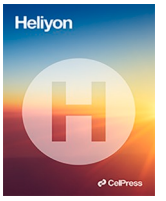
Heliyon 10 (2024) e24876

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Clinical decision support tool for breast cancer recurrence prediction using SHAP value in cooperative game theory

Ying Liu [a], Yating Fu [b], Yadong Peng [a,c], Jie Ming [d,*]

[a] Special Needs Comprehensive Department, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, Xinjiang, 830000, China
[b] Urumqi Stomatological Hospital, 196 Zhongshan Road, Tianshan District, 830002, Urumqi, China
[c] Oncology Department, The First Affiliated Hospital of Xiangnan University, Chenzhou, Hunan, 423000, China
[d] Medical Imaging Center, Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi, Xinjiang, 830000, China

ARTICLE INFO

Keywords:
Breast cancer
Cancer recurrence
SHAP
Explainable ensemble learning

ABSTRACT

Background: Recurrence remains the primary cause of death in patients with breast cancer. Although machine learning can efficiently predict the prognosis of breast cancer patients, the black-box nature of the model may result in a lack of evidence for clinicians when making critical decisions.

Methods: In this study, our main objective was twofold: (1) to develop a clinical decision support tool for predicting the prognosis of breast cancer and (2) to identify and explore the key factors that influence breast cancer recurrence. To achieve this, we employed an explainable ensemble learning method called Shapley additive explanation (SHAP), which leverages cooperative game theory. Using real-world data from 1629 breast cancer patients, we analyzed and uncovered the key factors associated with breast cancer recurrence. Subsequently, we used these identified factors to create a recurrence prediction model and establish a decision mechanism for the tool. The proposed method not only provides accurate recurrence predictions but also offers transparent explanations for these predictions.

Results: By utilizing four key factors, namely, tumor size, clinical stage III, number of lymph node metastases, and age, our decision support tool for predicting breast cancer recurrence achieved significant improvements. The extra-tree model exhibited an increased area under the receiver operating characteristic curve (AUC) of 0.97, while the Random Forest model demonstrated an improved AUC of 0.96. We also offer a decision mechanism for a recurrence prediction model based on the identified key factors. This transparent and interpretable decision-making process facilitated by our explainable ensemble learning model enhances trust and promotes its applicability in clinical settings.

Conclusions: The proposed explainable ensemble learning method shows promising results in predicting breast cancer recurrence, outperforming existing methods with high accuracy and transparency. This advancement has the potential to significantly improve clinical decision-making and patient outcomes in breast cancer treatment.

* Corresponding author.
E-mail address: mingjie1260@xjmu.edu.cn (J. Ming).

https://doi.org/10.1016/j.heliyon.2024.e24876
Received 15 August 2023; Received in revised form 15 January 2024; Accepted 16 January 2024
Available online 18 January 2024

1. Introduction

Breast cancer is a major global health threat to women [1]. In the United States alone, over 281,550 new cases were diagnosed in 2021, leading to approximately 43,600 cancer-related deaths [2]. Despite advancements in treatments and medications that have improved the 5-year relative survival rate to 90.6% [3,4], approximately 30% of breast cancer patients still experience relapse during follow-up [5]. The heterogeneity of breast cancer contributes to varied prognoses [6], and cancer recurrence remains a major concern as it not only affects patient survival but also drastically affects quality of life. The fear of cancer recurrence concerns many breast cancer patients, adversely affecting their well-being and potentially exacerbating the risk of recurrence in a vicious cycle [7–9].

Accurate early prediction of breast cancer recurrence risk is crucial to ease patient tension, enhance quality of life, and guide subsequent treatment decisions in cases of recurrence [10]. Numerous studies have identified important risk factors for breast cancer recurrence, including age, hormone receptor status, lymph node involvement, tumor size, Ki-67, menopausal status, and histological grade [11–14]. However, the complexity of individual factors makes precise prediction challenging because of the high heterogeneity among patients. Therefore, there is an urgent need to develop an effective predictive method that can assess the risk of breast cancer recurrence at the individual level to facilitate precision medicine approaches.

Since the 1990s, scientists have been working towards establishing models for breast cancer prediction, with widely used models including the Gail model proposed by Gail et al., in 1989 [15], the Tyrer-Cuzick model developed by Tyre and Cuzick in 2004 [16], and the BRCAPRO model introduced by Hall et al., in 2009 [17]. While the first two models rely on risk factors for prediction, the last model is built on gene-disease correlations. However, there are limited risk prediction models specifically designed for breast cancer recurrence.

In recent years, researchers have utilized artificial intelligence (AI) techniques to develop prediction and prognostic models for breast cancer [18–20]. In 2016, one study employed a machine learning model, specifically Tree Random Forest, and achieved a strong predictive performance with an AUC of up to 93% for the survival prediction of 900 breast cancer patients. However, subsequent research by Chen et al. proposed the deep neural network model AMND, which achieved an AUC of 87.04% for survival analysis of 1489 patients, showing a slightly reduced performance compared to the machine learning model. Another novel framework called GPDBN was introduced by Wang et al., which effectively integrates genomic data and pathological images for breast cancer prognosis prediction. Additionally, Lee et al. designed a neural network model that utilizes attention techniques to learn features of the heterogeneous tumor microenvironment, achieving an accuracy of 0.75 for breast cancer. Despite the effective prognostic capabilities of machine learning, its black-box nature may limit the transparency and evidence for clinicians when making critical decisions.

In recent years, with the rapid advancement of artificial intelligence, researchers have explored the development of predictive and prognostic models for breast cancer using AI techniques [18–21]. In 2016, a machine learning model was employed to predict the survival of 900 patients with breast cancer, achieving a strong predictive performance with an AUC of up to 93% using Tree Random Forest [19]. Subsequently, Chen et al. proposed the deep neural network model of MULTI-NMF DNN (AMND) through survival analysis of 1489 patients, achieving an AUC of 87.04% [22], which exhibited reduced performance compared to the machine learning model. Another novel framework, GPDBN, introduced by Wang et al., effectively integrates genomic data and pathological images for breast cancer prognosis prediction [23]. Additionally, a neural network model was designed that utilizes attention techniques to learn the features of a heterogeneous tumor microenvironment, achieving an accuracy of 0.75 for breast cancer [24]. Despite the effective prognostic capabilities of machine learning, the black-box nature of the model may result in a lack of transparency and evidence for clinicians when making critical decisions.

