

Multitask learning for multi-endpoint prognostic assessment in early breast cancer using routine electronic medical records

Ao Fan

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China

Rui Yuan

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China

Dan Zheng

Institute for Breast Health Medicine, Cancer Center, Breast Disease Center, West China Hospital, Sichuan University

Jingjing Xiao

Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth Military Medical University

Jiaheng Feng

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China

Zhongni Ye

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China

Chenyu Xiong

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China

Tingyu Qu

Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth Military Medical University

Junpeng Ren

Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth Military Medical University

Ting Luo

Institute for Breast Health Medicine, Cancer Center, Breast Disease Center, West China Hospital, Sichuan University

Jiajun Ding

Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth Military Medical University

Xiaorong Zhong

Institute for Breast Health Medicine, Cancer Center, Breast Disease Center, West China Hospital,
Sichuan University

Xiaowei Mao

xiaowei.mao@uestc.edu.cn

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of
China

Article

Keywords:

Posted Date: May 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9745774/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Multitask learning for multi-endpoint prognostic assessment in early breast cancer using routine electronic medical records

Ao Fan^{1†}, Rui Yuan^{1†}, Dan Zheng^{2, 3†}, Jingjing Xiao^{4†}, Jiaheng Feng¹, Zhongni Ye¹,
Chenyu Xiong¹, Tingyu Qu⁴, Junpeng Ren⁴, Ting Luo^{2, 3}, Jiajun Ding^{4*}, Xiaorong
Zhong^{2, 3*}, Xiaowei Mao^{1, 5*}

¹ Institute of Fundamental and Frontier Sciences, University of Electronic Science and
Technology of China, Chengdu, Sichuan, 610072, China

² Institute for Breast Health Medicine, Cancer Center, Breast Disease Center, West
China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu, Sichuan, 610041,
China

³ Sichuan Provincial Engineering Research Center of Intelligent Diagnosis and
Treatment of Breast Diseases, Chengdu, Sichuan, 610000, China

⁴ Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth
Military Medical University, Xi'an, Shaanxi Province, 710032, P.R. China

⁵ Tianfu Jincheng Laboratory, Chengdu, 610093, China

[†] **These authors contributed equally**

*** Corresponding authors:**

Xiaowei Mao (xiaowei.mao@uestc.edu.cn); Xiaorong
Zhong(zhongxiaorong@126.com); Jiajun Ding (yuki_ding1996@163.com)

ABSTRACT

Accurate prognostic prediction in early breast cancer is essential for tailoring adjuvant
therapy, yet existing clinical tools remain limited in their ability to integrate high-
dimensional electronic medical record (EMR) data for joint multi-endpoint risk

assessment. We developed EMR-BCEP, a multitask deep learning framework using routine breast cancer EMR data and incorporating the Multilabel Synthetic Minority Over-sampling Technique (MLSMOTE) and dynamic loss weighting optimization (DLWO) to simultaneously predict five clinically relevant prognostic endpoints. EMR-BCEP was developed and internally tested in a cohort of 7,457 patients and externally validated in an independent cohort of 1,125 patients. In the internal test set, EMR-BCEP outperformed corresponding single-task models, with AUROC gains of 0.12, 0.06, 0.04, 0.05 and 0.07 across the five endpoints, and its performance was largely preserved in external validation. To support clinical translation, EMR-BCEP has been deployed as an online interactive platform, freely accessible at https://maolab.tech/breast_cancer.

INTRODUCTION

Breast cancer is the most common malignancy among women worldwide and represents the second leading cause of cancer-related mortality in female patients. Its incidence has been continuously rising in recent years^{1,2}. According to projections, by 2040, there will be over 3 million new cases of breast cancer annually, with over 1 million deaths each year³. With the increasing emphasis on breast cancer screening and advancements in standardized diagnostic and therapeutic techniques, the cure rate for breast cancer patients has significantly improved^{4,5}. In this context, as survival outcomes continue to improve and patients live longer, breast cancer has increasingly transitioned into a condition requiring long-term disease management, in which the evaluation of treatment response and prognosis plays an important role in shaping survival outcomes and patient quality of life⁶. However, conventional prognostic assessment methods primarily rely on pathological characteristics, biomarkers, and imaging data, which require regular follow-ups and reassessments. These approaches are often time-consuming and susceptible to subjective interpretation. Therefore, there is an urgent need for more precise and efficient decision-support tools to improve the clinical management of breast cancer patients.

Artificial intelligence (AI) has been increasingly applied in breast cancer management, with its most notable impact in the field of early screening. By improving diagnostic accuracy and reducing healthcare costs, AI-based screening systems have demonstrated significant advantages over traditional approaches⁷⁻⁹. Recent studies have shown that AI techniques are also widely utilized in treatment evaluation and prognosis prediction. For instance, Boeri et al. performed an early evaluation of machine learning (ML) models for breast cancer prognosis, predicting local–regional recurrence, distant metastasis, and breast cancer–specific mortality using clinicopathological features¹⁰. Song et al. developed ML models leveraging 15 clinical variables from electronic medical record (EMR) data to predict patient survival¹¹. Teng et al. employed SEER population-based data to construct a nomogram predicting overall survival (OS) and breast cancer-specific survival (BCSS), facilitating individualized risk stratification. Collectively, these studies demonstrate the versatility of ML approaches in predicting diverse prognostic endpoints across clinical and population-based datasets¹². The ability to accurately predict cancer prognosis not only empowers patients with a clearer understanding of their condition but also supports clinicians in making evidence-based decisions and optimizing treatment strategies¹³⁻¹⁵. However, prognostic tasks in breast

cancer, including recurrence risk and survival prediction, are intrinsically correlated because they share overlapping clinical, pathological, and molecular determinants. Most existing approaches are based on separate single-task learning frameworks, in which individual prognostic endpoints are modeled independently or evaluated sequentially. This approach may limit the integration and effective utilization of multidimensional patient data, hindering shared information across related outcomes, and ultimately leading to the loss of valuable cross-task information. In contrast, multitask learning (MTL) enables the simultaneous optimization of related clinical objectives, facilitating knowledge sharing across tasks and improving model generalization. Recent studies have demonstrated that MTL can effectively capture complementary patterns among heterogeneous data modalities and outperform single-task models in predictive performance, offering significant advantages in multi-source information integration and precision medicine applications¹⁶⁻¹⁸.

Although multitask learning provides an effective framework for integrating heterogeneous information and enhancing prognostic modeling in breast cancer, its performance can still be hindered by the inherent class imbalance within medical datasets¹⁹⁻²¹. Such imbalance arises when certain clinical outcomes or disease states are underrepresented, resulting in a majority class of common cases and a minority class of rare but clinically significant events²². Standard AI algorithms are typically optimized under the assumption of balanced data, and this imbalance can bias models toward the majority class while neglecting minority instances, thereby significantly degrading predictive performance. This issue is particularly critical for clinically important but underrepresented outcomes, as it increases the risk of misclassification and limits the model's clinical utility^{23,24}. Consequently, addressing class imbalance is crucial for developing reliable predictive models that are both sensitive and equitable in medical applications. Beyond addressing class imbalance, an additional challenge lies in the effective utilization of EMR data within multitask learning frameworks. Despite the growing interest in data-driven prognostic modeling, few studies have explored the joint prediction of multiple prognostic endpoints using EMR data. Most existing approaches rely on population-based datasets (e.g., SEER) or single-center structured data, which constrains model generalizability and limits clinical transferability.

In this study, we developed EMR-BCEP, a multitask learning framework for predicting multi-endpoint prognosis in breast cancer. Leveraging clinical diagnostic and therapeutic data, the model simultaneously predicts five key outcomes—BCSS,

invasive disease-free survival (iDFS), distant disease-free survival (D-DFS),
locoregional relapse-free survival (L-RFS), and OS. Furthermore, to mitigate class
imbalance, we incorporated a multilabel synthetic over-sampling (MLSMOTE)
strategy with semantic preservation to ensure balanced learning across endpoints²⁵⁻²⁷.
The model was trained and validated using data from two independent centers, and its
predictions were further used to generate clinical decision support recommendations.

RESULTS

EMR-BCEP is a multitask learning model designed to simultaneously predict five clinically relevant prognostic outcomes in early breast cancer using EMR data. Fig.1 illustrates the schematic overview of this study. To rigorously evaluate its predictive capability, the model was trained and tested on the development cohort and subsequently assessed on an independent validation set, ensuring a comprehensive evaluation of its generalizability and robustness.

