Am J Cancer Res 2024;14(12):5885-5896
www.ajcr.us /ISSN:2156-6976/ajcr0158258

Original Article

Deep learning to optimize radiotherapy decisions for elderly patients with early-stage breast cancer: a novel approach for personalized treatment

Guangliang Yang[1*], Haiqi Chen[2*], Jinchao Yue[1]

*[1]Department of Oncology, Dongying District People's Hospital, 333 Jinan Road, Dongying District, Dongying, Shandong, China; [2]Department of General Surgery, Dongying District People's Hospital, 333 Jinan Road, Dongying District, Dongying, Shandong, China. *Equal contributors and co-first authors.*

Received June 4, 2024; Accepted November 27, 2024; Epub December 15, 2024; Published December 30, 2024

Abstract: The use of routine adjuvant radiotherapy (RT) after breast-conserving surgery (BCS) is controversial in elderly patients with early-stage breast cancer (EBC). This study aimed to evaluate the efficacy of adjuvant RT for elderly EBC patients using deep learning (DL) to personalize treatment plans. Five distinct DL models were developed to generate personalized treatment recommendations. Patients whose actual treatments aligned with the DL model suggestions were classified into the Consistent group, while those with divergent treatments were placed in the Inconsistent group. The efficacy of these models was assessed by comparing outcomes between the two groups. Multivariate logistic regression and Poisson regression analyses were used to visualize and quantify the influence of various features on adjuvant RT selection. In a cohort of 8,047 elderly EBC patients, treatment following the Deep Survival Regression with Mixture Effects (DSME) model's recommendations significantly improved survival, with inverse probability of treatment weighting (IPTW)-adjusted benefits, including a hazard ratio of 0.70 (95% CI, 0.58-0.86), a risk difference of 4.63% (95% CI, 1.59-7.66), and an extended mean survival time of 8.96 months (95% CI, 6.85-10.97), outperforming other models and the National Comprehensive Cancer Network (NCCN) guidelines. The DSME model identified elderly patients with larger tumors and more advanced disease stages as ideal candidates for adjuvant RT, though no benefit was seen in patients not recommended for it. This study introduces a novel DL-guided approach for selecting adjuvant RT in elderly EBC patients, enhancing treatment precision and potentially improving survival outcomes while minimizing unnecessary interventions.

Keywords: Early-stage breast cancer, radiotherapy, elderly patients, deep learning, causal inference

Introduction

Breast cancer is the most prevalent cancer among women, and its incidence increases with age [1]. Approximately 30% of invasive breast cancer diagnoses and half of all breast cancer-related deaths occur in women aged 70 years and older [2]. Despite advancements in medical care over the past two decades, mortality rates have significantly declined in younger patients, but survival rates for elderly patients have not improved [3]. This disparity is primarily due to older patients being more vulnerable to frailty and comorbidities, making it necessary to tailor their treatment differently from that of younger patients [4, 5].

Standard treatment for early-stage breast cancer (EBC) generally includes breast-conserving surgery (BCS) followed by adjuvant radiotherapy (RT) [6]. However, the suitability of this approach for elderly patients is still debated [7]. Notably, the PRIME II trial, reported by Kunkler et al. [8], showed no significant benefit from adjuvant RT in women over 65 with low-risk EBC (T1-2, node-negative, and estrogen receptor (ER)-positive), which is consistent with the National Comprehensive Cancer Network (NCCN) recommendations suggesting the potential omission of RT in similar cases [6]. Conversely, a study by Wang et al. [2], using data from the National Cancer Database, found that omitting RT may increase mortality in elderly patients,

https://doi.org/10.62347/TRNO3190

highlighting the complexities of determining effective treatment in this population. Precision medicine, which tailors healthcare based on individual patient characteristics, is becoming increasingly important in this context [9].

Traditional methods for evaluating treatment effects, such as randomized controlled trials (RCTs) and large-scale observational studies, often fail to account for individual differences, potentially leading to overgeneralizations [10]. To assess treatment heterogeneity, the conventional approach involves subdividing patients into representative subgroups and conducting RCTs within each group. However, this is costly, time-consuming, and ethically challenging [11]. Additionally, observational studies are particularly prone to biases, which complicate the inference of unbiased individual treatment effects (ITE) [10]. Prior research [12, 13] has demonstrated that deep learning (DL)-based treatment recommendation systems can effectively predict ITE, identify treatment heterogeneity, and select the most suitable treatments for individual patients. For this reason, we chose DL for the subsequent statistical analysis and treatment recommendation in this study.

By employing a DL model, our goal is to identify elderly EBC patients who would benefit from adjuvant RT, thereby optimizing treatment plans to enhance survival outcomes while minimizing unnecessary interventions.

Methods

Study design and setting

As a population-based retrospective cohort study, this study seeks to offer tailored treatment recommendations for elderly patients with EBC through the application of deep learning (DL). Participants were drawn from the Surveillance, Epidemiology, and End Results (SEER) 18 database, covering cancer patients across 18 U.S. regions, approximately representing 27.8% of the national population [14]. Adherence to the "Enhanced Guidelines for Reporting Observational Studies in Epidemiology" was maintained throughout this study [15].

This study focused on female patients aged 65 and older diagnosed with ductal, lobular, or mixed ductal-lobular carcinoma as their primary cancer between 2010 and 2015, who underwent BCS. Exclusion criteria included: (1) incomplete demographic data; (2) missing human epidermal growth factor receptor (HER), estrogen receptor (ER) or progesterone peceptor (PR) status; (3) carcinoma in situ; (4) unspecified laterality or presence of bilateral breast cancer; (5) undetermined TNM stage or tumor size; (6) metastatic breast cancer; (7) unspecified axillary lymph node status; (8) absence of data on adjuvant RT; (9) unknown histologic grades and types; and (10) incomplete follow-up or presence of multiple malignancies. The selection process is depicted in **Figure 1A**.

Data on demographic variables (sex, age, race, income, marital status), tumor attributes (location, size, laterality, histological grade, type, and TNM stage), and treatment specifics (administration of adjuvant RT) were extracted from the SEER database. Cases with any missing clinical characteristics were not included. Primary study endpoints were overall survival (OS) - the period from diagnosis to death from any cause - and breast cancer-specific survival (BCSS), the interval from diagnosis to breast cancer-related death. Subjects alive as of December 31, 2020, were censored. Tumor staging followed the guidelines of the 7th edition of the American Joint Committee on Cancer Staging Manual.

Algorithms

The T-learner employs dual models to estimate the ITE as $ITE = \mu_1(\chi) - \mu_0(\chi)$, where μ_1 and μ_0 are models trained on distinct treatment cohorts [16]. While the T-learner mitigates some confounding effects, its performance may still suffer due to unequal predictive accuracy [10] and skewed treatment assignments [17] resulting from disparities in patient numbers and baseline characteristics across treatment groups.