Therefore, in 2021, Torres explored the role of machine learning models in predicting the survival of breast cancer patients [25]. However, their study has several limitations. First, the breast cancer dataset had numerous missing values, and the use of simulations to fill in these missing data may have led to deviations from real-world situations. Second, the study involved three machine learning models, and the predictive accuracy of the XGB model was only moderate (C-index of approximately 0.73), making it challenging to apply the model in clinical practice. Moreover, the clinical significance of the study was limited, with the only notable finding being the crucial role of age in breast cancer survival. Consequently, the development of an explainable artificial intelligence model capable of predicting breast cancer prognosis would hold significant importance in a clinical setting.

To address this need, we utilized SHAP, an explainable ensemble learning method based on cooperative game theory, for breast cancer clinical research based on real-world clinical data. Cooperative game theory is a branch of game theory that studies how participants collaborate to achieve common goals and how to allocate the value generated by their cooperation [26]. In cooperative game theory, SHAP value is used to allocate values within alliances and measure the contribution of each participant to the alliance. This value is based on the average contribution of participants to various possible alliances [27]. SHAP is a method used to assess the contribution of each feature to a machine learning model's output. When using machine learning models to make predictions, such as the prediction of breast cancer recurrence in this study, it is important to understand how these predictions are generated. The SHAP method helps us explain the output of such machine learning models.

In this study, we aimed to develop a clinical decision support tool for predicting the prognosis of breast cancer and explore the key factors influencing breast cancer recurrence. Subsequently, we aimed to provide a decision mechanism for the recurrence prediction model based on the identified key factors. Our analysis involved real-world clinical data from 1629 breast cancer patients with minimal missing values. The study revealed that tumor size, clinical stage III, number of lymph node metastases, and age were key factors for predicting recurrence. By incorporating these top four key factors, the method demonstrated qualified predictive capability and provided a clear explanation of its predictions.

2. Materials and methods

Ethical approval

This study was approved by the institutional review board of the Xinjiang Medical University Affiliated Cancer Hospital and was conducted in accordance with the tenets of the Declaration of Helsinki (study number K-2021028). Informed consent was obtained from each participant after a detailed explanation of the nature and possible consequences of the study. The privacy rights of the human subjects were adhered to throughout the study.

2.1. Study population

Patients diagnosed with breast cancer at the Xinjiang Medical University Affiliated Cancer Hospital between August 1, 2012, and August 1, 2014, were selected as research subjects. A total of 1629 female patients with breast cancer were included in the subsequent data analysis after applying inclusion and exclusion criteria.

These criteria were as follows:

1) Female patients with primary, unilateral breast cancer confirmed by pathological biopsy. 2) No history of other malignant tumors. 3) Complete clinicopathological data and follow-up information were available. Patients with a history of other malignant tumors were excluded from the analysis. All treatments in this study were performed according to the relevant guidelines and regulations.

Follow-up: The study endpoint was February 1, 2017. The follow-up period ranged from 2.0 to 55.0 months, with a median follow-up of 47.0 months. Eventually, a total of 1629 patients' data were collected after ruling out cases with incomplete data collection and lack of follow-up. Follow-up information defined local recurrence as the diagnosis of ipsilateral breast, chest wall, or regional lymph node recurrence of breast cancer three months after the first treatment. Distant metastasis was defined as the diagnosis of distant metastasis based on clinical or imaging evidence.

2.2. General information and data analysis

Data were discarded according to the inclusion and exclusion criteria. From August 1, 2012, to August 1, 2014, a total of 1629 women were recorded. Demographic characteristics included registration number, age, occupation, height, weight, BMI, family history of breast cancer, history of benign breast disease, time points of menarche and menopause, living area, and education. Clinicopathological data included tumor size, number of lymph node metastases, clinical stage, tumor tissue Gleason score, type of molecular markers, risk of recurrence, estrogen and progesterone receptor status, proliferation status of p53, EGFR, CK5/6, CK14, Her-2, and Ki-67. According to the Chinese guidelines for diagnosis and treatment of breast cancer [28]the risk of recurrence is determined by multiple factors, such as postoperative pathological tumor size, HER-2, ki-67, grade, LVSI, number of positive lymph nodes, and age. These factors were used to calculate the risk of recurrence and classify patients into low-, moderate-, or high-risk categories. Treatment information included surgery, cycles of neoadjuvant chemotherapy and postoperative chemotherapy, radiotherapy, and targeted and endocrine therapy.

2.3. Data preprocessing

In our study, we used a continuous age scale to represent the menopausal age of patients who had experienced menopause. However, for patients who had not yet experienced menopause, we considered that there were many cases in our collected data where patients were above the average natural menopausal age of 48.7 years for Chinese women [29] but were still pre-menopausal. To better approximate the actual situation, we did not consider the average menopausal age of pre-menopausal patients. Instead, we implemented the following processing method to define the menopausal age of these patients: patients below 50 years of age had not yet reached menopause, and their menopausal age was set at 50 years. Patients aged between 50 and 54 years had not yet reached menopause, and their menopausal age was set at 55 years. Patients aged between 55 and 59 years had not yet reached menopause, and their menopausal age was set to 60 years.

After quantifying the three molecules associated with breast cancer patient proliferation status—ki-67, p53, and EGFR—their individual measurements were combined into a single column called "ki-67_p53_EGFR." To determine positivity, a threshold of greater than 10% was applied for each biomarker. If any one of the three molecules (ki-67, p53, or EGFR) had a measurement exceeding the 10% threshold, it was recorded as "1." If two of the three molecules had measurements surpassing the threshold, it was denoted as "2." Finally, if all three molecules exhibited measurements above the threshold, it was represented as "3." This approach allows the categorization of breast cancer patients based on the combined positivity of ki-67, p53, and EGFR, providing a simplified representation of their proliferation status.

