	14 (46.7)
	Negative	16 (53.3)
HER2	Negative	16 (57.1)
	Positive	6 (21.4)
	Inconclusive	6 (21.4)
Ki-67 (%)		17 (12-35)
ECOG	0	9 (39.1)
	1	12 (52.2)
	2	2 (8.7)
Fertility preservation	Yes	6 (20.0)
	No	24 (80.0)
Surgery	Breast-conserving surgery	22 (73.3)
	Mastectomy	8 (26.7)
Axillary management	Sentinel biopsy	28 (93.3)
	Axillary dissection	2 (6.7)
Neoadjuvant therapy	Yes	11 (36.7)
	No	19 (63.3)
pCR	Yes	4 (13.3)
	No	26 (86.7)

MDT: Multidisciplinary tumor board, ER: Estrogen receptor, PR: Progesterone receptor, HER2: Human epidermal growth factor receptor 2, ECOG: Eastern Cooperative Oncology Group, pCR: Pathological complete response

Neoadjuvant therapy had been administered to 11 patients (36.7%), whereas the remaining 19 patients (63.3%) underwent upfront surgery. Among those who received neoadjuvant treatment, 4 patients (36.6%) achieved a pathological complete

response. Breast-conserving surgery was performed on 22 patients (73.3%), while 8 patients (26.7%) underwent mastectomy. Axillary staging was performed by sentinel lymph node biopsy in 28 patients (93.3%), whereas axillary dissection was required in 2 patients (6.7%).

Concordance between AI-generated recommendations and MDT decisions varied according to treatment modality. The highest concordance was observed for radiotherapy, with 27 cases (90%) in agreement. Concordance was also relatively high for anti-HER2 therapy and endocrine therapy, with agreement rates of 73.3% and 70.0%, respectively. In contrast, concordance was lower for genomic testing and chemotherapy (53.3% and 33.3%, respectively) (Table 2).

Chemotherapy had the highest discordance rate: 20 cases (66.7%) showed disagreement. Among these discordant cases, the AI recommended chemotherapy in 6 cases (30%) while the MDT did not; conversely, the MDT recommended chemotherapy in 4 cases (20.0%) while the AI suggested genomic testing. In 5 additional cases (25%), the AI recommended genomic testing when the MDT did not, and in 5 cases (25%), the MDT recommended genomic testing when the AI did not.

Table 2. Concordance

Chemotherapy	Concordance	10 (33.3)
	Discordance	20 (66.7)
	MDT genomic, AI chemotherapy	5 (25)
	MDT chemotherapy, AI genomic	4 (20)
	MDT no chemotherapy, AI genomic	5 (25)
	MDT no chemotherapy, AI chemotherapy	6 (30)
Endocrine therapy	Concordance	21 (70)
	Discordance	9 (30)
	MDT recommended endocrine, AI did not	5 (55.6)
	AI recommended endocrine, MDT did not	4 (44.4)
Anti-HER2 therapy	Concordance	22 (73.3)
	Discordance	8 (26.7)
	MDT recommended anti-HER2, AI did not	3 (37.5)
	AI recommended anti-HER2, MDT did not	5 (62.5)
Radiotherapy	Concordance	27 (90)
	Discordance	3 (10)
	Clinical risk-based MDT decision	3 (100)
Genomic testing	Concordance	16 (53.3)
	Discordance	14 (46.7)
	MDT recommended genomic test, AI did not	5 (35.7)
	AI recommended genomic test, MDT did not	9 (64.3)

AI: Artificial intelligence, MDT: Multidisciplinary tumor board, HER2: Human epidermal growth factor receptor 2

For endocrine therapy, discordance was observed in 9 cases (30%). In 5 of these cases (55.6%), the MDT recommended endocrine therapy, whereas the AI did not; conversely, in 4 cases (44.4%), the AI recommended endocrine therapy, whereas the MDT did not. Discordance in anti-HER2 therapy was observed in 8 cases (26.7%), including 3 cases (37.5%) in which the MDT recommended anti-HER2 therapy but the AI did not, and 5 cases (62.5%) in which the AI recommended anti-HER2 therapy, but the MDT did not.

Radiotherapy had the lowest discordance rate, with 3 cases (10%) showing disagreement. Genomic testing showed discordance in 14 cases (46.7%). Overall, agreement was higher for radiotherapy, endocrine therapy, and anti-HER2 therapy, whereas it was lower for chemotherapy and genomic testing.

Discussion

In this study, we evaluated the concordance between AI-generated treatment recommendations and decisions made by an experienced MDT in patients with breast cancer. Our findings indicate that agreement between the AI system and the MDT varied across treatment modalities, highlighting several key factors.

Radiotherapy demonstrated the highest concordance rate (90%). This finding is not unexpected, as radiotherapy decisions in breast cancer are largely based on standardized treatment algorithms incorporating surgical procedure, tumor stage, and nodal status (8). In such relatively structured and guideline-driven settings, AI systems may be able to reproduce clinical decision pathways with a high degree of consistency.