Balanced Individual Treatment Effect for Survival data (BITES) [17], a semi-parametric DL survival regression model, optimizes treatment arm comparisons by maximizing the p-Wasserstein distance using Integral Probability Metrics. This approach effectively addresses imbalances in both covariate spaces [18] and latent representations [19]. Unlike the T-learner, which uses separate estimators, BITES employs a unified model architecture

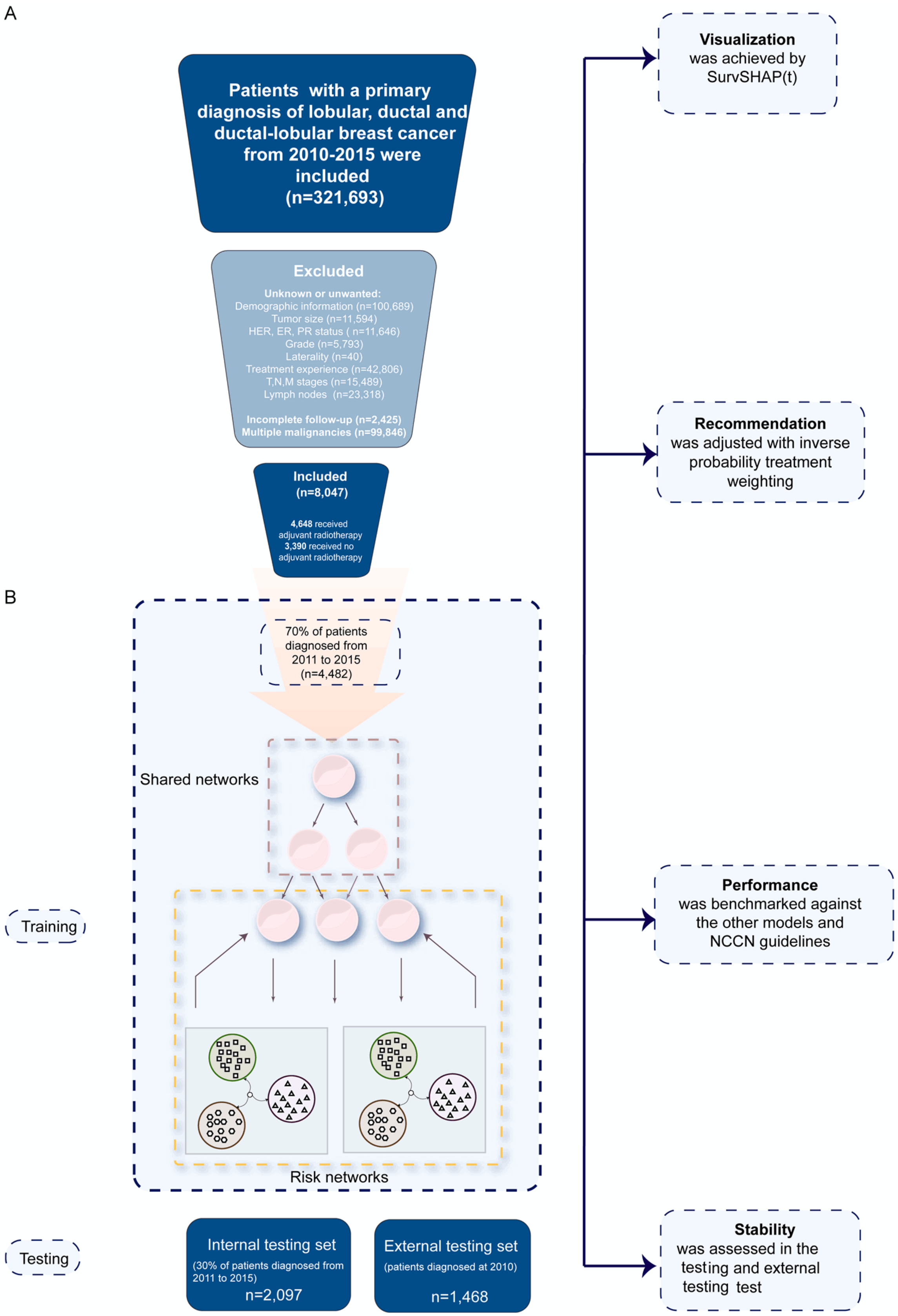

A
Patients with a primary diagnosis of lobular, ductal and ductal-lobular breast cancer from 2010-2015 were included (n=321,693)
Excluded
Unknown or unwanted:
Demographic information (n=100,689)
Tumor size (n=11,594)
HER, ER, PR status (n=11,646)
Grade (n=5,793)
Laterality (n=40)
Treatment experience (n=42,806)
T,N,M stages (n=15,489)
Lymph nodes (n=23,318)
Incomplete follow-up (n=2,425)
Multiple malignancies (n=99,846)
Included (n=8,047)
4,648 received adjuvant radiotherapy
3,390 received no adjuvant radiotherapy
B
70% of patients diagnosed from 2011 to 2015 (n=4,482)
Shared networks
Training
Risk networks
Testing
Internal testing set
(30% of patients diagnosed from 2011 to 2015)
n=2,097
External testing set
(patients diagnosed at 2010)
n=1,468
Visualization was achieved by SurvSHAP(t)
Recommendation was adjusted with inverse probability treatment weighting
Performance was benchmarked against the other models and NCCN guidelines
Stability was assessed in the testing and external testing test

Figure 1. Inclusion process and model architecture. A: Inclusion process. B: The model architecture of Deep Survival Regression with Mixture Effects.

with dual output heads, enhancing consistency across treatment group comparisons due to its end-to-end training.

The Deep Survival Regression with Mixture Effects (DSME) [20] synthesizes elements from the T-learner, representation-based, and subclassification causal inference strategies. Presented in **Figure 1B**, DSME comprises a shared network paired with dual risk networks. It generates balanced representations similar to BITES. Each risk network in DSME, representing a specific treatment group, consists of a finite mixture of K neural networks. Patient assignment to these networks is governed by a gating function *g(.)* [21], allowing for subgroup differentiation within the risk networks. DSME utilizes outputs from the shared network as a posterior and maximizes patient representation via the *Q(.)* function [21], enhancing risk stratification [10]. This model not only addresses heterogeneity within each subgroup but also refines the proportional hazard assumption to a conditional proportional hazard assumption, valid only within each latent group.

Calculation of individual treatment effect

In estimating the ITE, each patient provides only one observable outcome, with the counterfactual scenario remaining unobservable. Therefore, these outcomes must be inferred through predictive modeling. The individual survival distribution, derived from predicted log hazard ratios and treatment-specific baseline hazards, indicates changes in survival probability over time.

We conceptualized the potential outcome as the duration until a predetermined mortality threshold (50%), termed time at risk (TaR). The formulation is as follows: $ITE_{TaR}(X;P) = TaR_{do\ T=1;\ P=50\%} - TaR_{do\ T=0;\ P=50\%}$, where P represents the set mortality rate, X denotes the covariates, and $TaR_{T=1}$ and $TaR_{T=0}$ represent the predicted time intervals under two distinct treatment conditions, respectively. This approach facilitates the derivation of individualized treatment recommendations based on the ITE values.

Model development, validation, and treatment recommendation

We trained a total of five models: DSME, BITES, Cox Mixtures with Heterogeneous Effects (CMHE) [22], DeepSurv [23], and the Cox Proportional Hazards model (CPH). DeepSurv and CPH utilized the T-learner structure for training and application.

Initially, patients diagnosed in 2010 were designated as an external testing cohort and excluded from model exposure. From the remaining dataset, 70% of the patients were randomly selected for the training set used to develop the models, while the remaining 30% formed a testing set that remained unseen by the models to assess performance. During model development, fivefold cross-validation was employed to optimize hyperparameters, training models on four-fifths of the training data and validating on the remaining fifth. Training was automatically halted if there was no decrease in validation loss after 1,000 iterations.

