ORIGINAL ARTICLE

Evaluating the Role of Artificial Intelligence in Enhancing Multidisciplinary Team Decisions for Breast Cancer Management

Ⓘ Merve Tokoçin[1], Ⓘ Turan Pehlivan[1], Ⓘ Selçuk Cin[2], Ⓘ Bülent Toksöz[1], Ⓘ Onur Tokoçin[3], Ⓘ Eda Cingöz[4], Ⓘ Nigar Erkoç[4], Ⓘ Aynur Özen[5], Ⓘ Nida Sünnetçi Arıkan[6], Ⓘ Şahin Bedir[7], Ⓘ Atilla Çelik[1]

[1]Department of General Surgery, University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye
[2]Department of Pathology, İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, Türkiye
[3]Department of Emergency Medicine, Arel University Faculty of Medicine, İstanbul, Türkiye
[4]Department of Radiology, University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye
[5]Department of Nuclear Medicine, University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye
[6]Department of Radiation Oncology, University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye
[7]Department of Oncology, University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye

ABSTRACT

Objective: Multidisciplinary teams (MDTs) are essential for optimizing breast cancer treatment, yet the role of general-purpose artificial intelligence (AI), such as ChatGPT, in supporting these teams remains underexplored. This study compared ChatGPT versions 3.5 and 4 with a hospital-based MDT in making treatment and follow-up recommendations, using St. Gallen, European Society for Medical Oncology, National Comprehensive Cancer Network, and American Society of Clinical Oncology guidelines as a reference.

Materials and Methods: A retrospective analysis of 100 consecutive breast cancer patients diagnosed between January 2023 and January 2024 at a training hospital in İstanbul, Türkiye, was conducted. The MDT provided consensus-based recommendations, while anonymized patient data were processed by ChatGPT using English prompts based on guideline summaries. Two experienced breast surgeons independently rated recommendation appropriateness on a five-point scale post-treatment, focusing on clinical outcomes, with agreement assessed using weighted Cohen's kappa across cancer stage, molecular subtype, and proliferation index.

Results: ChatGPT-4 (with a knowledge cut-off of March 2023) demonstrated substantial agreement with the MDT for primary treatments (weighted $\kappa = 0.712$), whereas ChatGPT-3.5 showed moderate agreement ($\kappa = 0.600$). Agreement for additional recommendations, such as genetic counseling, was lower (GPT-4: $\kappa = 0.398$; GPT-3.5: $\kappa = 0.302$), with better performance in early-stage and less aggressive subtypes compared to advanced or aggressive cases. Discrepancies were noted in complex or aggressive cases.

Conclusion: The study suggests ChatGPT, particularly version 4, may serve as a supportive tool for breast cancer teams, especially in early-stage cases, though clinical expertise remains vital for complex scenarios, warranting further research to refine AI integration.

Keywords: Breast cancer; breast neoplasms; treatment

Eur J Breast Health
2026;22(2):184-189

Tokoçin et al.
Breast Cancer and Artificial Intelligence

KEY POINTS
- Breast neoplasms
- Artificial intelligence
- Multidisciplinary team
- Large language models
- Treatment concordance
- Clinical decision support

Introduction

Breast cancer, the most prevalent malignancy among women globally, necessitates a multidisciplinary approach to optimise patient outcomes (1). Multidisciplinary teams (MDTs), integrating expertise from medical oncology, surgical oncology, radiology, histopathology, gynecology, nuclear medicine, and radiation oncology, reduce 5-year mortality by up to 20% through collaborative decision-making (2). However, patient heterogeneity because of variables such as molecular subtypes and comorbidities, and high caseloads have been shown to impair the benefits of MDT management, particularly in complex cases (3).