Similarly, concordance was relatively high for endocrine therapy (70.0%) and anti-HER2 therapy (73.3%). These treatment decisions are primarily determined by receptor status and well-established therapeutic algorithms, which may explain the higher level of agreement between AI-generated recommendations and MDT decisions (9).

In contrast, chemotherapy demonstrated the lowest concordance rate (33.3%). Most discordant cases arose when the AI recommended chemotherapy, whereas the MDT favored genomic testing or clinical observation. This finding suggests that AI systems may rely more heavily on guideline-based risk stratification, while clinicians often incorporate additional contextual and patient-specific factors when determining the indication for chemotherapy (10). Variables such as patient

age, tumor grade, proliferative activity, comorbidities, and clinical judgment may affect treatment selection beyond what algorithm-based systems can fully capture.

A similar pattern was observed for genomic testing, with moderate concordance between AI and MDT (53.3%). The AI system recommended genomic testing more frequently than the MDT. This may reflect the tendency of AI models to favor molecular testing in borderline cases where guidelines support the use of genomic assays as adjunctive decision-making tools, whereas clinicians may choose to omit such testing when the result is unlikely to alter the treatment plan (10).

Taken together, these findings stress an important distinction between treatment decisions that are predominantly guideline-driven and those that require more individualized clinical interpretation. AI-based systems appear to perform well in structured decision-making contexts, although they show lower agreement in more detailed scenarios that depend on physician experience and broader clinical judgment.

Study Limitations

This study has several limitations. First, the sample size was relatively small, which may limit the generalizability of the findings. Second, the analysis was conducted within a single-institution MDT setting, and treatment decisions may vary across institutions according to local practice patterns, clinical expertise, and patient characteristics. In addition, the AI system was evaluated retrospectively in shadow mode and, therefore, was not assessed within a real-time clinical workflow.

Regardless of these limitations, the present study provides preliminary insight into the interaction between AI-based decision support tools and multidisciplinary clinical decision-making in breast cancer care. Rather than replacing clinicians, such systems may serve as complementary tools that support clinical reasoning and facilitate guideline-informed treatment planning.

Conclusion

AI demonstrated high concordance with MDT decisions across treatment modalities that were largely guided by standardized criteria, including radiotherapy, endocrine therapy, and anti-HER2 therapy. By contrast, lower agreement was observed in modalities requiring more individualized judgment, particularly in chemotherapy and genomic testing. These findings suggest that AI may be most useful as a decision-support tool integrated into multidisciplinary clinical practice rather than as a substitute for expert clinical evaluation.

Ethics

Ethics Committee Approval: The Acıbadem Mehmet Ali Aydınlar University Institutional Review Board approved the study (approval number: 2026/02, date: 22.01.2026).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Concept/Design: A.U.M., E.Ç., C.U., Data Collection or Processing: A.U.M., E.Ç., N.B., Analysis or Interpretation: A.U.M., N.B., Literature Review: A.U.M., E.Ç., H.K., O.D., N.B., C.U., Writing, Reviewing and Editing: A.U.M., E.Ç., H.K., O.D., N.B., C.U.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. Int J Cancer. 2021;149:778-89.
2. Pillay B, Wootten AC, Crowe H, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: a systematic review of the literature. Cancer Treat Rev. 2016;42:56-72.
3. Taylor C, Munro AJ, Glynne-Jones R, et al. Multidisciplinary team working in cancer: what is the evidence? BMJ. 2010;340:743-5.
4. Prades J, Remue E, van Hoof E, Borras JM. Is it worth reorganising cancer services on the basis of multidisciplinary teams (MDTs)? A systematic review of the objectives and organisation of MDTs and their impact on patient outcomes. Health Policy. 2015;119:464-74.
5. Chen SF, Alyakin A, Seas A, et al. LLM-assisted systematic review of large language models in clinical medicine. Nature Medicine. 2026;32:1152-9.
6. Pinninti R, Gullapalli R, Mallavarapu KM, et al. Comparing artificial intelligence and multidisciplinary tumor board decision making in real-world cancer care: a prospective blinded concordance study. ESMO Real World Data Digit Oncol. 2026;11:100688.
7. Mano MS, Çitaku FT, Barach P. Implementing multidisciplinary tumor boards in oncology: a narrative review. Future Oncol. 2022;18:375-84.
8. Smith BD, Bellon JR, Blitzblau R, et al. Radiation therapy for the whole breast: executive summary of an American Society for radiation oncology (ASTRO) evidence-based guideline. Pract Radiat Oncol. 2018;8:145-52.
9. Loibl S, André F, Bachelot T, et al. Early breast cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2024;35:159-82.
10. Khalid T, Cutress RI, Remer M, Copson ER. Clinical Impact of somatic genomic testing on breast cancer care. Clin Oncol. 2025;37:103665.