To assess the impact of model recommendations, patients were categorized into recommended (Consis.) and anti-recommended (Inconsis.) groups, based on the alignment between the actual treatment received and the model's recommendations. We calculated the multivariate hazard ratio (HR), 10-year risk difference (RD), and the difference in 10-year restricted mean survival time (DRMST) between these groups to evaluate the protective effects of the models. Inverse probability treatment weighting (IPTW) was employed to adjust for baseline imbalances between the Consis. and Inconsis. groups. All models were standardized in their approach to calculating ITE.

Statistical analyses

Statistical analyses were conducted using R version 4.1.3 and Python version 3.8. For reporting purposes, continuous variables were described using the median and interquartile range (IQR), while categorical variables were presented as counts and percentages (%). The comparison of Kaplan-Meier (KM) survival curves was facilitated through the application of the log-rank test.

Results

Patients

In this study, 8,047 elderly female breast cancer patients with complete follow-up data who satisfied the inclusion criteria were analyzed. The overall mortality rate was recorded at 18.4% (95% CI: 17.6%-19.3%) across a median follow-up duration of 75 months (IQR: 57-96). Patients had a median age of 73 years (IQR: 68-79) and a median tumor size of 11 mm (IQR: 7-15). Of these, 4,648 (57.8%) received adjuvant radiation therapy, while the remaining 3,399 (42.2%) did not undergo radiation treatment. Detailed baseline clinical characteristics are summarized in Table S1.

Model performance

The study included 2,097 patients in the testing set and 1,468 diagnosed in 2010 for the external testing set. We assessed performance metrics over a 10-year horizon for both sets. To mitigate potential biases from better prognostic factors in the Consis. group, IPTW was applied to adjust for baseline imbalances, encompassing demographic and tumor characteristics such as age, race, marital status, income, location, laterality, histology, grade, TNM stage, tumor size, breast cancer subtype, and axillary lymph node status. However, treatment variables were not adjusted to prevent the introduction of unmeasured confounding. Comprehensive model performances are detailed in **Table 1**.

Integrated Brier Score (IBS) was employed to quantify the discrepancies between the predicted and actual survival distributions in both factual and counterfactual scenarios. Within the testing sets, CPH exhibited superior discrimination, the IBS values were 0.12 (95% CI, 0.11-0.13) for the non-RT group (IBS^a) and 0.08 (95% CI, 0.07-0.09) for the RT group (IBS^b). Similarly, in the external testing sets, IBS^a was 0.13 (95% CI, 0.11-0.14) and IBS^b was 0.08 (95% CI, 0.07-0.09), which closely followed by BITES, with respective IBS values in the testing set of 0.12 (95% CI, 0.11-0.132) and 0.08 (95% CI, 0.07-0.09), and in the external testing set of 0.12 (95% CI, 0.11-0.14) and 0.08 (95% CI, 0.08-0.09).

The models predicted factual and counterfactual survival using baseline covariates, leading to ITE and treatment recommendations. Survival benefits from model recommendations were assessed by comparing the protective effects of Consis. group versus Inconsis. group. Metrics used to evaluate model performance were adjusted with IPTW, minimizing the influence of other prognostic factors. Additionally, comparisons were made with NCCN guidelines, which advise pT2 patients with grade 3 tumors to receive adjuvant RT. For HR+ and HER2- patients over 70 with pT1 disease or over 65 with pT ≤ 3 cm, adjuvant RT may be omitted. Comparisons were drawn between patients whose treatment aligned with NCCN guidelines and those who did not.

In the testing set, following the DSME recommendation led to the most significant survival improvement (IPTW-adjusted HR: 0.70, 95% CI, 0.58-0.86; IPTW-adjusted RD: 4.63, 95% CI, 1.59-7.66; IPTW-adjusted DRMST: 8.96, 95% CI, 6.85-10.97). CMHE showed the best HR results (0.70, 95% CI, 0.57-0.85; IPTW-adjusted HR: 0.70, 95% CI, 0.57-0.86). Adhering to NCCN guidelines resulted in an increase in 10-year survival (IPTW-adjusted DRMST: 4.09, 95% CI, 1.26-6.93).

In the external testing set, DSME demonstrated superior performance (IPTW-adjusted HR: 0.68, 95% CI, 0.56-0.83; IPTW-adjusted RD: 10.50, 95% CI, 5.82-15.10; IPTW-adjusted DRMST: 11.21, 95% CI, 8.35-14.78), outperforming BITES (IPTW-adjusted HR: 0.76, 95% CI, 0.62-0.92; IPTW-adjusted RD: 6.90, 95% CI, 2.26-11.50; IPTW-adjusted DRMST: 3.77, 95% CI, 0.47-6.87) and CPH (IPTW-adjusted HR: 0.81, 95% CI, 0.66-0.99; IPTW-adjusted RD: 9.26, 95% CI, 4.55-14.00; IPTW-adjusted DRMST: 10.47, 95% CI, 7.19-13.44). Following NCCN guidelines also extended restricted mean survival time (IPTW-adjusted DRMST: 4.41, 95% CI, 1.16-7.23), although NCCN guidelines did not show a protective effect in multivariate metrics such as HR (0.94, 95% CI, 0.85-1.17) and IPTW-adjusted HR (0.93, 95% CI, 0.90-1.35). Thus, DSME emerged as the most effective treatment recommendation tool, surpassing other models and NCCN guidelines in both testing environments.

The KM curves for DSME's recommended Consis. group versus the Inconsis. group are displayed in **Figure 2A** (OS in the testing set), **Figure 2B** (BCSS in the testing set), **Figure 2C** (OS in the external testing set), and **Figure 2D** (BCSS in the external testing set). The curves illustrate a significant survival advantage for the Consis. group, with better OS (*p* value of IPTW-adjusted Log-rank test in both testing

Table 1. Model performance

Model	IBS[a]	IBS[b]	HR	IPTW-adjusted HR	RD (%)	IPTW-adjusted RD (%)	DRMST (month)	IPTW-adjusted DRMST (month)
Performance in the testing set								
DSME	0.16 (0.11-0.13)	0.08 (0.07-0.09)	0.70 (0.57-0.85)	0.70 (0.58-0.86)	**8.60 (5.43-11.80)**	**4.63 (1.59-7.66)**	**8.64 (5.95-11.32)**	**8.96 (6.85-10.97)**
BITES	0.12 (0.11-0.13)	0.08 (0.07-0.09)	0.73 (0.60-0.89)	0.74 (0.60-0.90)	7.67 (4.52-10.80)	3.87 (0.83-6.91)	7.59 (4.93-10.25)	6.08 (4.20-8.95)
CMHE	0.19 (0.17-0.20)	0.16 (0.15-0.17)	**0.70 (0.57-0.85)**	**0.70 (0.57-0.86)**	8.34 (5.17-11.50)	4.49 (1.45-7.53)	8.49 (5.81-11.18)	1.97 (-2.17-5.04)
DeepSurv	0.17 (0.16-0.18)	0.16 (0.15-0.17)	0.72 (0.59-0.87)	7.30 (0.60-0.89)	8.02 (4.86-11.20)	4.16 (1.12-7.19)	8.11 (5.43-10.78)	7.96 (5.85-10.17)
CPH	**0.12 (0.11-0.13)**	**0.08 (0.07-0.09)**	0.80 (0.66-0.97)	0.83 (0.68-1.02)	6.98 (3.76-10.20)	3.53 (0.45-6.60)	6.84 (4.12-9.57)	6.99 (4.13-9.25)
NCCN			0.93 (0.81-1.15)	0.92 (0.81-1.10)	4.00 (0.81-7.37)	3.91 (-1.65-7.17)	4.62 (1.57-7.48)	4.09 (1.26-6.93)
Performance in the external testing set								
DSME	0.50 (0.49-0.53)	0.42 (0.41-0.43)	**0.68 (0.56-0.83)**	**0.68 (0.56-0.83)**	**16.00 (11.10-20.90)**	**10.50 (5.82-15.10)**	**11.06 (7.67-14.45)**	**11.21 (8.35-14.78)**
BITES	**0.12 (0.11-0.14)**	0.08 (0.08-0.09)	0.74 (0.61-0.91)	0.76 (0.62-0.92)	13.20 (8.23-18.10)	6.90 (2.26-11.50)	9.91 (6.51-13.31)	3.77 (0.47-6.87)
CMHE	0.18 (0.17-0.19)	0.14 (0.13-0.15)	0.74 (0.61-0.90)	0.75 (0.62-0.92)	13.90 (8.93-18.90)	8.30 (3.65-12.90)	10.42 (6.98-13.86)	-1.25 (-3.62-1.47)
DeepSurv	0.12 (0.11-0.14)	0.08 (0.07-0.09)	0.74 (0.61-0.90)	0.75 (0.62-0.92)	13.80 (8.88-18.80)	8.25 (3.59-12.90)	10.42 (6.98-13.86)	-1.29 (-4.37-1.50)
CPH	0.13 (0.11-0.14)	**0.08 (0.07-0.09)**	0.78 (0.64-0.96)	0.81 (0.66-0.99)	14.90 (9.84-19.90)	9.26 (4.55-14.00)	10.26 (6.80-13.73)	10.47 (7.19-13.44)
NCCN			0.94 (0.85-1.17)	0.93 (0.90-1.35)	4.80 (-0.41-7.37)	3.91 (-1.65-7.17)	3.96 (0.69-6.95)	4.41 (1.16-7.23)