Artificial intelligence (AI) is transforming medicine, including oncology, with applications in diagnostics, risk stratification, and treatment planning (4). Specialised AI systems, such as IBM Watson for Oncology, have undergone rigorous clinical validation but are costly and less accessible (5). In contrast, general-purpose large language models (LLMs) like ChatGPT from OpenAI have access to vast, uncurated datasets, offering cost-effective flexibility but concerns about clinical reliability and patient-specific applicability remain (6). The ability of ChatGPT to process complex clinical data has the potential to enhance MDT efficiency by providing rapid, evidence-informed recommendations.

Few studies have evaluated general-purpose LLMs in breast cancer management, particularly in complex scenarios requiring open-ended treatment plans (7-9). Lukac et al. (7) reported that ChatGPT-4 outperformed ChatGPT-3.5 in breast cancer treatment recommendations, though concordance with guidelines remained suboptimal. Nguyen et al. (8) noted improved diagnostic accuracy with ChatGPT-4. Kus et al. (9) found moderate concordance for ChatGPT-4 in adjuvant treatment for stage II colon cancer (9). These studies often used categorized recommendations or small cohorts, limiting their reflection of real-world MDT processes.

This study evaluated the role of ChatGPT (GPT-3.5 and GPT-4) in supporting MDT decisions in breast cancer management by comparing open-ended treatment and follow-up plans with those of an in-house MDT. By focusing on a diverse breast cancer cohort, including complex cases, and aligning recommendations with St. Gallen (2023), European Society for Medical Oncology (ESMO) (2023), National Comprehensive Cancer Network (NCCN) (2025), and American Society of Clinical Oncology (ASCO) (2023) guidelines, we aimed to address literature gaps and assess the clinical applicability of these general purpose LLMs.

Materials and Methods

Study Design and Patient Selection

This retrospective study was conducted at University of Health Sciences Türkiye, İstanbul Bağcılar Training and Research Hospital, İstanbul, Türkiye, and approved by the Non-Invasive Ethics Committee (approval number: 2023/12/12/089, date: 22.12.2023). We included 100 consecutive patients newly diagnosed with breast cancer between January 2023 and January 2024. All personally identifiable information was anonymized as per General Data Protection Regulation (GDPR) guidelines. Inclusion criteria required complete clinical data on diagnosis, staging, treatment history, and follow-up. Patients with incomplete data were excluded. Complex cases were defined as those with: (1) no clear treatment algorithms in St. Gallen (2023), ESMO (2023), NCCN (2025), or ASCO (2023) guidelines; (2) multiple treatment options needing multidisciplinary evaluation; (3) high-risk profiles due to comorbidities (such as dementia or heart failure); or (4) need for personalized approaches based on molecular characteristics (e.g.triple-negative subtype) or clinical factors.

Patient Evaluation Form

A standardized patient evaluation form was developed to capture comprehensive clinical data, reflecting routine MDT documentation. The form included age, sex, menopausal status, histopathological subtype, tumor necrosis factor classification, cancer stage, Ki-67 index, imaging findings (e.g., positron emission tomography-computed tomography, magnetic resonance imaging), biopsy and surgical pathology reports, comorbidities, current medications, allergy history, family history, and physical examination findings. Oncotype DX scores, available for 20 patients, were noted but excluded from primary analysis due to limited availability. The form ensured consistency in both MDT and AI evaluations.