Finally, the clinical staging, pathological classification, histological grade, and molecular classification in the dataset were encoded in a "one-hot" manner to generate a dataset for further analysis. One-hot encoding is a commonly used data encoding method that transforms categorical variables into a numerical format that can be processed using machine learning algorithms.

2.4. Evaluation of ensemble learning models

Ensemble learning is a technique that combines multiple machine learning models to perform learning tasks. It aims to reduce the variance among individual models and improve the overall prediction performance of the ensemble. In this study, we applied several classic algorithms to ensemble learning, including Bagging, Random Forest, Extra Trees, AdaBoost, and Gradient Boost. To evaluate the performance of these algorithms, we plotted the Receiver Operating Characteristic (ROC) curves of each.

Next, we partitioned the dataset into three subsets: training, testing, and validation at a ratio of 7:2:1. To address the issue of imbalanced samples in the dataset (212 recurrence cases and 1417 non-recurrence cases), we used a balanced accuracy rate as an evaluation metric for the models.

Among the algorithms tested, Random Forest and Extra Trees exhibited similar prediction capabilities as both of them are ensemble learning models composed of numerous decision tree models. To gain further insights into the models' behaviors and interpret their predictions, we employed the SHAP approach. This approach, which is rooted in cooperative game theory, has proven to be effective in explaining the underlying mechanisms and feature importance of machine learning models [30].

2.5. Explainability analysis of model

SHAP is a machine learning method that seeks an explanation model g different from f that is more concise in structure than the original model f (a black-box model), with the aim of explaining f.

For the original model f with output $f(x)$ for a single input x, assume that there is a mapping function $x = h_x(x')$ simplifying the original input x to x'. Then, SHAP aims to ensure $g(z') \approx f(h_x(z'))$ whenever $z' \approx x'$. According to the local accuracy (Property 1 in Ref. [31]), $f(x) = g(z')$ when $x = h_x(x')$. In SHAP, the explanation model $g(z')$ is expressed as

$$g(z') = \varphi_0 + \sum_{i=1}^{M} \varphi_i z_i'$$

where φ_i can be calculated by Equation 8 in Ref. [31].

The explanation model $g(z')$ is an approximation of the original model $f(x)$. The approximation process is shown in Fig. S1 (Supplementary Figure), where $[z_S] = f(h_x(z'))$, and S is the set of non-zero indices in z'. Initially, $E[f(z)]$ is the output of the original model when the information about the input is not available, and the value of $E[f(z)]$ is exactly φ_0; $\varphi_1 + \varphi_0$ is the output when only the first feature $z_1' \neq 0$ so that φ_1 measures the contribution of the first feature to push $g(z')$ forward to $f(x)$. Then, by setting the features to non-zero values one-by-one in the order of $z_1', z_2', \ldots, z_M'$, all the feature weights (or contributions) φ_i can be derived successively. Thus, the output of the explanatory model $g(z')$ will approach $f(x)$ eventually.

As shown in Fig. S1, the contributions of features may "push" the explanation model to either the right (along the blue arrow) or left

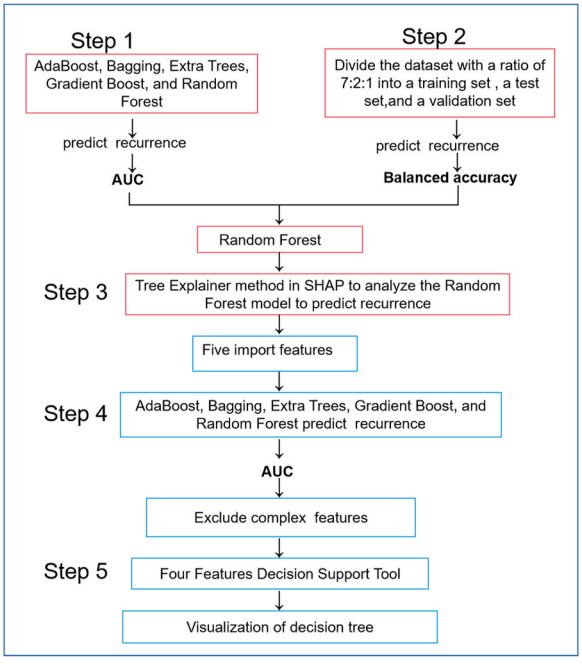


Fig. 1. Overview of study design.

(along the red arrow), which indicates the positive and negative contributions of features to the original model f.

2.6. Decision support model construction

The top five features identified as critical for breast cancer recurrence were selected based on their SHAP values, with a threshold of 0.04. These features were then used as inputs for five ensemble learning classification models to construct a decision support tool. The accuracies of the models were measured to evaluate the performance of the decision support tool in recurrence prediction.

To further analyze the impact of these features, a decision tree model was constructed using the selected features. The decision tree classifier had a maximum depth of 3, indicating that the tree would have three levels of branching to make predictions about breast cancer recurrence. To visualize the prediction process of the decision tree model, the dtreeviz library (available at https://github.com/parrt/dtreeviz) was utilized. This library provides methods to visually represent decision trees and interpret their predictions. By using dtreeviz, the decision tree model's prediction process for breast cancer recurrence could be effectively visualized, allowing for a clearer understanding of the logic and decision-making behind the model.

3. Results

In this study, we employed SHAP methods and analyzed real-world data from 1629 breast cancer patients to identify the key factors associated with breast cancer recurrence. Subsequently, we used these identified factors to create a recurrence prediction model and establish a decision mechanism for the tool. Fig. 1 presents an overview of the study design. The proposed explainable ensemble

Table 1
Clinicopathologic characteristics of breast cancer patients.