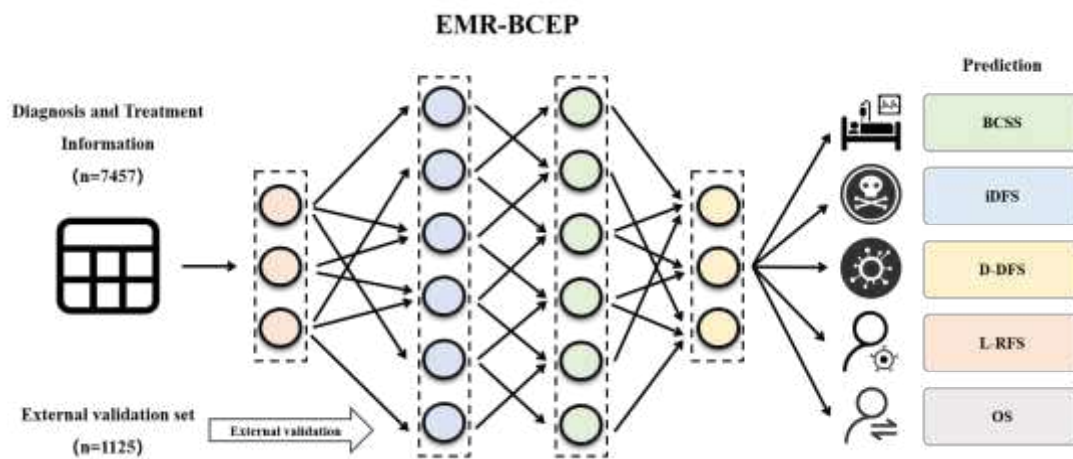


Fig. 1: Schematic Diagram of the Breast Cancer Prognostic Assessment Multitask Model.

EMR-BCEP is a multitask learning model that can predict five endpoint events: BCSS, iDFS, D-DFS, L-RFS, and OS. Performance was evaluated in an internal test set and an independent external validation cohort.

Preparation of breast cancer clinical data

We obtained EMR data from breast cancer patients diagnosed and undergoing postoperative adjuvant therapy from the breast cancer database at West China Hospital, Sichuan University, which were used for model development and performance evaluation. After sample screening and data preprocessing, a total of 7,457 patient records were included. The average age of patients was 48.17 ± 9.64 years, with 7,427 females (99.60%) and 30 males (0.40%). The dataset was randomly split into training and internal test sets at an 8:2 ratio. Additionally, EMR data from 1,125 breast cancer patients in the phase of postoperative adjuvant therapy were collected from the Xijing Hospital, serving as an external validation cohort. This cohort had a mean age of 49.09 ± 10.60 years, including 1,120 females (99.56%) and 5 males (0.44%). Demographic characteristics of the datasets are summarized in Table 1.

Characteristics of the breast cancer clinical data included in this study are as follows: sex, age, ethnicity, education level, weight (kg), height (cm), body surface area (m²), marital status, reimbursement ratio, menopausal status at diagnosis, primary or recurrent status, HER2 FISH, diagnosis, ER, ER percentage, PR, PR percentage, CK5/6, Ki67, HER2, WHO grade, surgery type, surgery name, chemotherapy regimen, radiotherapy mode, T, N, pN, M, and clinical stage. Among these, surgery name and diagnosis were unstructured textual features, while the remaining 28 features were structured. In addition, the frequency distribution of each categorical value was computed to characterize the underlying data composition (Supplementary Table 1).

Table 1 Demographic characteristics of the training, internal test and external validation sets.

Characteristics	Training (n = 5965)	Internal test (n = 1492)	External validation (n = 1125)
Age years, mean $\pm$ SD	48.11 $\pm$ 9.71	48.41 $\pm$ 9.38	49.09 $\pm$ 10.60
Age groups, <i>n</i> (%)			
11–20	1(0.02)	2(0.13)	1(0.09)
21–30	178(2.98)	36(2.41)	33(2.93)
31–40	1072(17.97)	241(16.15)	170(15.11)
41–50	2560(42.92)	655(43.90)	374(33.24)
51–60	1446(24.24)	395(26.47)	313(27.82)
61–70	633(10.61)	140(9.38)	163(14.49)
71–80	71(1.19)	23(1.56)	60(5.33)
81+	4(0.07)	0(0.00)	11(0.99)
Ethnicity, <i>n</i> (%)			
Han	5851(98.09)	1463(98.06)	1109(98.58)
Tibetan	68(1.14)	18(1.21)	2(0.18)
Yi	13(0.22)	4(0.27)	0(0)
Qiang	11(0.18)	3(0.20)	0(0)
Hui	7(0.12)	1(0.07)	12(1.07)
Others	15(0.25)	3(0.19)	2(0.17)
Sex, <i>n</i> (%)			
Female	5938(99.55)	1489(99.80)	1120(99.56)
Male	27(0.45)	3(0.20)	5(0.44)

n: Number of patients.

Preprocessing of prognostic endpoints

The five endpoints in this study—BCSS, iDFS, D-DFS, L-RFS, and OS—were modeled as binary classification tasks. Endpoint assessment was conducted within a fixed 5-year follow-up period after surgery, and all events were defined as those occurring within 5 years postoperatively. Patients without events during the 5-year follow-up were assigned a label of 0 (favorable outcome), whereas those who experienced events within this period were assigned a label of 1 (unfavorable outcome). Event determination was conducted by experienced breast cancer clinicians based on patient follow-up data. The date of surgery was considered to be the starting point for all endpoint events. Detailed label distributions for the five tasks are presented in Table 2.

All endpoints exhibited marked class imbalance. To address this issue, MLSMOTE was applied to the training set, after which the degree of class imbalance was reassessed. Fig. 2 depicts the label distributions for each task before and after MLSMOTE application, showing improvements in the proportion between common and less frequent labels. Additionally, after applying MLSMOTE, the maximum imbalance ratio (MaxIR) and mean imbalance ratio (MeanIR) decreased by 0.39 and 0.82, respectively, demonstrating a substantial mitigation of class imbalance (Supplementary Table 2). Model performance before and after MLSMOTE was subsequently compared. One set of results was obtained using the original, non-resampled data, while the other was based on MLSMOTE-augmented data. Performance was evaluated using Macro-F1 (MacroFM) and Micro-F1 (MicroFM) (Supplementary Table 3). In the internal test set, MacroFM increased from 0.03 to 0.30 and MicroFM from 0.04 to 0.31. In the external validation cohort, both MacroFM and MicroFM increased from 0.01 to 0.16. These consistent gains demonstrate that MLSMOTE effectively mitigates class imbalance and substantially enhances the model’s generalization capability.

Table 2 Label statistical analysis of five tasks: BCSS, iDFS, D-DFS, L-RFS, and OS.

Task	Training (n = 5965)	Internal test (n = 1492)	External validation (n = 1125)
BCSS, n(%)			
Death	330(5.53)	83(5.56)	22(1.96)
Survival	5635(94.47)	1409(94.44)	1103(98.04)

iDFS, $n(\%)$			
Recurrence	669(11.22)	167(11.19)	79(7.02)
No recurrence	5296(88.78)	1325(88.81)	1046(92.98)
D-DFS, $n(\%)$			
Recurrence	621(10.41)	155(10.39)	58(5.16)
No recurrence	5344(89.59)	1337(89.61)	1067(94.84)
L-RFS, $n(\%)$			
Recurrence	438(7.34)	109(7.31)	59(5.24)
No recurrence	5527(92.66)	1383(92.69)	1066(94.76)
OS, $n(\%)$			
Death	404(6.77)	101(6.77)	31(2.76)
Survival	5561(93.23)	1391(93.23)	1094(97.24)

n : Number of patients.

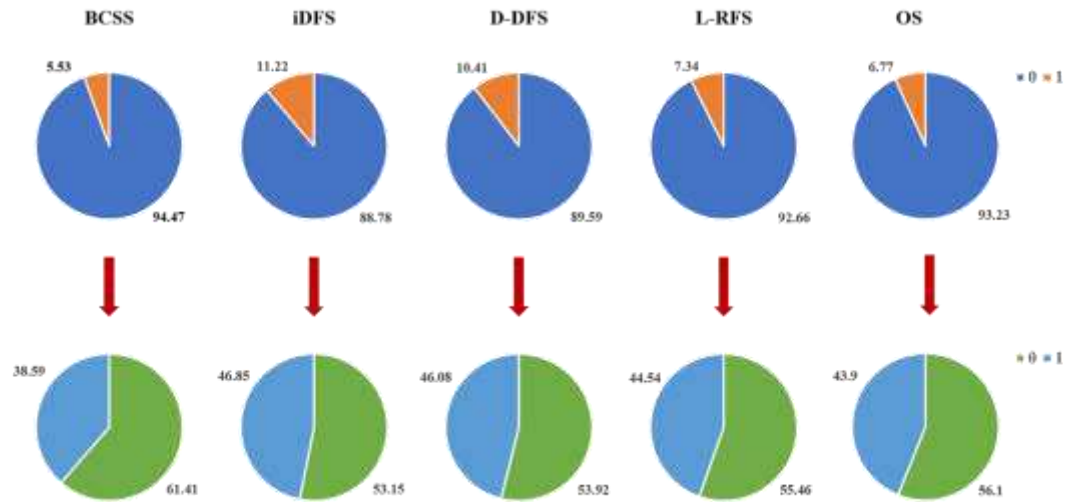


Fig. 2: Label distribution diagram before and after applying MLSMOTE.

The upper half shows the original label distribution for the five tasks, while the lower half shows the label distribution for the five tasks after applying MLSMOTE.