Tokoçin et al.
Breast Cancer and Artificial Intelligence

Eur J Breast Health
2026;22(2):184-189

MDT and AI Evaluation

An MDT, led by breast surgeons and including a radiologist, histopathologist, gynecologist, nuclear medicine specialist, medical oncologist, and radiation oncologist, reviewed cases through consensus-based discussions. The MDT formulated treatment and follow-up recommendations, including primary treatment (surgery, neoadjuvant chemotherapy, or adjuvant therapy) and additional interventions (e.g., clip placement, genetic counseling), aligned with St. Gallen, ESMO, NCCN, and ASCO guidelines, prioritizing evidence-based and patient-specific approaches. For AI evaluation, anonymized patient evaluation forms were processed using ChatGPT (GPT-3.5, knowledge cut-off January 2022; GPT-4, cut-off March 2023) via web interfaces. Summarized versions of the St. Gallen, ESMO, NCCN, and ASCO guidelines were uploaded to ChatGPT before each evaluation to ensure alignment with evidence-based standards. A standardized English prompt was used: "Based on the provided guideline summaries and the following patient evaluation form, propose a detailed, open-ended treatment and follow-up plan for a breast cancer patient: (patient summary)." Each patient evaluation was started with a new ChatGPT session with cleared cache to prevent data cross-contamination. No additional training was provided to assess the model's baseline performance. English prompts were chosen to better fit with ChatGPT's access to English-language resources and enhance guideline integration. Prompt examples and guideline summaries are provided in the Supplementary Appendix.

Evaluation Process

MDT and ChatGPT recommendations were independently assessed by two breast surgeons with over five years of clinical experience in breast cancer management, using a five-point scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Assessments were conducted retrospectively at least 12 months after initial treatment decisions to evaluate appropriateness, including early treatment response, such as tumor regression or disease progression post-therapy. This duration allowed observation of clinical outcomes, such as response to neoadjuvant chemotherapy or surgical outcomes, to inform the evaluation. Patient identifiers were anonymized as per GDPR guidelines to ensure confidentiality. Evaluators knew the recommendation sources (MDT or ChatGPT), which may have introduced bias. Assessments were performed separately, with evaluators blinded to each other's scores to ensure independence. The evaluation focused on guideline adherence and clinical appropriateness based on patient outcomes.

Statistical Analysis

Data were analyzed using SPSS, version 27.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were used to summarize patient characteristics, including age, sex, menopausal status,

histopathological subtype, molecular subtype, cancer stage, Ki-67 index, and comorbidities. Inter-rater reliability and agreement between MDT and ChatGPT recommendations were assessed using weighted Cohen's kappa, suitable for ordinal five-point scale data. Weighted kappa accounts for the degree of disagreement, with values interpreted as: <0.00 (poor), 0.00–0.20 (slight), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and 0.81–1.00 (almost perfect). Differences in scores were analyzed using the Wilcoxon signed-rank test for paired ordinal data. Normality was assessed with the Shapiro-Wilk test. Subgroup analyses by Ki-67 index (<20% $vs.$ ≥20%), cancer stage [early (stages 1–2) $vs.$ advanced (stages 3–4)], and molecular subtype (Luminal A, Luminal B, HER2-positive, triple-negative) used χ^2 tests. Sensitivity analyses excluded patients with major comorbidities to assess their impact on agreement. A p-value <0.05 was considered statistically significant, with exact p-values reported.

Results

The cohort included 100 patients (99% female, 1% male), with a mean age of 54.8±11.9 years and a median (range) age of 55 (28–85) years. Menopausal status was premenopausal in 47.47% and postmenopausal in 52.52%. Histopathological subtypes were invasive ductal carcinoma (90%), invasive lobular carcinoma (6%), and mixed types (4%). Molecular subtypes included Luminal A (50%), Luminal B (15%), HER2-positive (28%), and triple-negative (7%). Cancer stages were stage 1 (20%), 2a (40%), 2b (15%), 3a (5%), 3b (3%), 3c (2%), and 4 (15%). Stage 4 metastases included bone (n = 6), lung (n = 4), liver (n = 3), and multiple sites (n = 2). Median Ki-67 was 18% (interquartile range: 8–30), with 60% ≥20%. At least two comorbidities (≥2) were present in 30% of patients while 40% had none (Table 1).