DSME, Deep Survival regression with Mixture Effects; BITES, Balanced Individual Treatment Effect for Survival data; CMHE, Cox Mixtures with Heterogeneous Effects; CPH, Cox proportional hazards model; NCCN, National Comprehensive Cancer Network treatment guidelines; IBS, integrated Brier score; HR, hazard ratio; RD, 10-year risk difference; DRMST, the difference in the 10-year restricted mean survival time; a, integrated Brier score in the non-radiation group; b, integrated Brier score in the adjuvant radiation group. Bolded font indicates that the model performs best in this metric. NCCN guidelines recommend pT2 patients with grade 3 or grade 3 to receive adjuvant RT. As for patients with HR+ and HER2-, they are suggested to consider omit adjuvant RT if they are over 70 years old and with pT1 disease or they are over 65 years old and with pT no bigger than 3 cm.

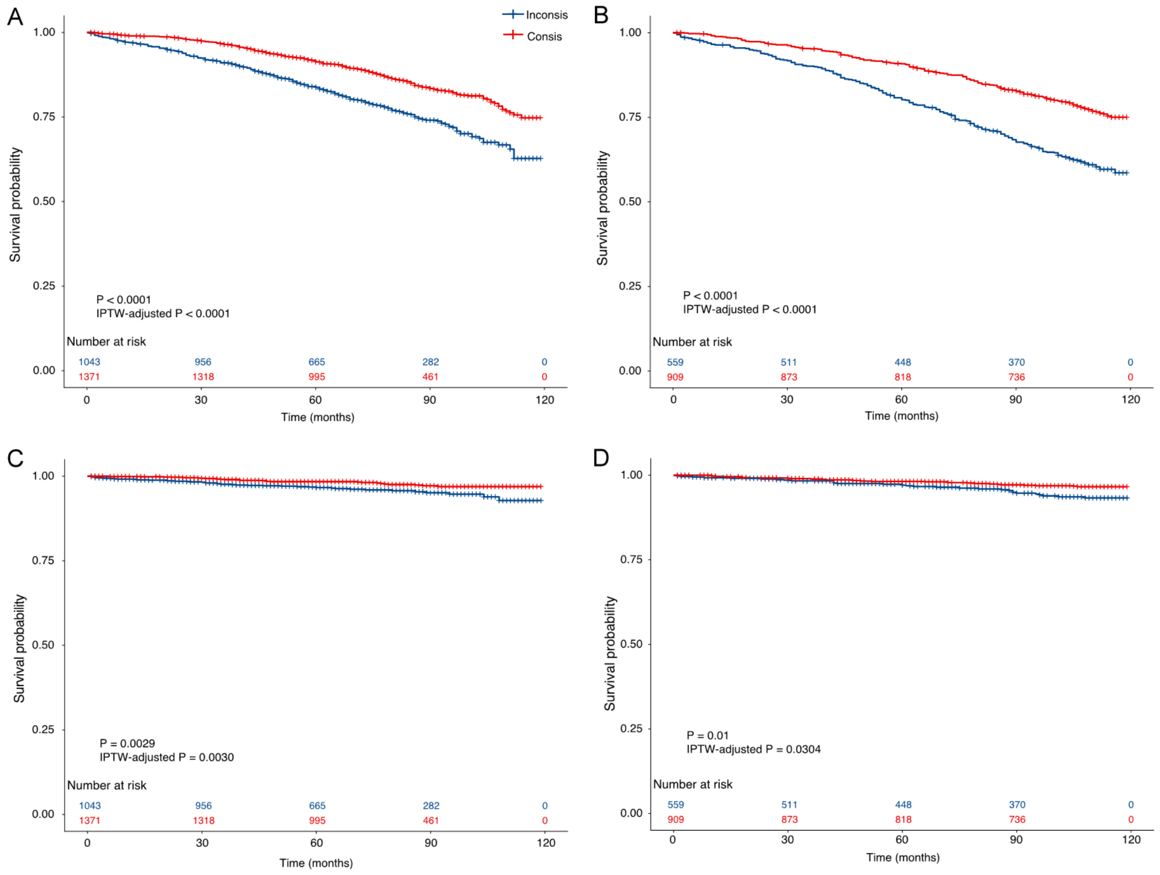


Figure 2. The Kaplan-Meier curves of Consis. and Inconsis. groups. A: The Kaplan-Meier curves of Consis. versus Inconsis. groups of overall survival in the testing set. B: The Kaplan-Meier curves of Consis versus Inconsis groups of breast cancer-specific survival in the testing set. C: The Kaplan-Meier curves of Consis. versus Inconsis. groups of overall survival in the external testing set. D: The Kaplan-Meier curves of Consis versus Inconsis groups of breast cancer-specific survival in the external testing set. *P* value was calculated using Log-rank test; IPTW-adjusted *P* values was calculated using inverse probability treatment weighting-adjusted Log-rank test.

sets < 0.0001) and BCSS (*p* value of IPTW-adjusted Log-rank test in the testing set =0.0030; in the external testing set =0.0304).

The potential impact of treatment proportion imbalances on DSME's protective effect was scrutinized by calculating the interventional natural direct effect (INDE) and the interventional natural indirect effect (INIE), as initially suggested by Diaz et al. [24]. Treatment variables were considered as mediators, and adjustments were made for baseline characteristics. Figure S1A and S1B display the INDE and INIE for the testing and external testing sets, respectively, expressed as slopes in a linear regression model. The DSME recommendation influenced OS directly, with INDE values of -0.15 (95% CI, -0.20 to -0.10) in the testing set and -0.24 (95% CI, -0.26 to -0.21) in the external testing set, and INIE values of 0.07 (95% CI, 0.01 to 0.12) and 0.08 (95% CI, 0.05 to 0.10), respectively, indicating that these effects were not mediated by the treatments administered.