EMR-BCEP performance on internal test set

Based on a multitask learning framework, EMR-BCEP utilizes 30 features as input, including 28 structured features and 2 unstructured features, and its performance was evaluated on the internal test set. For the five tasks—BCSS, iDFS, D-DFS, L-RFS, and OS—the multitask model achieved area under the receiver operating characteristic

curve (AUROC) values of 0.76, 0.72, 0.72, 0.75, and 0.76, respectively. Compared to the single-task models, which yielded AUROCs of 0.64, 0.66, 0.68, 0.70, and 0.69, this corresponds to improvements of 0.12, 0.06, 0.04, 0.05, and 0.07, demonstrating strong discriminative ability. AUROC curves for five tasks are shown in Fig. 3. Additionally, the multitask MacroFM and MicroFM scores were 0.30 and 0.31, substantially higher than the single-task scores of 0.23 and 0.24, indicating superior classification results for imbalanced multilabel data (Supplementary Table 4). Accuracy metrics for the model are reported in Supplementary Table 5.

Performance stratification of EMR-BCEP on the internal test set was conducted by age (Fig. 4a) and molecular subtype (Fig. 4b). For age stratification, within the primary clinical population aged 31-70 years, EMR-BCEP consistently outperformed the single-task model across all five tasks, demonstrating more stable and uniform predictive advantages. In contrast, in the two extreme age groups with relatively smaller sample sizes (21-30 and 71-80 years), the performance of EMR-BCEP showed greater variability, with inconsistent performance differences across tasks and reduced advantages or comparable performance to the single-task model in some endpoints. For molecular subtype stratification, across the three major molecular subtypes—Luminal A, Luminal B (HER2 negative), and Luminal B (HER2 positive)—EMR-BCEP achieved markedly higher AUROC values than the single-task model across all five tasks, representing the most prominent predictive advantage. For the Triple-negative and Others subtypes, EMR-BCEP maintained superior performance in most tasks, with relatively weaker results observed only for iDFS in the Triple-negative subtype and D-DFS in the Others subtype, while overall differences remained limited.

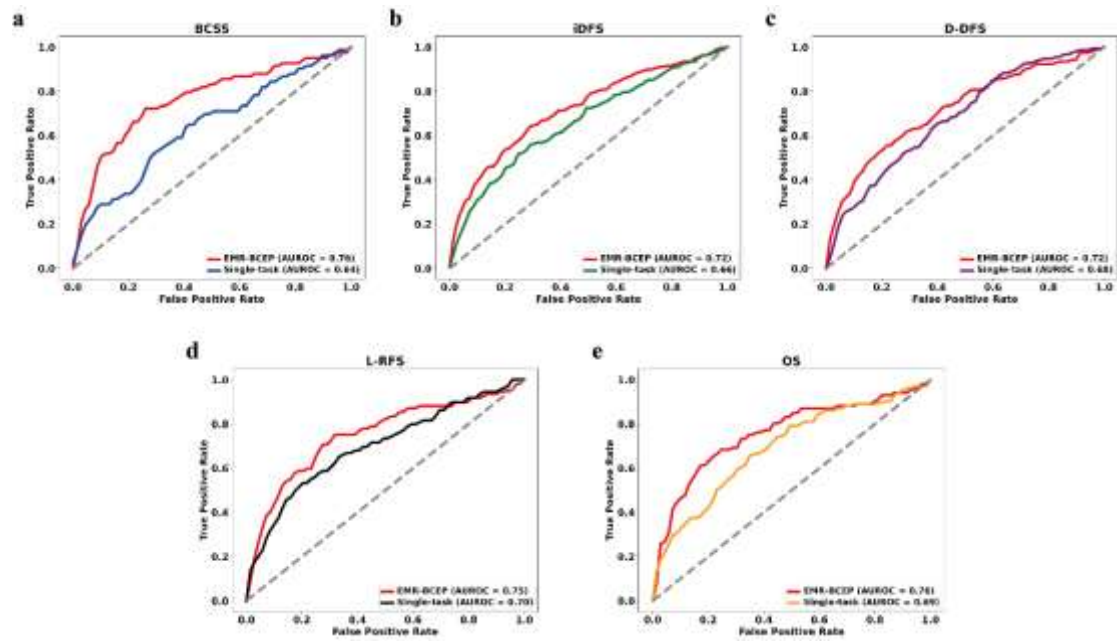


Fig. 3: Receiver operating characteristic (ROC) curves for the EMR-BCEP model and single-task model on internal test set.

a ROC curve for BCSS. **b** ROC curve for iDFS. **c** ROC curve for D-DFS. **d** ROC curve for L-RFS. **e** ROC curve for OS.

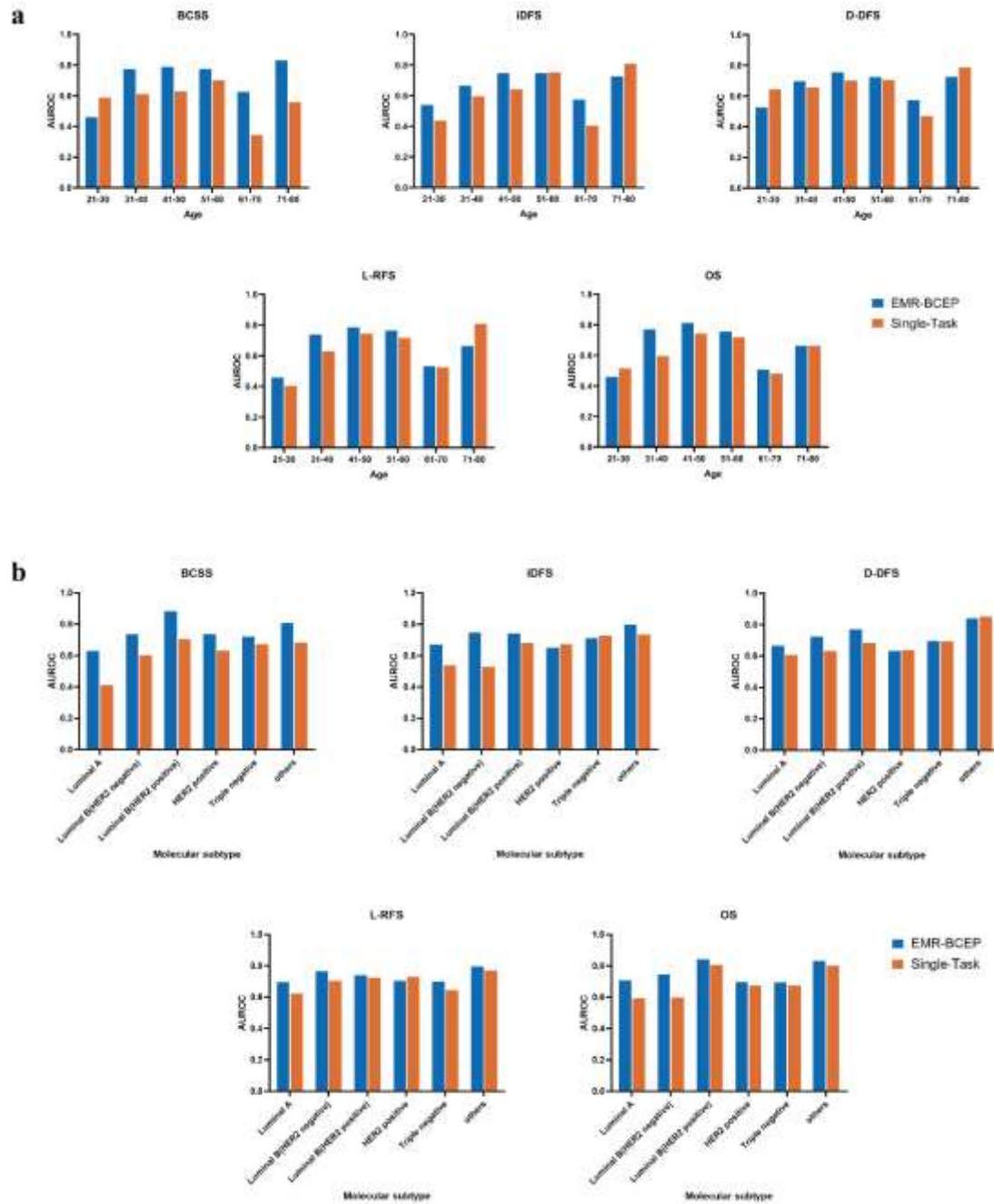


Fig. 4: Comparison of age stratification and molecular subtype stratification between the EMR-BCEP model and the single-task model on internal test set.

a Age-stratified comparison results between EMR-BCEP and single-task models for BCSS, iDFS, D-DFS, L-RFS, and OS. **b** Molecular subtype-stratified comparison results between EMR-BCEP and single-task models for BCSS, iDFS, D-DFS, L-RFS, and OS.