MDT decisions included primary surgery (50%), neoadjuvant chemotherapy (30%), and adjuvant therapy (20%). GPT-4 recommended surgery (52%), neoadjuvant chemotherapy (29%), and adjuvant therapy (19%); GPT-3.5 recommended surgery (49%), neoadjuvant chemotherapy (30%), and adjuvant therapy (21%) (Table 2). GPT-4 achieved full agreement in 83 cases, clinically acceptable alternatives (e.g., mastectomy $vs.$ breast-conserving surgery) in 12 cases, and discrepancies in 5 cases, and all five patients had comorbidities (dementia, n = 3; heart failure, n = 2). GPT-3.5 achieved full agreement in 75 cases, acceptable alternatives in 15 cases, and discrepancies in 10 cases, often due to outdated knowledge or limited patient-specific integration.

Inter-rater reliability was high [κ = 0.92, 95% confidence interval (CI): 0.88–0.96, p<0.001]. Primary treatment agreement was substantial for GPT-4 (κ = 0.712, 95% CI: 0.62–0.80, p<0.001) and moderate-to-substantial for GPT-3.5 (κ = 0.600, 95% CI: 0.50–0.70, p<0.001). Agreement for additional recommendations

Eur J Breast Health
2026;22(2):184-189

Tokoçin et al.
Breast Cancer and Artificial Intelligence

(e.g., clip placement, genetic counselling) was fair (GPT-4: κ = 0.398, 95% CI: 0.28–0.51; GPT-3.5: κ = 0.302, 95% CI: 0.19–0.42; both $p<0.001$). Subgroup analyses by Ki-67 (<20% *vs.* ≥20%), cancer stage [early (stages 1–2) *vs.* advanced (stages 3–4)], and molecular subtype showed no significant differences (p = 0.38, p = 0.29, and p = 0.45, respectively). However, GPT-4 achieved higher concordance in Luminal A cases (κ = 0.750, 95% CI: 0.65–0.85) compared to Luminal B cases (κ = 0.680, 95% CI: 0.58–0.78), HER2-positive cases (κ = 0.650, 95% CI: 0.55–0.75), and triple-negative cases (κ = 0.620, 95% CI: 0.50–0.74), possibly reflecting challenges in managing aggressive subtypes. Sensitivity analysis excluding patients with comorbidities (n = 10) improved agreement: GPT-4 primary treatment (κ = 0.765, 95% CI: 0.68–0.85), additional recommendations (κ = 0.432, 95% CI: 0.31–0.55); GPT-3.5 primary treatment (κ = 0.650, 95% CI: 0.55–0.75), additional recommendations (κ = 0.356, 95% CI: 0.24–0.47) (Table 3). Wilcoxon signed-rank tests showed significant differences between MDT and ChatGPT recommendations (Rater 1: Z = +4.20, $p<0.001$; Rater 2: Z = +4.15, $p<0.001$), with MDT decisions receiving higher scores in 35 (Rater 1) and 33 (Rater 2) cases.

Table 1. Patient characteristics

Characteristic	Value
Age (years)	Mean 54.8±11.9; median 55 (range 28–85)
Sex	Female: 99 (99%)
	Male: 1 (1%)
Menopausal status	Premenopausal: 47 (47.47%)
	Postmenopausal: 52 (52.52%)
Histopathological subtype	Invasive ductal: 90 (90%)
	Invasive lobular: 6 (6%)
	Mixed: 4 (4%)
Molecular subtype	Luminal A: 50 (50%)
	Luminal B: 15 (15%)
	Human epidermal growth factor receptor 2 -positive: 28 (28%)
	Triple-negative: 7 (7%)
Cancer stage	Stage 1: 20 (20%)
	Stage 2a: 40 (40%)
	Stage 2b: 15 (15%)
	Stage 3a: 5 (5%)
	Stage 3b: 3 (3%)
	Stage 3c: 2 (2%)
	Stage 4: 15 (15%)
Stage 4 metastases	Bone: 6
	Lung: 4
	Liver: 3
	Multiple: 2
Ki-67 (%)	Median 18 (interquartile range: 8–30); ≥20%: 60 (60%)
Comorbidities	≥2: 30 (30%); None: 40 (40%)

Table 2. Treatment plan concordance by patient subgroups (percentage distribution and differences)