Item	Number
Total	1629
Age (Year, Mean)	48.16 ± 10.45
Gravidity (N, Mean)	2.51 ± 1.71
Parity (N, Mean)	1.72 ± 1.28
Menarche age (Year, Mean)	13.71 ± 1.91
Menopausal age (Year, Mean)	49.44 ± 3.51
Urban areas (N, %)	1156 (70.96)
Rural areas (N, %)	473 (29.04)
Clinical stage (N, Mean) (N, %)	
In Situ	65 (4.00)
Stage I	443 (27.19)
Stage II	693 (42.54)
Stage III	346 (21.24)
Stage IV	82 (5.03)
Histological Grade (N, %)	
G1	60 (3.68)
G2	773 (47.45)
G3	283 (17.37)
No data	513 (31.50)
Pathological Type (N, %)	
Type I (Carcinoma in situ)	65 (3.99)
Type II (Medullary carcinoma, Mucinous carcinoma)	66 (4.05)
Type III (Invasive ductal carcinoma and Invasive lobular carcinoma)	1446 (88.77)
Type IV (Other type)	52 (3.19)
Molecular typing (N, %)	
Luminal A	285 (17.50)
Luminal B	797 (48.93)
Basel Type	252 (15.47)
Triple Negative	295 (18.11)
Tumor diameter (cm, Mean)	2.65 ± 2.10
≤2	720 (44.20)
2–5	209 (12.83)
≥5	167 (10.25)
Total Lymph Nodes Metastases (N, Mean)	3.13 ± 6.56
Total Lymph Nodes Metastases (N, %)	
0	830 (50.95)
0–3	446 (27.38)
4–9	172 (10.56)
≥10	180 (11.05)
The number of Chemotherapy cycles (N, Mean)	5.13 ± 2.16
Recurrence (N, %)	
Positive	212 (13.01)
Negative	1417 (86.99)
Recurrence time (Month, Mean)	16.75 ± 10.36

learning method demonstrated promising results in predicting breast cancer recurrence, outperforming existing methods with high accuracy and transparency.

3.1. General information

The clinical characteristics of the 1629 female breast cancer patients are summarized in Table 1. The mean ages of diagnosis, menarche, and menopause and their standard deviations were 48.16 ± 10.45, 13.71 ± 1.91, and 49.44 ± 3.51, respectively. A total of 173 (47.45%) and 1156 (70.96%) patients were from rural and urban areas, respectively. The two main pathological types were invasive ductal carcinoma and invasive lobular carcinoma. Recurrence occurred in 13.01% of patients, and the average recurrence time was 16.75 ± 10.36 months. A total of 44.2% of the patients had a tumor diameter of less than 2 cm, and 50.95% had lymph node metastasis.

3.2. Prediction performance of different ensemble learning methods

First, we used the five most common ensemble learning classification models to predict breast cancer recurrence: AdaBoost, Bagging, Extra Trees, Gradient Boost, and Random Forest. The AUCs of the five models were 0.91, 0.99, 0.99, 0.98, and 0.99, respectively (Fig. 2). These results indicate that the ensemble learning models performed well in predicting breast cancer recurrence.

Subsequently, we divided the data into training, test, and validation sets to better evaluate the five integrated learning models. To eliminate the impact of unbalanced samples on model evaluation (212 recurrences, 1417 no recurrence), we used a balanced readiness rate to evaluate the models, which could help avoid the exaggerated evaluation performance of unbalanced datasets:

$$balanced\ accuracy = \frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)=\frac{1}{2}(TPR+TNR)$$

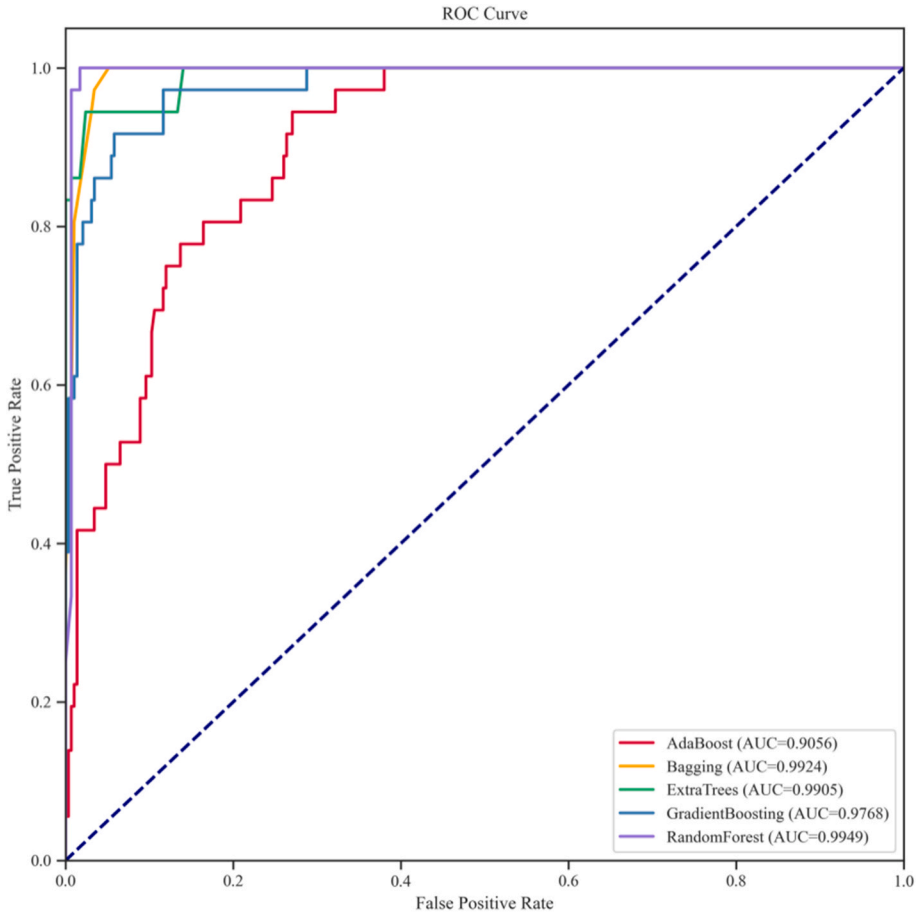


Fig. 2. ROC curves of five ensemble learning models to predict breast cancer recurrence. Red, orange, green, blue, and purple represent the five models of AdaBoost, Bagging, Extra Trees, Gradient Boosting, and Random Forest, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

where TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative, TPR is True Positive Rate, and TNR is True Negative Rate. The train_test_validation split method of sklearn was used to divide the dataset in a ratio of 7:2:1. When dividing the dataset, we set random_state from 0 to 99 and split the dataset 100 times repeatedly. The AdaBoost, Bagging, Extra Trees, Gradient Boost, and Random Forest models were scored, and the average accuracies over 100 iterations of the results of the five models were 0.60, 0.71, 0.81, 0.85, and 0.86, respectively. Furthermore, we employed a *t*-test to compare the differences in accuracy between different models. Our findings indicate that both the Extra Trees and Random Forest models significantly outperformed the remaining three models, as depicted in Fig. 3. However, due to the radical nature of the Extra Tree algorithm, which randomly selects a feature to split the decision tree, the selected feature may not always be the optimal splitting point. This can lead to higher variance in the prediction results of the model. Therefore, we adopted the Random Forest model to predict breast cancer recurrence.