External validation

In the external validation cohort, EMR-BCEP continued to demonstrate robust predictive performance. For the five tasks—BCSS, iDFS, D-DFS, L-RFS, and OS—the multitask model achieved AUROC values of 0.76, 0.62, 0.66, 0.63, and 0.69, respectively. Compared to single-task models, improvements of 0.05, 0.02, 0.01, and 0.04 were observed for BCSS, iDFS, L-RFS, and OS, respectively, whereas performance on D-DFS was maintained without deterioration, indicating strong cross-center generalizability (Supplementary Fig. 1). The multitask model also achieved MacroFM and MicroFM scores of 0.16, both higher than those of the single-task models (Supplementary Table 6). Overall, these results indicate that EMR-BCEP maintains stable and superior classification performance on an independent dataset. We further conducted age-stratified and molecular subtype-stratified analyses in the external validation cohort, and the overall trends were consistent with those observed in the internal test set (Supplementary Fig. 2).

Comparison with machine learning model

We further compared the performance of EMR-BCEP with that of the traditional machine learning model, extreme gradient boosting (XGBoost)²⁸. Since XGBoost cannot directly model unstructured text data, only the same 28 structured features were used for training and evaluation to ensure comparability. Under this structured-feature-only setting, EMR-BCEP achieved AUROCs of 0.67, 0.61, 0.64, 0.61, and 0.65 across the five tasks on the internal test set, and 0.70, 0.60, 0.62, 0.62, and 0.64 on the external validation set, respectively. Compared with the version incorporating unstructured features, overall AUROC values were reduced, indicating a measurable contribution of text information to model discrimination. For XGBoost, the AUROCs on the internal test set were 0.79, 0.75, 0.74, 0.79, and 0.79, whereas those on the external validation set decreased to 0.49, 0.51, 0.52, 0.51, and 0.50, respectively. Although XGBoost achieved higher performance on the internal test set than EMR-BCEP trained without unstructured data, this advantage did not generalize to the external validation cohort. Overall, EMR-BCEP demonstrated more stable discriminative performance across internal and external datasets under the same structured input setting.

Comparison of dynamic loss weight optimization methods (DLWO)

To highlight the impact of dynamic loss weight optimization (DLWO), we performed an ablation analysis by comparing the multitask model trained with and without DLWO, while keeping the model architecture, input features, and training settings identical

across all five prognostic tasks (Fig. 5). Incorporation of DLWO led to consistent improvements in predictive performance, with AUROC increases of 0.01, 0.05, 0.05, 0.06, and 0.05 for BCSS, iDFS, D-DFS, L-RFS, and OS, respectively, compared with the model trained using fixed loss weights (without DLWO). These results demonstrate that dynamic loss weight allocation facilitates more balanced optimization across tasks and improves overall classification performance.

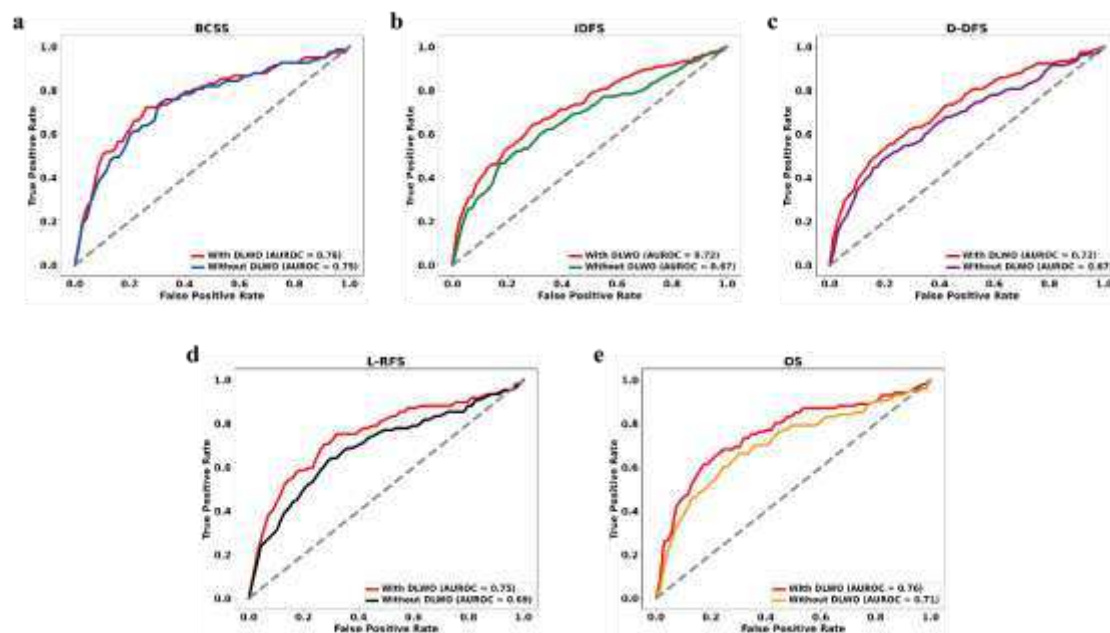


Fig. 5: ROC curves with and without dynamic loss weight optimization method.

a ROC curve for BCSS. **b** ROC curve for iDFS. **c** ROC curve for D-DFS. **d** ROC curve for L-RFS. **e** ROC curve for OS.

Online platform for clinical application

To facilitate clinical adoption and decision support, we integrated the multitask breast cancer prognostic prediction model into a web-based interactive platform. Clinical users can input the 30 required features, and the system will generate real-time individualized risk predictions for the five prognostic tasks—BCSS, iDFS, D-DFS, L-RFS, and OS. The web application also supports visual presentation of prediction results, enabling straightforward interpretation and practical use by clinicians. In addition, the model has been deployed as a mobile application, which is accessible for download via the web-based platform.

DISCUSSION

This study proposes EMR-BCEP, a multitask learning-based model for breast cancer prognostic assessment, which demonstrated robust predictive performance and cross-center generalizability in a dual-center retrospective cohort. The main contributions are as follows: (1) we developed a multitask learning framework to jointly encode structured and unstructured clinical data for predicting five interrelated breast cancer prognostic indicators (BCSS, iDFS, D-DFS, L-RFS, OS); (2) we implemented a dynamic loss-weight mechanism to balance learning across tasks, ensuring consistent performance improvements for all prognostic endpoints; (3) by combining MLSMOTE with a semantics-preserving text augmentation strategy, we mitigated multilabel imbalance and enhanced prediction of rare events; (4) the optimal model was deployed as a web-based interactive platform for real-time personalized risk prediction in clinical practice.

Prognostic assessment is an often overlooked yet critically important component of breast cancer management. EMR-BCEP employs a multitask learning framework to predict five key prognostic endpoints in breast cancer. Our results demonstrate that, compared with single-task models, the multitask approach yields consistently superior performance in the internal test cohort, improving the AUROC for BCSS, iDFS, D-DFS, L-RFS, and OS by 0.12, 0.06, 0.04, 0.05, and 0.07, respectively. It also achieved higher scores on imbalance-sensitive metrics such as MacroFM and MicroFM. This advantage arises from the high inter-task correlations among breast cancer prognostic indicators: by enabling knowledge and representation sharing across related tasks, MTL reduces redundant feature learning and improves data efficiency, offers a more holistic patient assessment, mitigates the information-loss limitations inherent to single-task models, and enhances generalization and predictive performance²⁹⁻³². Moreover, existing breast cancer prognostic studies still predominantly rely on TNM staging, molecular markers, and genomic assays. By leveraging EMR data that encompass both structured (demographic, pathological, and treatment features) and unstructured (physician narrative notes) data, our study demonstrates that accurate, multi-endpoint prognostic prediction can be generated from routinely collected EMR data. This indicates that effective prognostic modeling does not necessarily depend on costly or specialized testing. In resource-limited settings, such a model may serve as a practical tool to support clinical decision-making and optimize healthcare resource allocation, thereby helping reduce disparities between high- and low-resource environments³³⁻³⁶.

Class imbalance remains a major challenge in medical artificial intelligence. In breast cancer prognostic tasks, adverse events such as death, recurrence, or distant metastasis occur at inherently low rates, resulting in a highly skewed label distribution in real-world clinical data. Traditional training procedures tend to favor the majority class, overlooking minority-class patterns and leading to biased predictions. To address this issue, EMR-BCEP incorporates an MLSMOTE-based strategy combined with semantics-preserving text augmentation. Our results show that this approach effectively balances label distributions, improves AUROC for rare events, and enhances MacroFM and MicroFM performance, indicating better recognition of minority outcomes and greater model stability. Moreover, EMR-BCEP is compatible with advanced AI and deep learning pipelines, providing a scalable framework for multi-label medical data analysis.