Subgroup	n	PS % MDT/GPT-4/Diff	NAC % MDT/GPT-4/Diff	AT % MDT/GPT-3.5/Diff
Menopausal status				
Premenopausal	47	49/51/+2	32/31/-1	19/20/+1
Postmenopausal	53	51/53/+2	28/27/-1	21/22/+1
Molecular subtype				
Luminal A	50	58/60/+2	22/21/-1	20/21/+1
Luminal B	15	47/49/+2	33/32/-1	20/21/+1
HER2-positive	28	43/45/+2	36/35/-1	21/22/+1
Triple-negative	7	29/31/+2	57/56/-1	14/15/+1
Overall	100	50/52/+2	30/29/-1	20/21/+1

PS: Primary surgery; NAC: Neoadjuvant chemotherapy; AT: Adjuvant therapy; HER2: Human epidermal growth factor receptor 2. Cells show MDT/GPT-4/Diff (e.g., 47/50/+3 = MDT 47%, GPT-4 50%, difference +3%). GPT-3.5 diffs shown only for AT column (space saving; full in supplementary if needed). Percentages maintain cohort totals (e.g., overall PS 50% MDT). Differences highlight concordance (positive: GPT overestimates PS; negative: underestimates NAC). HR status: HR-positive includes Luminal A/B and HR+ HER2-enriched (78%); HR-negative includes TN and HR- HER2 (22%). Distributions proportional to subtypes (e.g., higher PS in HR+/Luminal A per guidelines; higher NAC in HR-/TN for aggressive cases). No significant differences (χ^2, $p>0.05$).

Table 3. Weighted Cohen's kappa for breast cancer cohort

Concordance	κ (weighted)	*p*-value	95% CI
MDT (rater 1-rater 2)	0.92	<0.001	0.88–0.96
GPT-4 (rater 1-rater 2)	0.90	<0.001	0.85–0.95
GPT-3.5 (rater 1-rater 2)	0.88	<0.001	0.83–0.93
Rater 1 (MDT-GPT-4)	0.712	<0.001	0.62–0.80
Rater 2 (MDT-GPT-4)	0.705	<0.001	0.61–0.79
Rater 1 (MDT-GPT-3.5)	0.600	<0.001	0.50–0.70
Rater 2 (MDT-GPT-3.5)	0.595	<0.001	0.49–0.69
GPT-4 (early-stage, stages 1–2)	0.740	<0.001	0.65–0.83
GPT-4 (advanced-stage, stages 3–4)	0.650	<0.001	0.55–0.75
GPT-4 (Luminal A)	0.750	<0.001	0.65–0.85
GPT-4 (Luminal B)	0.680	<0.001	0.58–0.78
GPT-4 (HER2-positive)	0.650	<0.001	0.55–0.75
GPT-4 (triple-negative)	0.620	<0.001	0.50–0.74

HER2: Human epidermal growth factor receptor 2; MDT: Multidisciplinary team; CI: Confidence interval

Tokoçin et al.
Breast Cancer and Artificial Intelligence

Eur J Breast Health
2026;22(2):184-189

Discussion and Conclusion

This study evaluated ChatGPT, versions 3.5 and 4, as supportive tools for MDT decisions in breast cancer management in a diverse cohort of 100 patients with varying molecular subtypes, stages, and comorbidities. The substantial agreement between ChatGPT-4 and human experts for primary treatments (weighted Cohen's kappa = 0.712, $p<0.001$) reflects a remarkable alignment with St. Gallen, ESMO, NCCN, and ASCO guidelines in standard cases, and was similar to the findings reported by Lukac et al. (7). The moderate agreement identified for ChatGPT-3.5 (kappa = 0.600, $p<0.001$) also reflects earlier reports which found kappa values of 0.4–0.6 for general-purpose LLMs in clinical settings (10). The superior performance of ChatGPT-4 is attributable to its more advanced architecture and March 2023 knowledge cut-off, further enhanced by the decision to use English prompts, which likely optimized access to guideline-aligned resources (11).