Further, the standardized mean difference (SMD) before and after IPTW adjustment is depicted in Figure S2A and S2B for the testing and external testing sets. The IPTW correction successfully balanced the covariates, achieving an SMD of less than 0.1 between groups, thereby demonstrating effective control of confounding variables across both sets [25].

Treatment heterogeneity

The heterogeneity of treatment effects was explored by assessing variations in the average treatment effect (ATE) across different patient

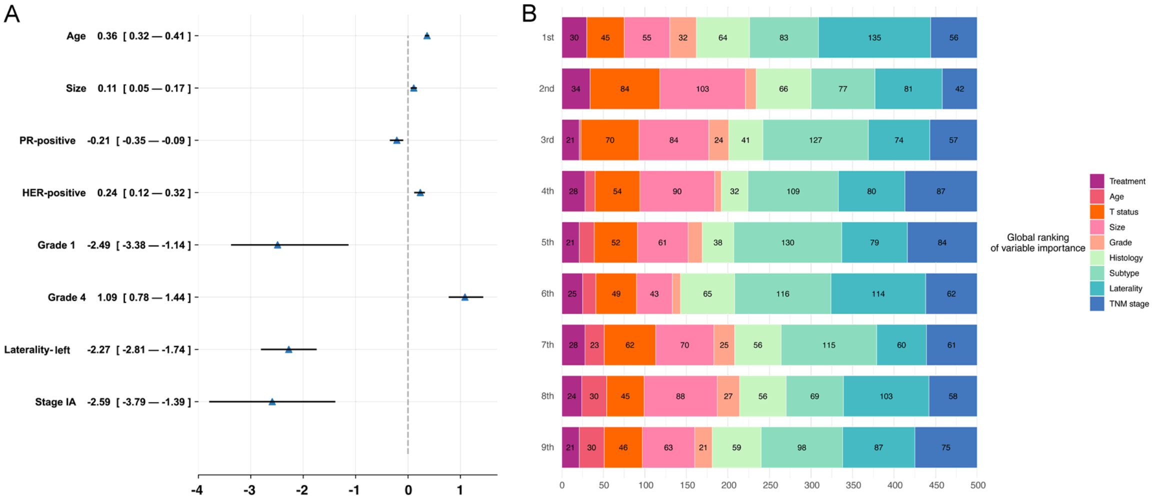


Figure 3. Model interpretation. A: Interpretation of model recommendation behavior. B: Interpretation of overall output using SurvSHAP(t).

subgroups, revealing differential responses to the same treatment based on distinct patient characteristics. Patients were categorized into groups based on whether DSME and NCCN guidelines recommended radiotherapy. This analysis encompassed both the testing and external testing populations. Figure S3A and S3B illustrate the HR and IPTW-adjusted HR for patients recommended and not recommended for RT, respectively.

For patients advised to receive radiotherapy, both DSME and NCCN guidelines showed a protective effect, with IPTW-adjusted HR of 0.71 (95% CI, 0.62-0.82) and 0.40 (95% CI, 0.17-0.94) respectively. Conversely, among those advised against radiotherapy, it emerged as a risk factor when DSME indicated a negative ITE, with an IPTW-adjusted HR of 1.84 (95% CI, 1.09-2.90). Notably, the NCCN guidelines failed to discern patients who would not benefit from RT, highlighting a critical limitation in identifying suitable candidates for this treatment.

Deep learning-based treatment insights

ITE values quantify the difference in TaR between patients undergoing RT versus those who do not, reflecting the additional time before a patient's mortality risk reaches 50% due to RT. Considering regional differences among patients, a mixed-effect linear regression model was employed. This model, applied to the aggregate data from both testing and external testing sets, predicts ITE based on covariates with the reporting region as a random effect. In this context, the beta coefficients indicate that when holding other variables constant, the presence of a specific covariate or a one-unit increase in it, extends the time until a patient's mortality risk reaches 50% when receiving RT compared to not receiving it. These findings are detailed in **Figure 3A**.

It was observed that for every 1 mm increase in tumor size, adjuvant RT extended a patient's survival by an additional 0.11 months (95% CI: 0.05-0.17) over 10 years. Furthermore, adjuvant RT was more effective in patients of advanced age (0.36, 95% CI, 0.32-0.41), those with HER-positive status (0.24, 95% CI, 0.12-0.32), and those with grade 4 tumors (1.09, 95% CI, 0.78-1.44). Conversely, adjuvant RT was less beneficial, and not recommended, for patients with PR-positive status (-0.21, 95% CI, -0.35 to -0.09), grade 1 tumors (-2.49, 95% CI, -3.38 to -1.14), stage IA cancer (-2.59, 95% CI, -3.79 to -1.39), and tumors located in the left breast (-2.27, 95% CI, -2.81 to -1.74).

Model interpretation

We employed SurvSHAP(t) to analyze the functional outputs of DSME, marking the first instance of using this method to provide a time-dependent interpretation based on a robust theoretical foundation [26]. **Figure 3B** displays the accumulated influence of the eight most critical variables, sorted by aggregated Shapley values across 500 observations. The horizontal bars chart the frequency with which each variable ranks in importance, from highest to lowest, with distinct colors marking the ranks.

Laterality emerged as the paramount prognostic factor in 135 cases, surpassing other significant factors such as tumor size, breast cancer subtype, and TNM stage.

Additionally, we conducted a case study on a randomly selected patient from the testing set, with the findings presented in Figure S4. Through DSME analysis, this patient's survival probabilities under various treatment scenarios were distinctly illustrated. This approach enables the calculation of several survival metrics, such as differences in mortality, TaR, and restricted survival time, aiding patients in making informed decisions about the most suitable treatment options.

Discussion

The routine use of adjuvant RT after BCS in elderly patients with EBC remains a subject of debate [8]. This controversy arises from the variable absolute benefits of RT, which depend significantly on individual patient characteristics and are particularly contentious among elderly patients due to their increased frailty and risk of complications [27]. Omitting adjuvant RT can undoubtedly spare patients from side effects such as breast pain, dermatitis, and potential cardiac and pulmonary risks, while also reducing time and financial burdens [28]. Therefore, it is crucial to balance the need to avoid overtreatment with the necessity of not compromising patient survival.

In this study, we thoroughly evaluated the DSME model, which demonstrated superior performance compared to state-of-the-art or commonly used models, real-world physician choices, and NCCN guidelines. After rigorously adjusting for bias, following DSME recommendations extended patient survival by 11 months within a 10-year period, a significant improvement compared to those who did not follow the recommendations. Although NCCN guidelines also extended survival and effectively identified suitable candidates for radiotherapy, these benefits were statistically significant only in univariate analyses.

Treatment decisions often require an understanding of complex interactions among features rather than reliance on static guidelines [29]. Our study highlights that DL models, such as DSME, are particularly adept at managing this complexity, as evidenced by their more robust protective effects compared to NCCN guidelines. While CPH showed better discrimination, it did not surpass the protective capabilities of other models incorporating advanced causal inference techniques. This suggests that accurate prognosis prediction is crucial, but integrating statistical methods to derive unbiased ITE from observational data is equally vital for effective treatment recommendations.