3.3. Interpretability analysis of Random Forest model based on SHAP

We used the Tree Explainer method in SHAP to analyze the Random Forest model to predict breast cancer recurrence. By generating an Explainer for the model, we were able to identify key features in the dataset that contribute to the prediction of breast cancer recurrence. In a binary classification problem like ours, the output of the tree model corresponds to the sample labels being classified as either 0 (no recurrence) or 1 (recurrence). The Explainer calculates the contribution of each feature to the model's prediction outcome when the tree model output is 0 or 1. This contribution is represented as a SHAP value. To identify the most influential features, we evaluated the sum of the absolute average contributions (average SHAP values) of both output types (0 and 1). As depicted in Fig. 4, we identified the top five significant features, which were tumor size, clinical stage III, total metastatic lymph nodes, risk of recurrence, and age. It is reasonable to conclude that these features play a crucial role in improving the prediction accuracy of the model.

3.4. Using features discovered by the explainable ensemble learning model to predict breast cancer recurrence

To predict breast cancer recurrence in the validation dataset, we initially used tumor size, clinical stage III, and total metastatic lymph nodes as features. The decision tree model constructed using these three features achieved an AUC ranging from approximately 0.73 to 0.84, with limited clinical significance in distinguishing between different risk levels. To improve the predictive capability of the model, we decided to incorporate two additional features. However, considering the convenience of applying this model in a clinical setting, we aimed to select easily obtainable features. Therefore, we compared the predictive abilities of the model when including risk of recurrence and age as features (Table 2 and Fig. S2). After removing the complex calculation of risk of recurrence from the model parameters, we observed that its exclusion did not significantly impact the predictive ability of the model, as indicated in Table 2 and Fig. S2. Additionally, the calculation method for this risk factor is specific to Chinese guidelines and may not be widely recognized internationally. Consequently, we chose to include age as a readily obtainable feature. The inclusion of age as a feature

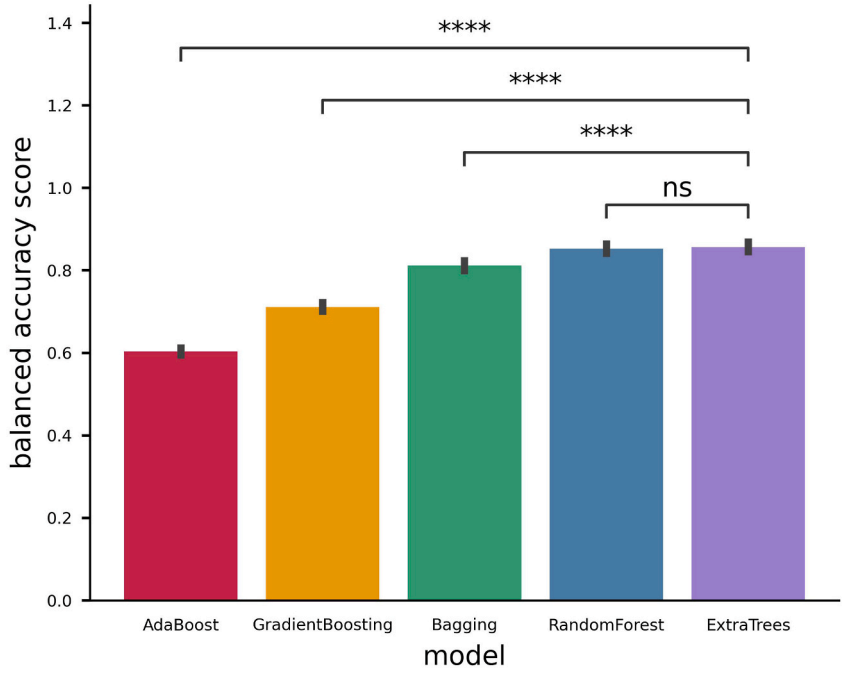


Fig. 3. Comparison of five ensemble learning models in predicting breast cancer recurrence.
ns: p≤1.00e+00, *: 1.00e-02 < p≤5.00e-02, **: 1.00e-03 < p≤1.00e-02, ***: 1.00e-04 < p≤1.00e-03, ****: p≤1.00e-04.

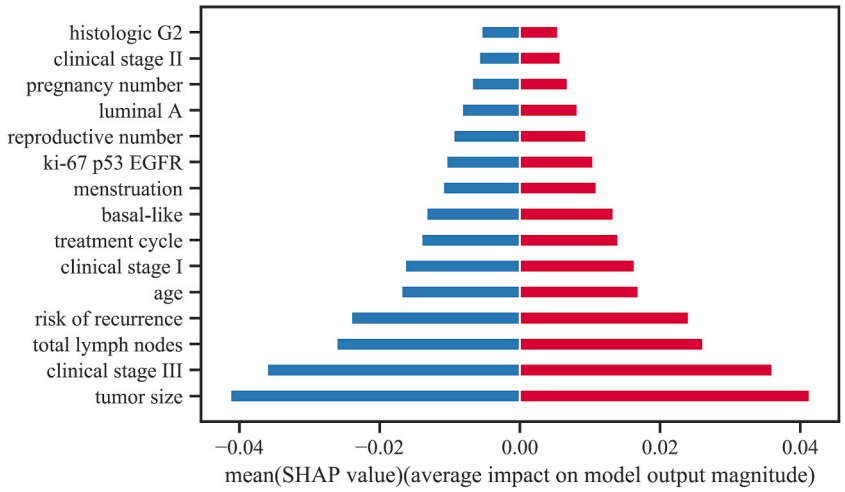


Fig. 4. Key features identified using the SHAP value analysis are represented by the colors red and blue, indicating labels 0 and 1, respectively. The horizontal axis represents the average SHAP value, while the vertical axis represents different clinical features. In this context, the term "Menstruation" refers to the age of menarche. "Basel-like" represents the Basel Type in molecular subtyping. "Total lymph nodes" refers to the total number of metastatic lymph nodes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Influence of the top five features on the AUC values across the five different algorithms.