In the age-stratified analysis, EMR-BCEP demonstrated stable and balanced predictive advantages in the primary clinical population aged 31–70 years, which is closely related to the sufficient sample size and relatively concentrated clinical characteristics within this age group. Through multitask learning with shared representations, the model can capture key associations from multidimensional features across tasks, resulting in superior performance compared with single-task models across multiple prognostic endpoints. For the younger (21–30 years) and older (71–80 years) groups, the model exhibited greater performance variability, likely due to limited sample sizes, increased disease heterogeneity, and uneven distribution of clinical features, which restrict the learning of shared representations and lead to more pronounced differences across tasks^{37,38}. This result suggests that optimizing model strategies for different age groups could help enhance the personalized predictive capability for breast cancer^{39,40}. In the molecular subtype-stratified analysis, EMR-BCEP outperformed single-task models in most tasks and subtypes, indicating that the multitask shared representation strategy enhances model generalizability across different biological contexts. For the three major subtypes—Luminal A, Luminal B (HER2 negative), and Luminal B (HER2 positive)—sufficient sample sizes and balanced feature distributions allowed the model to fully learn cross-task shared representations, resulting in substantial AUROC improvements across all five tasks and demonstrating the most pronounced predictive advantages. For other molecular subtypes, due to limited sample sizes and higher heterogeneity, EMR-BCEP showed slight performance decreases in certain tasks, and in some cases performed lower than single-task models. However, the differences were

minimal, indicating that multitask shared representations still provide effective predictive capabilities in these subtypes without evident performance degradation.

External validation in an independent cohort further supported the generalizability and stability of EMR-BCEP, with favorable performance across all prognostic tasks and consistent adaptability to diverse patient subgroups. Beyond demonstrating predictive performance in a separate dataset, these results suggest a certain degree of robustness to center shift and case-mix shift, which may arise from differences in institutional practice patterns, patient characteristics, treatment strategies, and outcome distributions. This is particularly important for clinical translation, as such heterogeneity is unavoidable in real-world settings. Therefore, the external validation findings support the potential applicability of EMR-BCEP beyond the development center and provide preliminary evidence for its deployment across institutions and regions.

From a clinical application perspective, EMR-BCEP offers several advantages for personalized breast cancer management. By integrating predictions across the five prognostic tasks of breast cancer, the model generates personalized survival curves that enable intuitive long-term risk assessment and facilitate more precise treatment planning. For high-risk patients, it can prompt therapy intensification (e.g., optimizing endocrine therapy, adding immunotherapy, or enhancing chemo/radiotherapy), while avoiding overtreatment in low-risk cases. It also optimizes follow-up scheduling via risk stratification, prioritizing high-risk patients and reducing unnecessary monitoring for low-risk patients, alleviating resource burden—particularly valuable in resource-limited settings. Moreover, it empowers patients to participate in informed decision-making by understanding their individualized prognosis, improving adherence and long-term health outcomes.

ML models such as XGBoost have demonstrated strong performance in structured clinical prediction tasks. In this study, although XGBoost outperformed the EMR-BCEP model on the internal test set, its AUROC scores in the external validation cohort were substantially lower. This discrepancy suggests that while tree-based ensemble methods can effectively capture dataset-specific patterns, they may overfit the training distribution and fail to maintain discriminative performance when confronted with distributional shifts. In contrast, EMR-BCEP, built upon a unified multitask learning framework, jointly optimizes multiple clinical endpoints to learn shared representations, effectively mitigating task-specific overfitting and promoting the extraction of domain-

invariant features—thereby enabling more stable predictions across heterogeneous data sources. Moreover, the integration of unstructured clinical text enhances the model’s ability to encode semantic and contextual information that cannot be captured by structured variables alone. Traditional machine learning models, however, are inherently limited in this regard, as they cannot directly model unstructured clinical narratives—hindering their capacity to leverage the rich contextual and semantic information embedded within EMR data.

The main limitations of our study concern both the number of data centers and the characteristics of the dataset. First, our analyses relied on data from only two institutions—West China Hospital of Sichuan University and Xijing Hospital—which may inevitably introduce regional and institutional biases. We recognize that breast cancer prognosis is shaped not only by biological heterogeneity, including molecular subtypes and clinical stages, but also by variations in treatment strategies, healthcare resources, and patient demographics. A dual-center cohort is unlikely to fully capture this complexity, and therefore the generalizability of our findings to broader populations remains limited. Second, because our dataset was retrospectively derived from EMR data, it inevitably suffers from missing values, inconsistent documentation, and underrepresentation of rare subtypes. These issues may affect the robustness of our model and its potential clinical utility. To overcome these limitations, future work should aim to incorporate larger, prospective, multicenter cohorts that better reflect the diversity of real-world clinical settings. We believe that only through such broader validation can prognostic models be ensured to be not only technically feasible but also genuinely valuable in clinical practice.

In conclusion, EMR-BCEP provides accurate predictions for breast cancer prognosis, including BCSS, iDFS, D-DFS, L-RFS and OS, with potential to guide individualized treatment decisions and resource allocation. MTL enhances the use of deep learning (DL) on EMR data for prognostic assessment. While performance varies across breast cancer subtypes, the model’s ability to integrate multimodal data improves prognostic reliability. Future research should include diverse populations and disease stages, and incorporate prospective clinical data to further validate applicability and accelerate clinical implementation.

METHODS

Study design and dataset

This study is a retrospective, dual-center investigation utilizing data from the breast cancer databases of West China Hospital, Sichuan University, and Xijing Hospital. We retrospectively retrieved EMR data of patients diagnosed and treated for breast cancer at West China Hospital, Sichuan University between January 2010 and December 2019, and at Xijing Hospital between January 2013 and December 2018. Collected clinical information included patient demographics, medical history and diagnosis, pathology results, surgical records, chemotherapy and radiotherapy. All participants provided informed consent, and all data were de-identified to protect patient privacy.

The objective of this study was to develop and evaluate EMR-BCEP. During model development, a total of 13,545 breast cancer treatment records from 10,727 patients at West China Hospital, Sichuan University were initially included. Records related to preoperative neoadjuvant therapy and recurrence/metastasis salvage treatment (3,647 records) were excluded first, followed by exclusion of 2,441 records with less than five years of follow-up, resulting in a final cohort of 7,457 patients with corresponding postoperative adjuvant treatment records, with only one treatment record included for each patient. The dataset was split into training and testing sets at an 8:2 ratio by patient identifiers, ensuring independence between the two groups. Specifically, 5,965 records were used for model training and 1,492 for testing (Fig. 6). Breast cancer treatment records of 1,125 patients from Xijing Hospital were used as an external validation cohort for model evaluation.

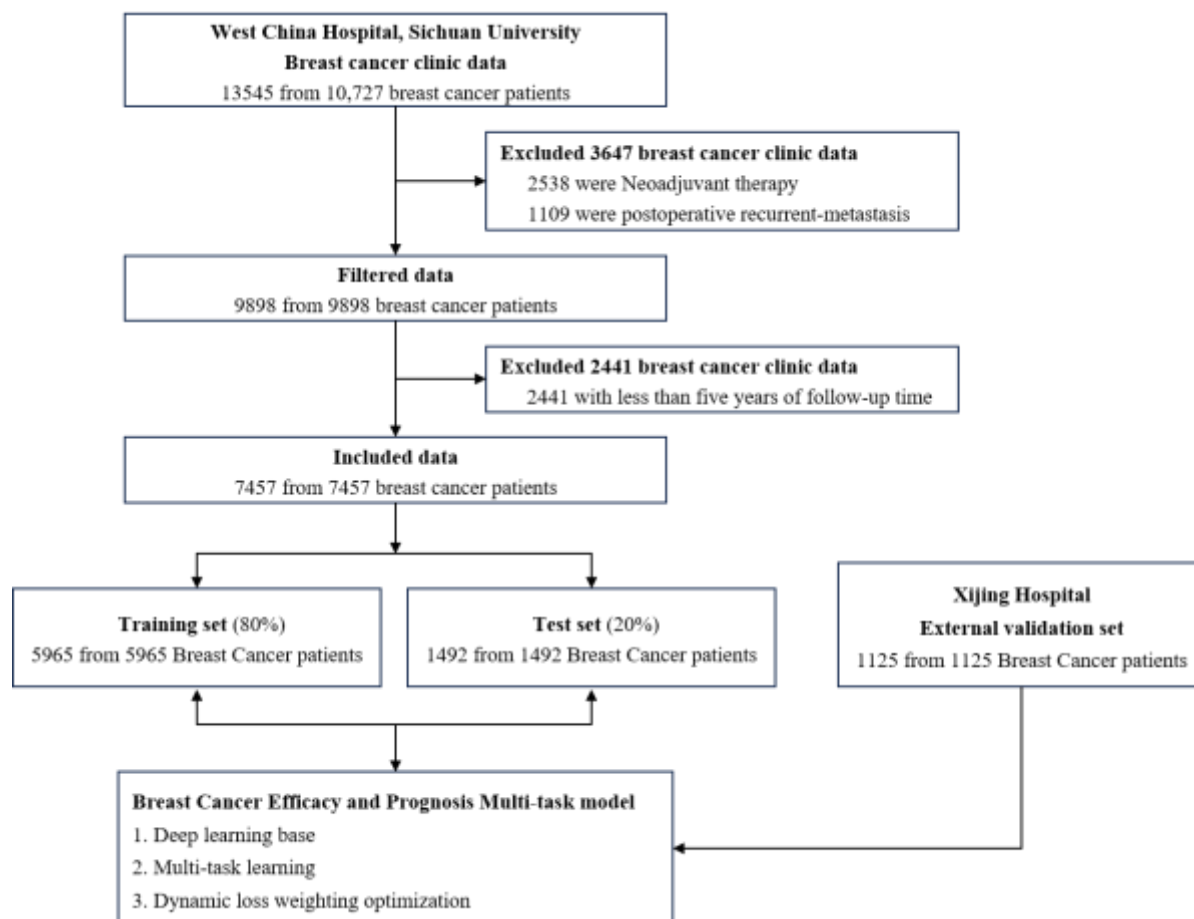


Fig. 6: Summary of clinical datasets used for development, testing, and external validation of EMR-BCEP.