The nature of artificial intelligence-guided intervention studies allows us to glean insights into DL-based treatment recommendations by analyzing model behaviors associated with ITE values. In our study, we controlled for potential confounders by maintaining other covariates constant. Consequently, our findings are largely independent of confounding variables, compared to outcomes derived from traditional methods. This independence from confounders not only makes the results quantifiable but also provides a crucial foundation for visualizing how baseline characteristics influence the relative efficacy of RT. This approach enhances our understanding of treatment dynamics in a way that is directly applicable to clinical decision-making.

In line with clinical consensus [3, 6], elderly patients with higher-risk features, such as larger tumor sizes, HER2-positive status, and grade 4 tumors, are found to benefit from adjuvant RT. Patients in older age groups with more advanced disease are even more likely to benefit, possibly due to a balance between their life expectancy and RT toxicity. Conversely, lower-risk patients [8], including those with hormone receptor (HR)-positive status, grade 1 tumors, and stage IA, are generally advised to forgo adjuvant RT. Additionally, elderly EBC patients with tumors located in the left breast may suffer adverse effects from adjuvant RT, potentially due to the increased risk of cardiac disease linked to left-sided RT [30].

Clinicians and patients need effective tools to discuss various treatment options, particularly those that clearly highlight survival benefits. Developing a graphical treatment recommendation system that showcases individual survival metrics and comparative analyses can greatly simplify the understanding of complex data for patients, their families, and healthcare providers. Traditionally, it has been challenging to precisely identify which patients would benefit most from specific treatments and to provide individualized post-treatment outcome forecasts [9, 31]. Most existing models tend to rely on patient characteristics to establish prognostic factors, potentially introducing biases toward certain treatments [32]. The DSME model excels in addressing these challenges by offering a clear method to communicate tailored outcomes based on different treatment scenarios, thus fulfilling a critical need in clinical decision-making.

Limitations

This study has several noteworthy limitations. To begin with, the SEER database, while extensive, lacks access to several critical variables, including Ki67, BRCA status, positive margin presence, and detailed information on RT administration. These biological markers play a crucial role in accurately assessing the prognosis and treatment outcomes of elderly breast cancer patients. Without them, the depth of our analysis is somewhat restricted, and the precision of the DSME model's treatment recommendations may be compromised. In addition, the SEER database does not provide data on quality of life (QoL) or progression-free survival, which are essential for evaluating the broader impact of treatments, especially for elderly patients. As a result, our study focuses primarily on survival outcomes, leaving a gap in understanding how treatments affect patient well-being.

Another limitation stems from the retrospective design of this study, which inherently introduces potential selection bias and unmeasured confounding factors. Relying on previously collected data limits control over the quality and consistency of the variables, which could influence the accuracy of treatment outcome assessments. Moreover, the absence of important patient information, such as comorbidities, which are particularly relevant in older populations, further complicates the ability to control for all variables and fully account for patient health profiles.

Finally, while the DSME model demonstrated robust statistical performance, it has not yet been validated in real-world clinical settings. The absence of prospective clinical trials or real-world studies limits the model's current generalizability and its practical application in clinical decision-making. Moving forward, future research should prioritize validating the DSME model through such trials and real-world implementations to ensure its effectiveness in broader clinical environments. Additionally, incorporating missing biological markers, QoL data, and prospective validation would greatly enhance the model's reliability and improve the comprehensiveness of its recommendations for elderly breast cancer patients.

Conclusions

This study is the first to evaluate a DL-guided method for selecting adjuvant RT in elderly EBC patients. The DSME model provides crucial insights, suggesting that adjuvant RT benefits patients with larger tumors, advanced age, or later stages, while those with early-stage, HR-positive disease, or left-sided tumors might omit it.

Disclosure of conflict of interest

None.

Address correspondence to: Jinchao Yue, Department of Oncology, Dongying District People's Hospital, 333 Jinan Road, Dongying District, Dongying 257000, Shandong, China. E-mail: yjcqiang@163.com

References

[1] Singh D, Vignat J, Lorenzoni V, Eslahi M, Ginsburg O, Lauby-Secretan B, Arbyn M, Basu P, Bray F and Vaccarella S. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO global cervical cancer elimination initiative. Lancet Glob Health 2023; 11: e197-e206.

[2] Wang F, Meszoely I, Pal T, Mayer IA, Bailey CE, Zheng W and Shu XO. Radiotherapy after breast-conserving surgery for elderly patients with early-stage breast cancer: a national registry-based study. Int J Cancer 2021; 148: 857-867.

[3] Stueber TN, Diessner J, Bartmann C, Leinert E, Janni W, Herr D, Kreienberg R, Woeckel A and Wischnewsky M. Effect of adjuvant radiotherapy in elderly patients with breast cancer. PLoS One 2020; 15: e0229518.

[4] Saiki H, Petersen IA, Scott CG, Bailey KR, Dunlay SM, Finley RR, Ruddy KJ, Yan E and Redfield MM. Risk of heart failure with preserved ejection fraction in older women after contemporary radiotherapy for breast cancer. Circulation 2017; 135: 1388-1396.

[5] Land LH, Dalton SO, Jensen MB and Ewertz M. Influence of comorbidity on the effect of adjuvant treatment and age in patients with early-stage breast cancer. Br J Cancer 2012; 107: 1901-1907.

[6] Gradishar WJ, Moran MS, Abraham J, Abramson V, Aft R, Agnese D, Allison KH, Anderson B, Burstein HJ, Chew H, Dang C, Elias AD, Giordano SH, Goetz MP, Goldstein LJ, Hurvitz SA, Jankowitz RC, Javid SH, Krishnamurthy J, Leitch AM, Lyons J, Mortimer J, Patel SA, Pierce LJ, Rosenberger LH, Rugo HS, Schneider B, Smith ML, Soliman H, Stringer-Reasor EM, Telli ML, Wei M, Wisinski KB, Young JS, Yeung K, Dwyer MA and Kumar R. NCCN guidelines® insights: breast cancer, version 4.2023. J Natl Compr Canc Netw 2023; 21: 594-608.

[7] Tang L, Matsushita H and Jingu K. Controversial issues in radiotherapy after breast-conserving surgery for early breast cancer in older patients: a systematic review. J Radiat Res 2018; 59: 789-793.

[8] Kunkler IH, Williams LJ, Jack WJL, Cameron DA and Dixon JM. Breast-conserving surgery with or without irradiation in early breast cancer. N Engl J Med 2023; 388: 585-594.

[9] Lei L and Candès EJ. Conformal inference of counterfactuals and individual treatment effects. J R Stat Soc Series B Stat Methodol 2020; 83.

[10] Yao L, Chu Z, Li S, Li Y, Gao J and Zhang A. A survey on causal inference. ACM Trans Knowl Discov Data 2020; 15: 1-46.

[11] Zhu E, Chen Z, Ai P, Wang J, Zhu M, Xu Z, Liu J and Ai Z. Analyzing and predicting the risk of death in stroke patients using machine learning. Front Neurol 2023; 14: 1096153.

[12] She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, Jiang G, Liu H, Xie D, Cao N, Ren Y and Chen C. Development and Validation of a deep learning model for non-small cell lung cancer survival. JAMA Netw Open 2020; 3: e205842.

[13] Zhu E, Shi W, Chen Z, Wang J, Ai P, Wang X, Zhu M, Xu Z, Xu L, Sun X, Liu J, Xu X and Shan D. Reasoning and causal inference regarding surgical options for patients with low-grade gliomas using machine learning: a SEER-based study. Cancer Med 2023; 12: 20878-20891.

[14] Hankey BF, Ries LA and Edwards BK. The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev 1999; 8: 1117-1121.