Algorithm	AdaBoost	Bagging	Extra Trees	Gradient Boost	Random Forest
Three features (tumor size, clinical stage III, and total metastatic lymph nodes)	0.7320	0.8110	0.8436	0.7928	0.8191
Four features (tumor size, clinical stage III, and total metastatic lymph nodes, **risk of recurrence**)	0.7991	0.8794	0.9144	0.8210	0.8860
Five features (tumor size, clinical stage III, and total metastatic lymph nodes, **risk of recurrence, age**)	0.7858	0.9468	0.9770	0.8330	0.9655
Four features (tumor size, clinical stage III, and total metastatic lymph nodes, **age**)	0.7617	0.9344	0.9709	0.8217	0.9607

significantly improved the performance of the Extra Trees and Random Forest models. The AUC for the Extra Trees model increased to 0.97, while that of the Random Forest model improved to 0.96. This modification allowed us to enhance the accuracy of the model in predicting breast cancer recurrence by incorporating a more easily accessible feature while maintaining a high level of performance.

To illustrate the decision tree process more clearly, we visualized the decision tree model using dtreeviz (https://github.com/parrt/dtreeviz), as shown in Fig. 5. This figure shows the decision-making process using Random Forest. First, patients were divided into clinical stage III and non-clinical stage III groups, which were further divided according to tumor size. Subgroups in clinical stage III could be further divided into categories of recurrence and no recurrence based on the total metastatic lymph nodes, while one non-clinical stage III subgroup was divided by age and the other by the total number of metastatic lymph nodes.

4. Discussion

In this study, we introduced an explainable ensemble learning method that demonstrates remarkable performance in predicting breast cancer recurrence using only a small number of factors. Initially, we compared five common ensemble learning classification models — AdaBoost, Bagging, Extra Trees, Gradient Boost, and Random Forest — to determine their predictive capabilities for breast cancer recurrence. Through this comparison, we identified Random Forest as the most suitable model owing to its high prediction accuracy and AUC. To gain deeper insight into the predictions made by the Random Forest model, we employed the Tree Explainer method in SHAP. This analysis identified five key factors strongly associated with cancer recurrence: tumor size, clinical stage III, total lymph nodes, risk of recurrence, and age. By utilizing these four key factors, our decision support tool for breast cancer recurrence prediction achieved significant improvements. After removing the complex calculation of risk of recurrence from the model parameters, we observed that its exclusion did not significantly impact the predictive ability of the model. Therefore, we chose tumor size, clinical stage III, total lymph nodes, risk of recurrence, and age as features for our clinical decision support tool for breast cancer recurrence prediction. The Extra Trees model exhibited an increased AUC of 0.97, while the Random Forest model demonstrated an improved AUC of 0.96.

In the medical field, where AI decision-making can significantly impact patient care, there are concerns about theoretical flaws in machine learning decision mechanisms. Data-driven machine learning may establish associations between input data and outcomes,