To develop and evaluate EMR-BCEP, this study included 7,457 postoperative adjuvant treatment phase medical treatment data from 7,457 patients for model training and testing, and included 1,125 breast cancer medical treatment data from 1,125 patients for external validation.

Data preprocessing

In this study, all clinical breast cancer treatment data were exclusively from patients diagnosed with breast cancer, healthy individuals were not included. All patients underwent pathological confirmation and received one or more treatment modalities, including chemotherapy, radiotherapy, endocrine therapy, and targeted therapy.

To ensure high-quality data, we performed data preprocessing. First, all treatment records were aligned by matching patient identifiers with corresponding diagnosis and treatment dates, and outliers were removed by setting acceptable value ranges and formats. Next, unstructured clinical text data were preserved in their original form for subsequent model input. Structured data were divided into continuous and categorical variables. Continuous variables, such as age and body surface area, were normalized to

the [0, 1] range using min-max scaling, while categorical variables were processed using label encoding, assigning each category as an integer value starting from 0 and increasing sequentially. Finally, missing values were imputed with -1 as a dedicated sentinel value to explicitly indicate missingness. This fixed-value imputation strategy is a simple and practical preprocessing approach in machine learning, and it allowed the model to distinguish missing entries from valid observed values. The dataset was divided into training and internal test sets at an 8:2 ratio at the patient level. To prevent information leakage, all preprocessing parameters were obtained from the training set only, and the same transformation rules were subsequently applied to the internal test set and the external validation cohort.

Definition of breast cancer prognostic endpoints

The complex task of evaluating breast cancer prognostic assessment can be decomposed into multiple related subproblems, which are interconnected through shared factors or representations. Treating these subproblems independently as separate single-task models overlooks the rich interrelations among them, potentially compromising evaluation performance. In this study, we selected five highly correlated tasks: BCSS, iDFS, D-DFS, L-RFS, and OS. The criteria for each endpoint were based on the guidelines reported by Gourgou-Bourgade et al. in the breast cancer randomized clinical trial (RCT) framework⁴¹.

BCSS reflects mortality specifically attributable to breast cancer, serving as a measure of breast cancer-specific prognosis. iDFS indicates the occurrence of invasive recurrence, distant metastasis, or death, representing an important metric for assessing post-treatment disease control. D-DFS measures the incidence of distant metastasis or death, evaluating systemic tumor control. L-RFS denotes whether local or regional recurrence occurred, a key indicator of local treatment efficacy. OS represents overall patient mortality regardless of cause and is a core endpoint for assessing overall survival and treatment benefit. All endpoint events were defined within a 5-year follow-up period, and only events occurring within 5 years after surgery were considered in this study.

Selection of EMR-derived predictive features for model development

A total of 99 clinical breast cancer features were initially collected. While a large number of features can provide the model with sufficient information for improved

prediction performance, high-dimensional data in practical applications may introduce redundancy or noise, potentially degrading predictive accuracy. Furthermore, collecting multi-dimensional features can be challenging.

In this study, a stepwise feature selection approach was employed. First, feature inclusion was guided by clinically validated multidimensional prognostic models reported in prior large-scale breast cancer studies^{42,43}. Based on these established feature systems and considering data availability within our cohort, 30 core clinical features with demonstrated prognostic relevance were selected for subsequent analysis. Second, correlation analysis was performed on 28 numerical features. Only three pairs of features exhibited correlations exceeding 80%, therefore no features were removed. Additionally, two unstructured features were retained.

Data synthesis using the MLSMOTE algorithm

In this study, to address the prevalent label imbalance across the five breast cancer prognostic tasks (BCSS, iDFS, D-DFS, L-RFS, and OS), we adopted a multi-label data augmentation strategy tailored for joint modeling of structured and unstructured features. This approach is based on MLSMOTE combined with a neighbor-based sampling method for textual features, aiming to generate augmented training data that preserves both label correlations and semantic consistency. We constructed the feature space using 28 structured features and synthesized samples based on MLSMOTE to effectively increase the proportion of positive classes in the multi-label training set. For the two unstructured textual features, given the non-differentiable nature of text interpolation, we designed a semantic-preserving strategy based on sample neighborhood. Specifically, the textual fields were directly inherited from nearest-neighbor cases instead of being artificially fabricated, thereby avoiding semantic noise or invalid text generation while preserving authentic medical language representations within the real clinical context. The number of synthetic samples was determined according to the imbalance ratio of each prognostic task. Specifically, minority class samples were oversampled until the class imbalance ratio (IR) was reduced below 1.5, rather than forcing the class distribution to become fully balanced. This strategy aimed to mitigate class imbalance while minimizing the risk of introducing excessive synthetic samples. To determine the optimal augmentation scale, synthetic samples were generated at five preset levels (3,000, 3,500, 4,000, 4,500, and 5,000), and model performance was evaluated under each setting. The final synthetic sample size was then

chosen according to the configuration that achieved the best overall performance. To maintain methodological rigor and comparability across experimental settings, the MLSMOTE-based augmentation procedure was consistently applied to both multitask and single-task training datasets.

Construction of Early Breast Cancer Prognostic Assessment models

The multitask model for evaluating the prognosis of early breast cancer consists of three components: a shared encoder module, a multitask decoder module, and a dynamic loss weight optimization module. The details of each component are illustrated in Fig.7. The shared encoder module comprises a structured data encoder and an unstructured data encoder. It takes breast cancer diagnosis and treatment data as input, extracts features from both structured and unstructured data separately, and integrates them. The multitask decoder module then uses task-specific prediction heads for each endpoint to decode the integrated features and generate task-specific predictions. Finally, to optimize the training process in multitask learning, we introduce a dynamic loss optimization module, where task-specific loss weights are adaptively adjusted throughout training. In the following sections, we will provide a detailed description of each component.

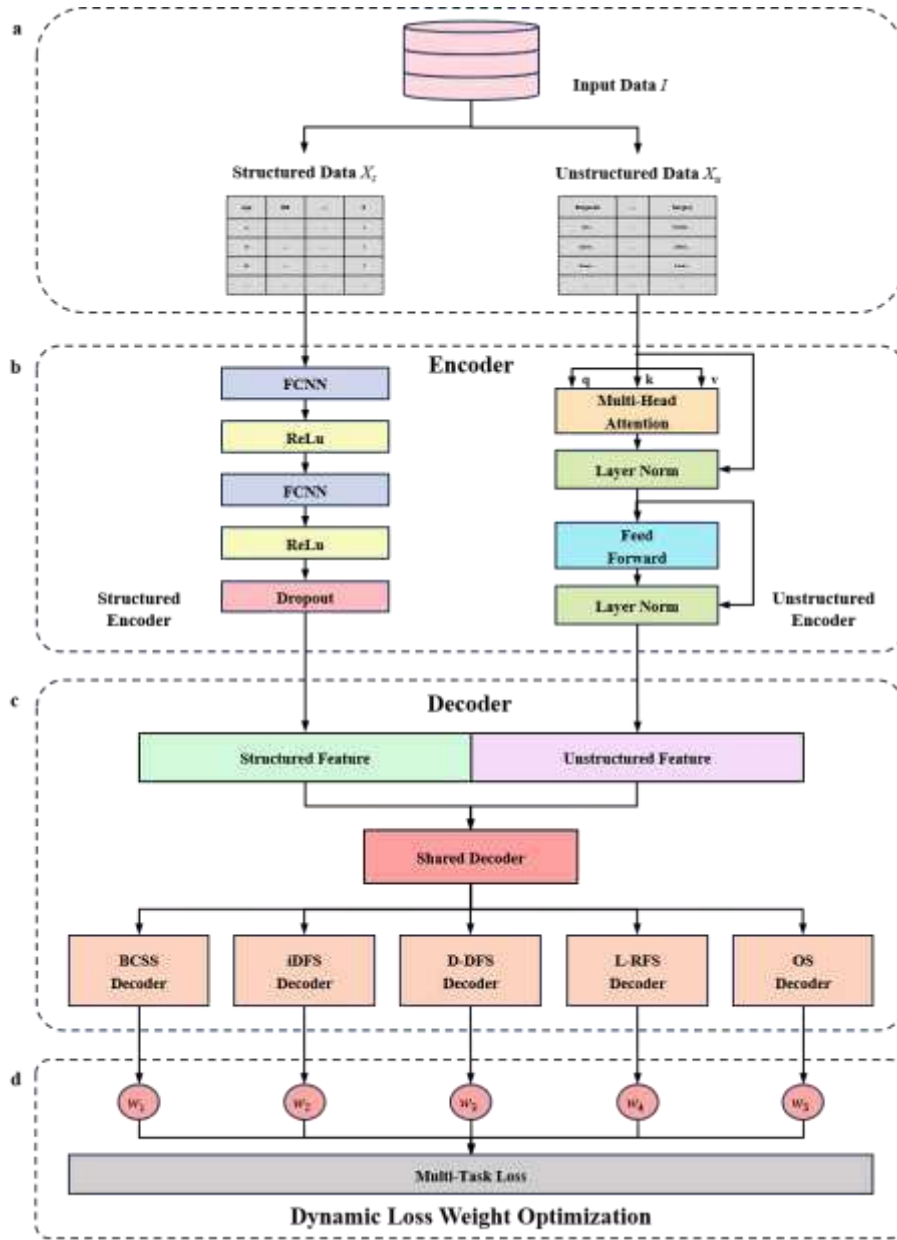


Fig. 7: EMR-BCEP model architecture.

a Input module of the EMR-BCEP model. X_s represents structured data, and X_u represents unstructured data. **b** Shared encoder module. It consists of two parts: a structured encoder and an unstructured encoder. **c** Multitask decoder module. Different tasks use specific decoders to decode the fused features. **d** Dynamic loss weight optimization module. w_i represents the weights of different tasks.