[15] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC and Vandenbroucke JP; STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370: 1453-1457.

[16] Künzel SR, Sekhon JS, Bickel PJ and Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci U S A 2019; 116: 4156-4165.

[17] Schrod S, Schäfer A, Solbrig S, Lohmayer R, Gronwald W, Oefner PJ, Beissbarth T, Spang R, Zacharias HU and Altenbuchinger M. BITES: balanced individual treatment effect for survival data. Bioinformatics 2022; 38 Suppl 1: i60-i67.

[18] Li F, Morgan KL and Zaslavsky AM. Balancing covariates via propensity score weighting. J Am Stat Assoc 2014; 113: 390-400.

[19] Johansson FD, Shalit U, Kallus N and Sontag DA. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. J Mach Learn Res 2022; 23: 1-50.

[20] Zhu E, Zhang L, Wang J, Hu C, Jing Q, Shi W, Xu Z, Ai P, Dai Z, Shan D and Ai Z. Personalized surgical recommendations and quantitative therapeutic insights for patients with metastatic breast cancer: insights from deep learning. Cancer Innov 2024; 3: e119.

[21] Nagpal C, Yadlowsky S, Rostamzadeh N and Heller KA. Deep cox mixtures for survival regression. ArXiv 2021; abs/2101.06536.

[22] Nagpal C, Goswami M, Dufendach KA and Dubrawski AW. Counterfactual phenotyping with censored time-to-events. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2022; 3634-3644.

[23] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T and Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2016; 18: 24.

[24] Díaz I, Hejazi NS, Rudolph KE and Laan MJVD. Non-parametric efficient causal mediation with intermediate confounders. Biometrika 2021; 108: 627-641.

[25] Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biom J 2009; 51: 171-184.

[26] Krzyzi'nski M, Spytek M, Baniecki H and Biecek P. SurvSHAP(t): time-dependent explanations of machine learning survival models. Knowl Based Syst 2022; 262: 110234.

[27] Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, Cutter D, Da-

vies C, Ewertz M, Godwin J, Gray R, Pierce L, Whelan T, Wang Y and Peto R. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet 2011; 378: 1707-1716.

[28] Ho AY and Bellon JR. Overcoming resistance - omission of radiotherapy for low-risk breast cancer. N Engl J Med 2023; 388: 652-653.

[29] Pan H, Wang J, Shi W, Xu Z and Zhu E. Quantified treatment effect at the individual level is more indicative for personalized radical prostatectomy recommendation: implications for prostate cancer treatment using deep learning. J Cancer Res Clin Oncol 2024; 150: 67.

[30] Cheng YJ, Nie XY, Ji CC, Lin XX, Liu LJ, Chen XM, Yao H and Wu SH. Long-term cardiovascular risk after radiotherapy in women with breast cancer. J Am Heart Assoc 2017; 6: e005633.

[31] Zhu E, Zhang L, Wang J, Hu C, Pan H, Shi W, Xu Z, Ai P, Shan D and Ai Z. Deep learning-guided adjuvant chemotherapy selection for elderly patients with breast cancer. Breast Cancer Res Treat 2024; 205: 97-107.

[32] Di Ieva A. AI-augmented multidisciplinary teams: hype or hope? Lancet 2019; 394: 1801.

Table S1. Patients

	No radiotherapy (n=3,399)	Radiotherapy (n=4,648)
Age, median (IQR), y	74.0 (69.0-81.0)	72.0 (68.0-77.0)
Tumor size, median (IQR), mm	11.0 (8.0-16.0)	10.0 (7.0-15.0)
Race-White	2,908 (85.6)	4,066 (87.5)
Marriage-Married	1,431 (42.1)	2,324 (50.0)
Income-Higher than 70,000$	1,092 (32.1)	1,725 (37.1)
Laterality-Right	1,640 (48.2)	2,276 (49.0)
Laterality-Left	1,759 (51.8)	2,372 (51.0)
Human epidermal growth factor receptor 2-positive	263 (7.7)	268 (5.8)
Estrogen receptor-positive	3,091 (90.9)	3.834 (82.5)
Progesterone receptor-positive	2,762 (81.3)	3,444 (74.1)
Grade		
G1	1,209 (35.6)	1,587 (34.1)
G2	1,598 (47.0)	2,174 (46.8)
G3	586 (17.2)	882 (19.0)
G4	6 (0.2)	5 (0.1)
Histology		
Ductal	2,898 (85.3)	4,004 (86.1)
Lobular	314 (9.2)	405 (8.7)
Ductal-lobular	187 (5.5)	239 (5.1)
Location		
Upper outer quadrant	1,160 (34.1)	1,759 (37.8)
Upper inner quadrant	536 (15.8)	698 (15.0)
Lower outer quadrant	241 (7.1)	311 (6.7)
Lower inner quadrant	221 (6.5)	338 (7.3)
Central/overlapping	923 (27.2)	1,281 (27.6)
Nipple/axillary tail	20 (0.6)	34 (0.7)
Breast/NOS	298 (8.8)	227 (4.9)
T stage		
T1a	412 (12.1)	735 (15.8)
T1b	1,119 (32.9)	1,677 (36.1)
T1c	1,450 (42.7)	1,779 (38.3)
T1mic	27 (0.8)	50 (1.1)
T2	391 (11.5)	407 (8.8)
TNM stage		
IA	3,008 (88.5)	4,241 (91.2)
IB	0 (0.0)	0 (0.0)
IIA	391 (11.5)	407 (8.8)

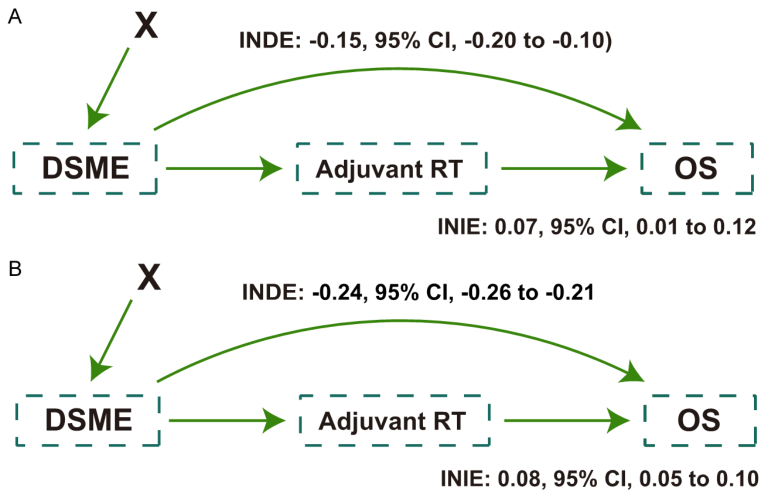

A
X
INDE: -0.15, 95% CI, -0.20 to -0.10)
DSME
Adjuvant RT
OS
INIE: 0.07, 95% CI, 0.01 to 0.12
B
X
INDE: -0.24, 95% CI, -0.26 to -0.21
DSME
Adjuvant RT
OS
INIE: 0.08, 95% CI, 0.05 to 0.10

Figure S1. Causal path of Self-Normalizing Balanced individual treatment effect for survival data recommendations. A: Causal path of Self-Normalizing Balanced individual treatment effect for survival data recommendations in the testing set. B: Causal path of Self-Normalizing Balanced individual treatment effect for survival data recommendations in the external testing set. DSME, Deep Survival Regression with Mixture Effects; INDE, interventional natural direct effect; INIE, interventional natural indirect effect; X indicates patients' baseline features, which were adjusted as intermediate confounders.