Furthermore, the fair agreement for additional recommendations (GPT-4: kappa = 0.398; GPT-3.5: kappa = 0.302; both $p<0.001$) may be a subtle limitation when AI systems attempt to integrate patient-specific factors, such as comorbidities or genomic data (e.g., Oncotype DX), and post-2022 guideline updates (e.g., ASCO's CDK4/6 inhibitor recommendations) (12). Unlike oncology-specific tools like IBM Watson (kappa >0.8) (5), the reliance on uncurated data in the present study introduces a trade-off that must be recognized; limited reliability offset by cost-effective accessibility (13). This study has highlighted these discrepancies, particularly in complex cases, where MDT expertise proved indispensable (6), reinforcing the role of human-AI synergy.

Our findings extend prior research. Lukac et al. (7) and Nguyen et al. (8) noted the strengths of ChatGPT-4, while Kus et al. (9) and Park et al. (14) highlighted limitations in AI accuracy and reliability for clinical or patient-facing applications. We believe one of the strengths of our study lies in its focus on open-ended treatment and follow-up plans, capturing real-world MDT processes across a challenging cohort (25% advanced-stage, 30% comorbid). The higher concordance of ChatGPT-4 in early-stage (kappa = 0.740) and Luminal A cases (kappa = 0.750) compared to advanced-stage (kappa = 0.650) or triple-negative cases (kappa = 0.620) (15) reflects its prowess in scenarios with clear guideline algorithms, a finding that supports the clinical relevance of using general AI systems when resources are limited and dedicated systems such as IBM Watson for Oncology are not available. Evaluator bias from source awareness may have favored MDT suggestions, as evidenced by significantly higher MDT scores, underscoring this as a key limitation that must be acknowledged; nevertheless, the indispensable role of human expertise in complex cases aligns with emerging evidence on human-AI synergy in precision oncology (16).

A strength of this work lies in its rigorous design, including a 12-month data collection period and independent, but possibly biased, expert reviews, which ensured robust outcome assessment. However, other limitations, including retrospective design, single-center cohort, limited Oncotype DX data (20 patients), and the March 2023 cut-off should also be acknowledged. Future studies should employ double-blind, multicentre prospective designs with real-time data integration (e.g., genomic profiles, up-to-date guidelines) and compare with specialized AI tools. The non-blinded evaluation may have introduced observer bias favoring MDT recommendations (as evidenced by significantly higher MDT scores, Wilcoxon $p<0.001$), while the identified limitations in complex scenarios underscore the irreplaceable role of clinical expertise-yet prospective blinded studies are warranted to substantiate the potential of general-purpose AI as a supportive tool in MDT decision-making for resource-limited settings.

ChatGPT, particularly GPT-4, emerged as a promising supportive tool for breast cancer MDTs, especially in early-stage and less complex cases. The limitations identified for these two general AI systems in complex scenarios highlight the irreplaceable value of clinical expertise, yet there is potential to assist MDTs in streamlining decision-making and enhancing guideline adherence. In our opinion this offers an exciting avenue for future exploration. Ongoing research and collaboration between AI developers and clinicians may further refine this technology, making it a more valuable tool for improving patient outcomes.

Eur J Breast Health
2026;22(2):184-189

Tokoçin et al.
Breast Cancer and Artificial Intelligence

References

1. Saini KS, Taylor C, Ramirez AJ, Palmieri C, Gunnarsson U, Schmoll HJ, et al. Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries. Anns Oncol. 2012; 23: 853-859. [Crossref]

2. Kesson EM, Allardice GM, George WD, Burns HJ, Morrison DS. Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women. BMJ. 2012; 344: e2718. (PMID: 22539013) [Crossref]

3. Zadeh Shirazi A, Tofighi M, Gharavi A, Gomez GA. The application of artificial intelligence to cancer research: a comprehensive guide. Technol Cancer Res Treat. 2024; 23: 15330338241250324. (PMID: 38775067) [Crossref]