Shared encoder module

The shared encoder module is designed to extract features from both structured and unstructured data and integrate them into a unified representation space. Structured data are encoded using a fully connected neural network (FCNN), which is well suited to

high-dimensional inputs and supports nonlinear classification and regression across large feature spaces. The model takes structured clinical data of breast cancer patients, X_s , as input and outputs the structured feature representation, h_s . The unstructured data encoder is based on a multi-head self-attention mechanism (Multi-Head Self-Attention, MHSA). It takes unstructured data, X_u , as input and outputs the unstructured feature representation, h_u . Before being processed by the MHSA module, the unstructured data is first transformed into a vector representation of length 256 through an embedding layer, ensuring that the model can effectively learn semantic features in a high-dimensional space. Next, deep feature modeling is performed using six self-attention Transformer layers, each containing eight independent attention heads. Each attention head captures global dependencies in different subspaces of the text. After multi-head attention computation, the extracted features are concatenated and passed through a linear transformation to integrate information, forming a comprehensive contextual representation. Following each MHSA layer, a feedforward network (FFN) is applied to further enhance feature representation capability. The FFN consists of two fully connected layers, with a hidden dimension of 2048. Overall, the model alternates between six layers of MHSA and FFN, effectively capturing deep semantic information from the data to provide robust feature representations for downstream tasks. Finally, the structured features h_s and unstructured features h_u are concatenated to form the fused feature representation H , which is used for downstream task decoding:

$$H = h_s + h_u \quad (1)$$

Multitask decoder module

During the multitask decoding phase, all tasks share a common base decoder module (Fig. 7), which is built on a convolutional neural network (CNN). The module employs convolutional layers with a kernel size of 3 to capture local feature patterns, followed by pooling operations to progressively extract higher-level representations, and ReLU activation for nonlinear transformation. In this study, task type was found to have a substantial impact on model generalizability and performance. To accommodate the specific requirements of different tasks, the model adopts task-specific fully connected neural networks (FCNNs) in the hidden layers. The outputs are mapped through a Sigmoid activation function to yield class probabilities. This design allows each task to benefit from shared feature representations while preserving task-specific structures

and prediction objectives, thereby enabling efficient modeling and prediction across multiple prognostic endpoints in breast cancer.

Dynamic loss weighting module

Multitask learning improves performance by sharing representation information, leading to better results compared to single-task training. It helps alleviate model overfitting and enhances generalization ability. However, the scale of losses for different tasks may vary, causing one task to dominate the overall loss, while the other tasks fail to influence the learning process of the shared layers. To address this issue, we employed a dynamic loss weight optimization strategy^{44,45}. This strategy automatically adjusts the weight of each task based on its learning progress, overcoming the limitations of traditional static weight allocation. First, for each task, we define a learnable uncertainty parameter σ , and use the softplus function to ensure it remains positive. This ensures stable convergence for the weighted sum of all task losses. The weight w for each task is then computed as follows:

$$w = \log(1 + \sigma^2) \quad (2)$$

Next, we propose a dynamic weight optimization method based on task weight normalization, calculated as follows:

$$L = \frac{1}{2 * w^2} * loss + w \quad (3)$$

Finally, the individual task losses, including the dynamic weights and regularization adjustments, are summed to obtain the final loss:

$$L_{total} = \sum_{i=1}^n L_i \quad (4)$$

Model evaluation

In this study, the data used for training and internal test the multitask model for early breast cancer prognostic assessment was randomly split. To ensure fairness in comparison, both the multitask model and single-task models were trained and internal tested on the same training and internal test set. All models were implemented using PyTorch 2.4.1 on an Ubuntu 24.04 LTS system. The host was equipped with an Intel(R) Xeon(R) Gold 6226R CPU and an NVIDIA RTX A6000. During training, the batch sizes were set to 128 with a total of 100 epochs.

MeanIR and MaxIR are used to estimate the imbalance rate in a dataset, as defined in the appendix. We employed macro-average and micro-average F1 scores (MacroFM and MicroFM) to analyze the performance of the model on the original dataset versus the MLSMOTE-augmented dataset, MacroFM equally weights performance across tasks, MicroFM emphasizes overall performance by aggregating samples across classes. For given TP, FP, FN, and TN denote the number of true positives, false positives, false negatives, and true negatives for a single task, precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

The F1 score is calculated as:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

MacroFM and MicroFM are obtained by averaging the F1 scores across L tasks:

$$MacroFM = \frac{1}{|L|} \sum_{i=1}^{|L|} F1(TP_i, FP_i, TN_i, FN_i) \quad (8)$$

$$MicroFM = F1\left(\sum_{i=1}^{|L|} TP_i, \sum_{i=1}^{|L|} FP_i, \sum_{i=1}^{|L|} TN_i, \sum_{i=1}^{|L|} FN_i\right) \quad (9)$$

Model performance was evaluated using AUROC, ROC curves, accuracy, MacroFM, and MicroFM.

Statistical analysis

In this study, all statistical analyses were performed using Python 3.8. For continuous variables, we calculated the mean and standard deviation (SD). For discrete variables, we recorded their frequency of occurrence.

AUROC was used as the primary performance metric. Each experiment was performed five times with different random seeds while keeping the same data split and parameter settings.

Ethics and informed consent

This study was approved by the Clinical Test and Biomedical Ethics Committee of West China Hospital, Sichuan University (reference number 2020 (427)) on 14 May, 2020. All study participants signed consent forms electronically.

Data availability

Limited deidentified data used for the analysis presented in this work can be made available to qualified researchers upon request. Please contact Dr. Xiaowei Mao, the corresponding author, via email at xiaowei.mao@uestc.edu.cn.

Code availability

The code for EMR-BCEP is available at https://github.com/MaoLab-CD/Breast_cancer.

REFERENCES

- 1 Nolan, E., Lindeman, G. J. & Visvader, J. E. Deciphering breast cancer: from biology to the clinic. *Cell* **186**, 1708-1728 (2023).
<https://doi.org/10.1016/j.cell.2023.01.040>
- 2 Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* **72**, 7-33 (2022).
<https://doi.org/10.3322/caac.21708>
- 3 Kim, J. *et al.* Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nat Med* **31**, 1154-1162 (2025).
<https://doi.org/10.1038/s41591-025-03502-3>
- 4 Allemani, C. *et al.* Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* **385**, 977-1010 (2015).
[https://doi.org/10.1016/s0140-6736\(14\)62038-9](https://doi.org/10.1016/s0140-6736(14)62038-9)
- 5 Autier, P. *et al.* Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *Bmj* **341**, c3620 (2010). <https://doi.org/10.1136/bmj.c3620>
- 6 Runowicz, C. D. *et al.* American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. *Journal of Clinical Oncology* **34**, 611-635 (2016). <https://doi.org/10.1200/jco.2015.64.3809>
- 7 Eisemann, N. *et al.* Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nature Medicine* **31**, 917-924 (2025). <https://doi.org/10.1038/s41591-024-03408-6>
- 8 Dembrower, K. *et al.* Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *The Lancet Digital Health* **2**, e468-e474 (2020). [https://doi.org/10.1016/S2589-7500\(20\)30185-0](https://doi.org/10.1016/S2589-7500(20)30185-0)
- 9 Freeman, K. *et al.* Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *Bmj* **374**, n1872 (2021). <https://doi.org/10.1136/bmj.n1872>
- 10 Boeri, C. *et al.* Machine Learning techniques in breast cancer prognosis prediction: A primary evaluation. *Cancer Med* **9**, 3234-3243 (2020).
<https://doi.org/10.1002/cam4.2811>
- 11 Song, X. *et al.* Prognostic prediction of breast cancer patients using machine learning models: a retrospective analysis. *Gland Surg* **13**, 1575-1587 (2024).
<https://doi.org/10.21037/gs-24-106>
- 12 Teng, L. *et al.* A novel nomogram and survival analysis for different lymph node status in breast cancer based on the SEER database. *Breast Cancer* **31**, 769-786 (2024). <https://doi.org/10.1007/s12282-024-01591-5>
- 13 Sun, Y., Goodison, S., Li, J., Liu, L. & Farmerie, W. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics* **23**, 30-37 (2006). <https://doi.org/10.1093/bioinformatics/btl543>
- 14 Altman, D. G. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest* **27**, 235-243 (2009).