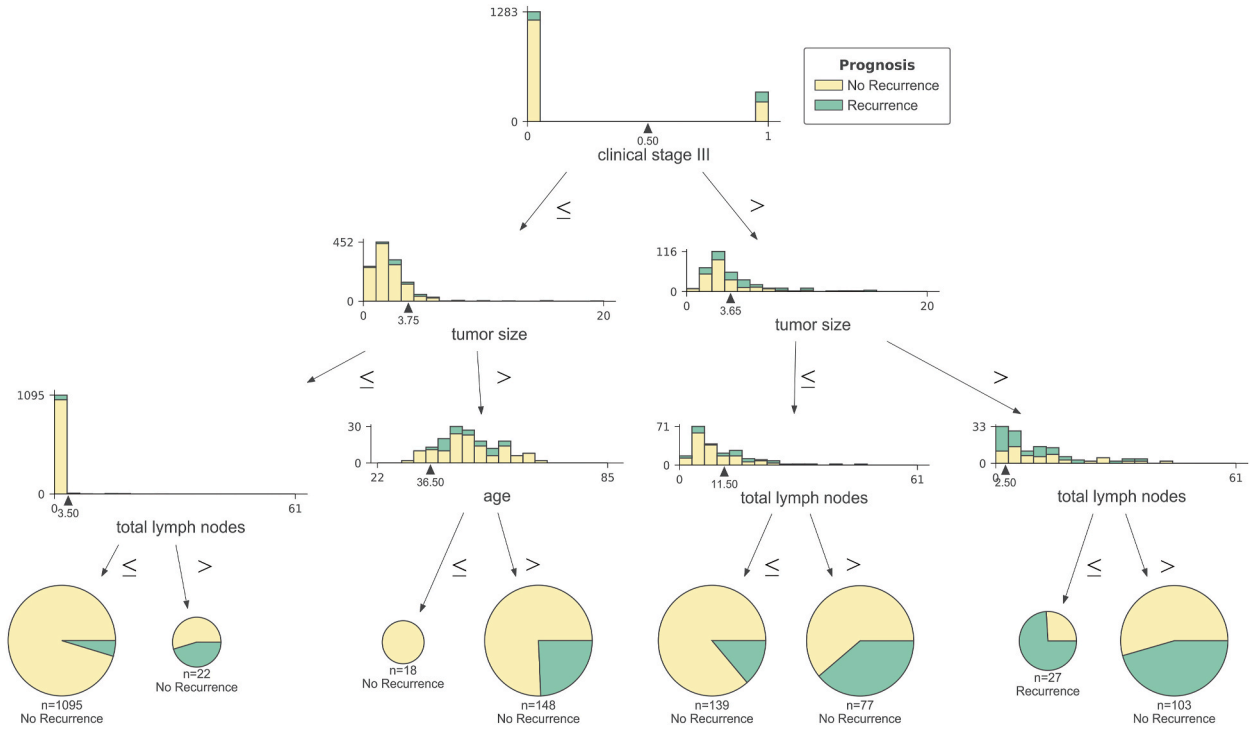


Fig. 5. Visualization of decision tree for predicting breast cancer recurrence using key clinical features. Yellow represents the non-recurrence patients, and green represents the recurrence population. Clinical state III, tumor size, age, and total lymph nodes are the reference indicators for decision-making. An arrow with '≤' denotes that the classification is below the threshold of random classification decision, while an arrow with '>' indicates it is above the threshold of classification decision. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

but it may not accurately identify true cause-effect relationships due to data limitations and biases. Additionally, black-box models, such as deep neural networks, can lead to low-level errors, raising security risks in certain applications [32]. Consequently, there is a growing need for AI systems in medicine to provide reasonable explanations of decision-making processes.

The main idea of the SHAP method is to calculate the SHAP values for each feature based on permutations and combinations of feature values. The SHAP value can be understood as the average marginal contribution of a feature value to the output. By taking a weighted average of all possible combinations of feature values, the SHAP method quantifies the impact of each feature value on the output [33–35]. In this study, the top five features with the highest SHAP values were tumor size, clinical stage III, total metastatic lymph nodes, risk of recurrence, and age. This indicates that these five features significantly influence the prediction of breast cancer recurrence. When using the Random Forest model with the four selected features to predict recurrence, an AUC of 0.96 was achieved, further confirming the significant impact of these features on the prediction. Additionally, our model provides a decision mechanism for recurrence prediction through the use of the Tree Explainer method in SHAP, which allows us to gain a clear understanding of the decision-making process at each step. This is in contrast to the black-box decision-making process of previous AI models. Our findings not only demonstrate the predictive power of the four features but also provide transparency and interpretability in the decision-making process. By utilizing the Tree Explainer method in SHAP, we can gain insights into how each feature contributes to the overall prediction, enabling clinicians to better understand and trust the model's predictions. This transparency enhances the clinical utility and acceptance of our model in real-world applications.

This study collected reliable and complete follow-up data over a 55-month period. In contrast, Moncada et al. conducted a similar study [25] using data from the Netherlands Cancer Registry from 2002 to 2005, which is outdated compared with the current study. Their data collection was incomplete, lacking information such as age, tumor characteristics, (hormonal) receptor status, clinical and pathological TNM staging, and number of removed and positive lymph nodes. The risk factors identified through SHAP in our study are consistent with clinical guidelines and medical expertise [36–38]. Our findings have significant implications for clinicians, as they can utilize the four-feature-predicted risk of breast cancer recurrence to develop personalized treatment options. This information enables clinicians to make informed decisions regarding treatment strategies, including performing needle biopsies or transferring patients to specialized hospitals for more effective care. Ultimately, these personalized approaches contribute to improved patient outcomes [39, 40].

Our study had some limitations. First, the study has a limited research duration of up to 55.0 months, which may not be sufficient for examining longer breast cancer survival periods. Due to the imbalance in our dataset, with only 13.01% of patients experiencing recurrence, there is a potential for bias in the predictive model. This is because there is a relatively smaller proportion of recurrent

events compared to non-recurrent events. Second, the study was conducted at a single center without multi-center validation. Third, the SHAP method focuses on the individual feature contributions and may overlook the differences between samples, which can lead to potential errors. In the future, in addition to analyzing the feature contributions of individual samples, exploring the feature contributions at the sample set level can provide a more comprehensive understanding of the model's behavior.

In conclusion, our study introduced an explainable ensemble learning method for breast cancer recurrence prediction and identified key factors with strong associations. The SHAP method was validated using real-world clinical data, and the transparent decision-making process of the model enhances trust and facilitates its application in clinical settings, aiding clinicians in developing personalized treatment plans based on accurate predictions. These findings have significant implications for the improvement of patient care and outcomes in breast cancer management.

Ethics approval and consent to participate

This study was reviewed and approved by Xinjiang Medical University Affiliated Cancer Hospital, with the approval number [study number K-2021028]. All participants/patients provided informed consent to participate in the study.

Consent for publication

Not applicable.

Funding

This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2022D01A01 and 2022D01A02).

Data availability statement

The patient datasets used and/or analyzed during the current study are not available due to patient data privacy protection. Other data without privacy concerns are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Ying Liu: Writing – review & editing, Writing – original draft, Funding acquisition. **Yating Fu:** Methodology, Formal analysis. **Yadong Peng:** Formal analysis, Data curation. **Jie Ming:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24876.

References

[1] M.P. Coleman, M. Quaresma, F. Berrino, J.-M. Lutz, R. De Angelis, R. Capocaccia, et al., Cancer survival in five continents: a worldwide population-based study (CONCORD), Lancet Oncol. 9 (2008) 730–756.
[2] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2021, CA A Cancer J. Clin. 71 (2021) 7–33.
[3] F.A. Vicini, M. Sharpe, L. Kestin, A. Martinez, C.K. Mitchell, M.F. Wallace, et al., Optimizing breast cancer treatment efficacy with intensity-modulated radiotherapy, Int. J. Radiat. Oncol. Biol. Phys. 54 (2002) 1336–1344.
[4] L.N. Shulman, W. Willett, A. Sievers, F.M. Knaul, Breast cancer in developing countries: opportunities for improved survival, JAMA Oncol. 2010 (2010) 595167.
[5] J.-Y. Kim, Y.S. Lee, J. Yu, Y. Park, S.K. Lee, M. Lee, et al., Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer Cohort in tertiary cancer center Registry, Front. Oncol. 11 (2021) 596364.
[6] K. Polyak, Heterogeneity in breast cancer, J. Clin. Invest. 121 (2011) 3786–3788.
[7] S. Simard, B. Thewes, G. Humphris, M. Dixon, C. Hayden, S. Mireskandari, et al., Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies, J. Cancer Surviv. 7 (2013) 300–322.
[8] A.L. Corter, M. Findlay, R. Broom, D. Porter, K.J. Petrie, Beliefs about medicine and illness are associated with fear of cancer recurrence in women taking adjuvant endocrine therapy for breast cancer, Br. J. Health Psychol. 18 (2013) 168–181.
[9] I. Spronk, F.G. Schellevis, J.S. Burgers, G.H. de Bock, J.C. Korevaar, Incidence of isolated local breast cancer recurrence and contralateral breast cancer: a systematic review, Breast 39 (2018) 70–79.
[10] C.L. Park, D. Cho, T.O. Blank, J.H. Wortmann, Cognitive and emotional aspects of fear of recurrence: predictors and relations with adjustment in young to middle-aged cancer survivors, Psycho Oncol. 22 (2013) 1630–1638.
[11] S. Loibl, P. Poortmans, M. Morrow, C. Denkert, G. Curigliano, Breast cancer, Lancet 397 (2021) 1750–1769.