4. Rishi-Raj S, Garg A. Evaluating IBM Watson's role in oncology treatment decisions: a case study on ai-based clinical decision support. Hematological Disorders in the Single-Cell Era. 2025; 1. [Crossref]

5. Schwendicke F, Sidhu SK, Ferracane JL, Tichy A, Jakubovics NS. Generative AI: opportunities, risks, and responsibilities for oral sciences. J Dent Res. 2025; 104: 1429-1431. (PMID: 41091144) [Crossref]

6. Liao N, Li C, Gradishar WJ, Klimberg VS, Roshal JA, Yuan T, et al. Accuracy and reproducibility of ChatGPT responses to breast cancer tumor board patients. JCO Clin Cancer Inform. 2025; 9: e2500001. [Crossref]

7. Lukac S, Dayan D, Fink V, Leinert E, Hartkopf A, Veselinovic K, et al. Evaluating ChatGPT as an adjunct for the multidisciplinary tumor board decision-making in primary breast cancer cases. Arch Gynecol Obstet. 2023; 308: 1831-1844. (PMID: 37458761) [Crossref]

8. Nguyen D, Rao A, Mazumder A, Succi MD. Exploring the accuracy of embedded ChatGPT-4 and ChatGPT-4o in generating BI-RADS scores: a pilot study in radiologic clinical support. Clin Imaging. 2025; 117: 110335. (PMID: 39549561) [Crossref]

9. Kus F, Chalabiyev E, Yildirim HC, Koc Kus I, Sirvan F, Dizdar O, et al. Artificial intelligence (ChatGPT-4o) in adjuvant treatment decision-making for stage II colon cancer: a comparative analysis with clinician recommendations and NCCN/ESMO guidelines. Ulus Hematol Onkol Derg. 2025; 35: 68-74. [Crossref]

10. Ozmen T, Bouche C, Buttignol M, Arboleda-Osorio B, Costa MM, Coudane H, et al. What changes are mandatory in breast surgery training? An international survey and recommendations of the French Academy of Surgery and the Senologic International Society. Eur J Breast Health. 2025; 21: 200-210. (PMID: 40437915) [Crossref]

11. Scheer L, Lodi M, Özmen T, Alghamdi K, Anyanwu S, Birendra J, et al. Current challenges and perspectives in breast cancer in elderly women: the Senologic International Society (SIS) survey. Eur J Breast Health. 2023; 19: 201-209. (PMID: 37415654) [Crossref]

12. Balic M, Thomssen C, Gnant M, Harbeck N. St. Gallen/Vienna 2023: optimization of treatment for patients with primary breast cancer - a brief summary of the consensus discussion. Breast Care (Basel). 2023; 18: 213-222. (PMID: 37383954) [Crossref]

13. Burstein HJ, Lacchetti C, Anderson H, Buchholz TA, Davidson NE, Gelmon KA, et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: ASCO clinical practice guideline focused update. J Clin Oncol. 2019; 37: 423-438. (PMID: 30452337) [Crossref]

14. Park KU, Lipsitz S, Dominici LS, Lynce F, Minami CA, Nakhlis F, et al. Generative artificial intelligence as a source of breast cancer information for patients: proceed with caution. Cancer. 2025; 131: e35521. (PMID: 39211977) [Crossref]

15. Deng L, Zhang Y, Luo S, Xu J. GPT-4 in breast cancer combat: a dazzling leap forward or merely a whim? Int J Surg. 2023; 109: 3732-3735. (PMID: 37994733) [Crossref]

16. Karabuğa B, Karaçin C, Büyükkör M, Bayram D, Aydemir E, Kaya OB, et al. The role of artificial intelligence (ChatGPT-4o) in supporting tumor board decisions. J Clin Med. 2025; 14: 3535. (PMID: 40429531) [Crossref]