- <https://doi.org/10.1080/07357900802572110>
- 15 Clark, G. M. Do we really need prognostic factors for breast cancer? *Breast Cancer Res Treat* **30**, 117-126 (1994). <https://doi.org/10.1007/bf00666054>
 - 16 Ma, Q. *et al.* Multitask deep learning model based on multimodal data for predicting prognosis of rectal cancer: a multicenter retrospective study. *BMC Medical Informatics and Decision Making* **25**, 209 (2025). <https://doi.org/10.1186/s12911-025-03050-3>
 - 17 Priyadarshi, S. *et al.* OncoMark: a high-throughput neural multi-task learning framework for comprehensive cancer hallmark quantification. *Communications Biology* **8**, 1434 (2025). <https://doi.org/10.1038/s42003-025-08727-z>
 - 18 Lin, T. W., Arriagada, S. O. & Lin, C. in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 1-5.
 - 19 Fotouhi, S., Asadi, S. & Kattan, M. W. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *Journal of Biomedical Informatics* **90**, 103089 (2019). <https://doi.org/10.1016/j.jbi.2018.12.003>
 - 20 Fernando, C. D., Weerasinghe, P. T. & Walgampaya, C. K. Heart Disease Risk Identification using Machine Learning Techniques for a Highly Imbalanced Dataset: a Comparative Study. *KDU Journal of Multidisciplinary Studies* (2022). <https://doi.org/10.4038/kjms.v4i2.50>
 - 21 Alamsyah, A. R. B., Anisa, S. R., Belinda, N. S. & Setiawan, A. SMOTE and Nearmiss Methods for Disease Classification with Unbalanced Data : Case Study: IFLS 5. *Proceedings of The International Conference on Data Science and Official Statistics* **2021**, 305-314 (2022). <https://doi.org/10.34123/icdsos.v2021i1.240>
 - 22 Heymans, M. W. & Twisk, J. W. R. Handling missing data in clinical research. *Journal of Clinical Epidemiology* **151**, 185-188 (2022). <https://doi.org/10.1016/j.jclinepi.2022.08.016>
 - 23 Sun, Y., Wong, A. K. C. & Kamel, M. S. CLASSIFICATION OF IMBALANCED DATA: A REVIEW. *International Journal of Pattern Recognition and Artificial Intelligence* **23**, 687-719 (2009). <https://doi.org/10.1142/S0218001409007326>
 - 24 Krawczyk, B. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence* **5**, 221-232 (2016). <https://doi.org/10.1007/s13748-016-0094-0>
 - 25 Charte, F., Rivera, A. J., del Jesus, M. J. & Herrera, F. MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation. *Knowledge-Based Systems* **89**, 385-397 (2015). <https://doi.org/10.1016/j.knosys.2015.07.019>
 - 26 Das, P., Thakran, Y., Anal, S. R. N., Pal, V. & Yadav, A. BRMCF: Binary Relevance and MLSMOTE Based Computational Framework to Predict Drug Functions From Chemical and Biological Properties of Drugs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **20**, 1761-1773

- (2023). <https://doi.org/10.1109/TCBB.2022.3215645>
- 27 Lucky, H., Roslynia & Suhartono, D. in *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*. 346-350.
- 28 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, San Francisco, California, USA, 2016).
- 29 Caruana, R. Multitask Learning. *Machine Learning* **28**, 41-75 (1997). <https://doi.org/10.1023/A:1007379606734>
- 30 Amyar, A., Modzelewski, R., Li, H. & Ruan, S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine* **126**, 104037 (2020). <https://doi.org/10.1016/j.compbiomed.2020.104037>
- 31 Lin, C.-H. *et al.* A multitask deep learning model utilizing electrocardiograms for major cardiovascular adverse events prediction. *npj Digital Medicine* **8**, 1 (2025). <https://doi.org/10.1038/s41746-024-01410-3>
- 32 Wang, B. *et al.* in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 930-935.
- 33 Siotos, C. *et al.* Tumor Site and Breast Cancer Prognosis. *Clinical Breast Cancer* **18**, e1045-e1052 (2018). <https://doi.org/10.1016/j.clbc.2018.05.007>
- 34 Aavula, R. & Bhramaramba, R. in *ICCCE 2020*. (eds Amit Kumar & Stefan Mozar) 1517-1533 (Springer Nature Singapore).
- 35 Finek, J. *et al.* The importance of prognostic factors in premenopausal women with breast cancer. *Anticancer Res* **27**, 1893-1896 (2007).
- 36 Stickeler, E. Prognostic and Predictive Markers for Treatment Decisions in Early Breast Cancer. *Breast Care (Basel)* **6**, 193-198 (2011). <https://doi.org/10.1159/000329471>
- 37 Sarfati, D., Koczwara, B. & Jackson, C. The impact of comorbidity on cancer and its treatment. *CA Cancer J Clin* **66**, 337-350 (2016). <https://doi.org/10.3322/caac.21342>
- 38 Dialla, P. O. *et al.* In breast cancer, are treatments and survival the same whatever a patient's age? A population-based study over the period 1998-2009. *Geriatr Gerontol Int* **15**, 617-626 (2015). <https://doi.org/10.1111/ggi.12327>
- 39 Kang, H. Y. J., Ko, M. & Ryu, K. S. Prediction model for survival of younger patients with breast cancer using the breast cancer public staging database. *Scientific Reports* **14**, 25723 (2024). <https://doi.org/10.1038/s41598-024-76331-y>
- 40 Gao, X., Zhang, F., Zhou, Q., Xu, H. & Bian, J. Metastasis, characteristic, and treatment of breast cancer in young women and older women: A study from the Surveillance, Epidemiology, and End Results registration database. *PLoS One* **18**, e0293830 (2023). <https://doi.org/10.1371/journal.pone.0293830>
- 41 Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Annals of Oncology* **26**, 2505-2506 (2015).

<https://doi.org/10.1093/annonc/mdv478>

- 42 Zhong, X. *et al.* Multidimensional Machine Learning Personalized Prognostic Model in an Early Invasive Breast Cancer Population-Based Cohort in China: Algorithm Validation Study. *JMIR Medical Informatics* **8** (2020).
<https://doi.org/10.2196/19069>
- 43 Sun, D., Wang, M. & Li, A. A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **16**, 841-850 (2019). <https://doi.org/10.1109/TCBB.2018.2806438>
- 44 Cipolla, R., Gal, Y. & Kendall, A. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7482-7491.
- 45 Chen, Z., Badrinarayanan, V., Lee, C.-Y. & Rabinovich, A. in *Proceedings of the 35th International Conference on Machine Learning* Vol. 80 (eds Dy Jennifer & Krause Andreas) 794--803 (PMLR, Proceedings of Machine Learning Research, 2018).

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2023ZD0502300 [2023ZD0502303]), National Natural Science Foundation of China (T2322002), the Sichuan Science and Technology Program (2024NSFTD0032, 2025ZNSFSC0003 and 2025ZDZX0012), the National Science and Technology Major Project (2024ZD0523900). We acknowledge computational resources provided by the High-Performance Computing Center at the University of Electronic Science and Technology of China (UESTC) for model training and validation.

Author information

These authors contributed equally: Ao Fan, Rui Yuan, Dan Zheng, Jingjing Xiao.

Authors and Affiliations

Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610072, China

Ao Fan, Rui Yuan, Jiaheng Feng, Zhongni Ye, Chenyu Xiong & Xiaowei Mao

Institute for Breast Health Medicine, Cancer Center, Breast Disease Center, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu, Sichuan, 610041, China

Dan Zheng, Ting Luo & Xiaorong Zhong

Sichuan Provincial Engineering Research Center of Intelligent Diagnosis and Treatment of Breast Diseases

Dan Zheng, Ting Luo & Xiaorong Zhong

Department of Thyroid, Breast, and Vascular Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi Province, 710032, P.R. China

Jingjing Xiao, Tingyu Qu, Junpeng Ren & Jiajun Ding

Tianfu Jincheng Laboratory, Chengdu, 610093, China

Xiaowei Mao

Contributions

A.F. and X.M. conceived and designed the study. D.Z., J.X., T.Q., T.L., J.D. and X.Z. contributed to data acquisition. A.F., J.X., T.L. and X.Z. performed data curation and preprocessing. A.F., J.F., Z.Y. and C.X. developed the computational code. A.F. and R.Y. developed the online platform. A.F. wrote the original draft. J.D., X.Z. and X.M. critically reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Corresponding authors

Correspondence to Xiaowei Mao, Xiaorong Zhong or Jiajun Ding.

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)