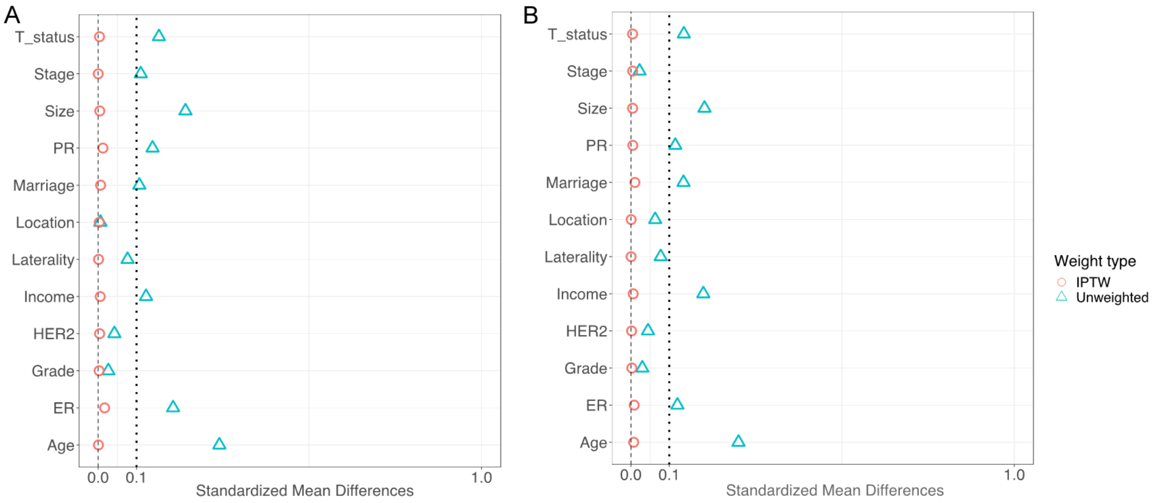


Figure S2. The standardized mean difference before and after inverse probability treatment weighting. A: The standardized mean difference before and after inverse probability treatment weighting in the testing set. B: The standardized mean difference before and after inverse probability treatment weighting in the external testing set. IPTW, inverse probability treatment weighting.

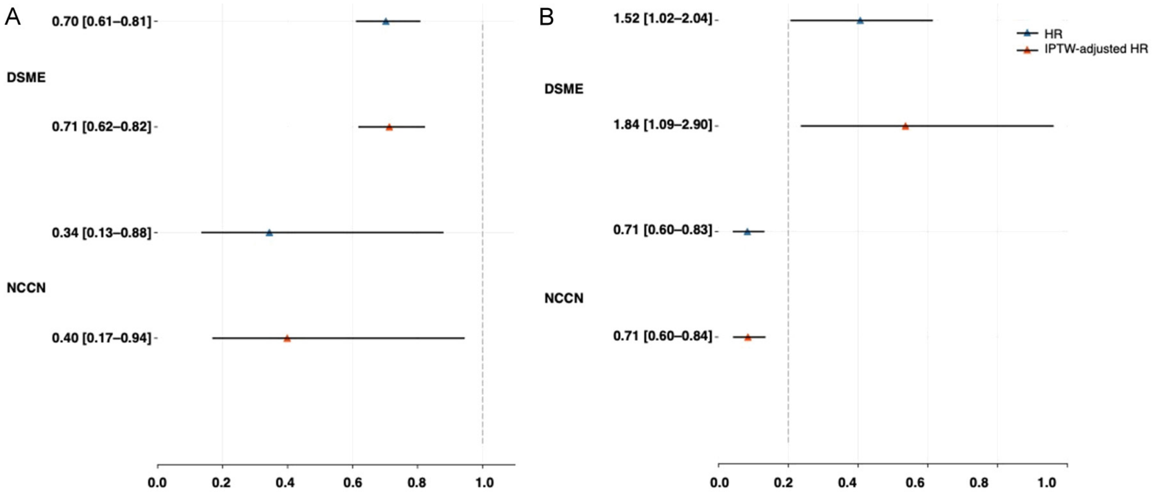
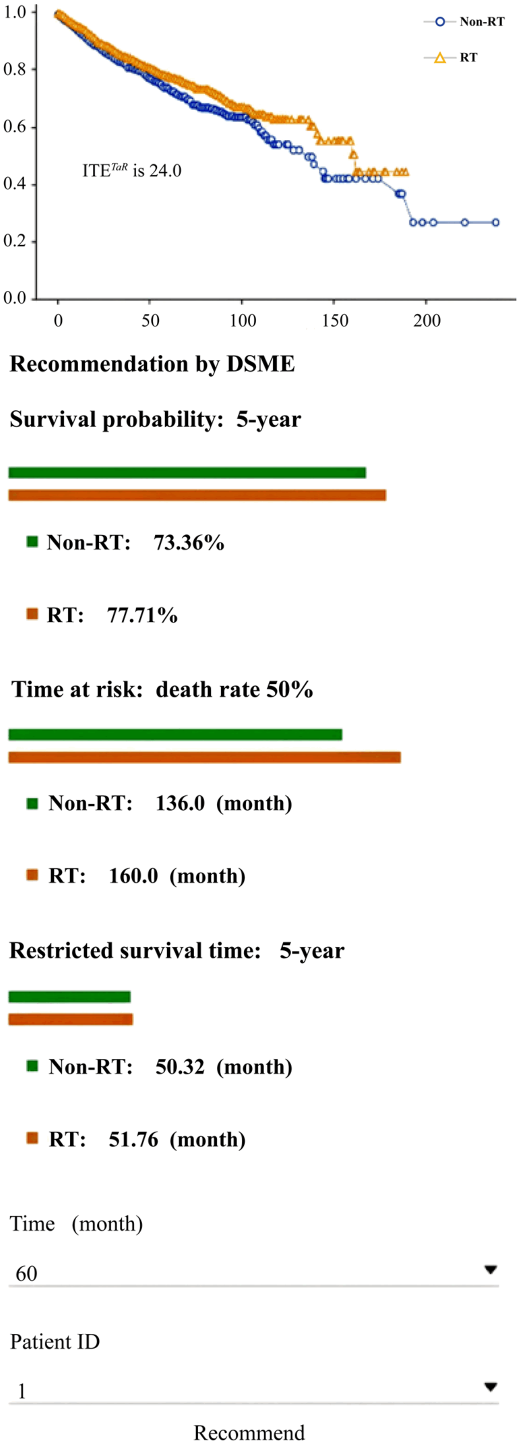


Figure S3. Average treatment effect and treatment heterogeneity. A: Average treatment effect and treatment heterogeneity in the testing set. B: Average treatment effect and treatment heterogeneity in the external testing set.



Recommendation by DSME

Survival probability: 5-year

🟩 **Non-RT: 73.36%**

🟧 **RT: 77.71%**

Time at risk: death rate 50%

🟩 **Non-RT: 136.0 (month)**

🟧 **RT: 160.0 (month)**

Restricted survival time: 5-year

🟩 **Non-RT: 50.32 (month)**

🟧 **RT: 51.76 (month)**

Time (month)

60 ▼

Patient ID

1 ▼

Recommend

Figure S4. The individual survival distribution predicted by model. RT, radiotherapy; ITE, individual treatment effect; TAR, time at risk; DSME, Deep Survival Regression with Mixture Effects.