[12] Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies, Lancet Oncol. 13 (2012) 1141–1151.
[13] D.H. Morris, M.E. Jones, M.J. Schoemaker, A. Ashworth, A.J. Swerdlow, Secular trends in age at menarche in women in the UK born 1908-93: results from the Breakthrough Generations Study, Paediatr. Perinat. Epidemiol. 25 (2011) 394–400.
[14] E.B. Gold, The timing of the age at which natural menopause occurs, Obstet. Gynecol. Clin. N. Am. 38 (2011) 425–440.
[15] M.H. Gail, L.A. Brinton, D.P. Byar, D.K. Corle, S.B. Green, C. Schairer, et al., Projecting individualized probabilities of developing breast cancer for white females who are being examined annually, J. Natl. Cancer Inst. 81 (1989) 1879–1886.
[16] J. Cuzick, A breast cancer prediction model incorporating familial and personal risk factors, Hered. Cancer Clin. Pract. 10 (2012), https://doi.org/10.1186/1897-4287-10-s2-a29.
[17] M.J. Hall, J.E. Reid, L.A. Burbidge, D. Pruss, A.M. Deffenbaugh, C. Frye, et al., BRCA1 and BRCA2 mutations in women of different ethnicities undergoing testing for hereditary breast-ovarian cancer, Cancer 115 (2009) 2222–2233.
[18] X. Yang, X. Ai, J.M. Cunningham, Computational prognostic indicators for breast cancer, Cancer Manag. Res. 6 (2014) 301–312.
[19] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13 (2015) 8–17.
[20] Z. Huang, X. Zhan, S. Xiang, T.S. Johnson, B. Helm, C.Y. Yu, et al., SALMON: survival analysis learning with multi-omics neural networks on breast cancer, Front. Genet. 10 (2019) 166.
[21] R. Li, X. Wu, A. Li, M. Wang, HFBSurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction, Bioinformatics 38 (2022) 2587–2594.
[22] H. Chen, M. Gao, Y. Zhang, W. Liang, X. Zou, Attention-based Multi-NMF deep neural network with multimodality data for breast cancer prognosis model, BioMed Res. Int. 2019 (2019) 9523719.
[23] Z. Wang, R. Li, M. Wang, A. Li, GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction, Bioinformatics 37 (2021) 2963–2970.
[24] Y. Lee, J.H. Park, S. Oh, K. Shin, J. Sun, M. Jung, et al., Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning, Nat. Biomed. Eng. (2022), https://doi.org/10.1038/s41551-022-00923-0.
[25] A. Moncada-Torres, M.C. van Maaren, M.P. Hendriks, S. Siesling, G. Geleijnse, Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival, Sci. Rep. 11 (2021) 6968.
[26] S. Lipovetsky, M. Conklin, Analysis of regression in game theory approach, Appl. Stoch Model Bus. Ind. 17 (2001) 319–330.
[27] E. Štrumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, Knowl. Inf. Syst. 41 (2014) 647–665.
[28] National Health Commission of PRC, National Health Commission of the People's Republic of China, Chinese guidelines for diagnosis and treatment of breast cancer 2018 (English version), Chin. J. Cancer Res. 31 (2019) 259–277.
[29] M. Wang, C. Kartsonaki, Y. Guo, J. Lv, W. Gan, Z.-M. Chen, et al., Factors related to age at natural menopause in China: results from the China Kadoorie Biobank, Menopause 28 (2021) 1130–1142.
[30] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, et al., Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, Nat. Biomed. Eng. 2 (2018) 749–760.
[31] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, arXiv [cs.AI]. 2017. Available: http://arxiv.org/abs/1705.07874.
[32] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (2019) 206–215.
[33] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, arXiv [cs.CV]. 2017. Available: http://arxiv.org/abs/1704.02685.
[34] A. Datta, S. Sen, Y. Zick, Algorithmic transparency via quantitative input influence: theory and experiments with learning systems, in: 2016 IEEE Symposium on Security and Privacy (SP), IEEE, 2016, https://doi.org/10.1109/sp.2016.42.
[35] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLoS One 10 (2015) e0130140.
[36] C.E. DeSantis, J. Ma, M.M. Gaudet, L.A. Newman, K.D. Miller, A. Goding Sauer, et al., Breast cancer statistics, CA A Cancer J. Clin. 69 (2019) 438–451, 2019.
[37] S.B. Edge, C.C. Compton, The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM, Ann. Surg Oncol. 17 (2010) 1471–1474.
[38] T.J. Eberlein, Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC trial, Year Bk. Surg. 2007 (2007) 297–298.
[39] L. Yin, J.-J. Duan, X.-W. Bian, S.-C. Yu, Triple-negative breast cancer molecular subtyping and treatment progress, Breast Cancer Res. 22 (2020) 61.
[40] W. Haque, V. Verma, S. Hatch, V. Suzanne Klimberg, E. Brian Butler, B.S. Teh, Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy, Breast Cancer Res. Treat. 170 (2018) 559–567.