

Prognostic Model and Influencing Factors for Breast Cancer Patients

Qing Zhang *

School of Economics and Management, Hebei University of Technology, Tianjin, China

* Corresponding author Email: zhangqingaxx@163.com

Abstract: Breast cancer is a common disease that affects women's life and health. Survival analysis of breast cancer patients can help doctors and patients understand the prognosis of patients and provide guidance for clinical treatment. In this study, experiments were conducted based on SEER breast cancer patient data, and feature selection was performed first, followed by the construction of prognostic models using four survival analysis methods. The C-Index, BS, and IBS indexes of the RSF model were 0.8535, 0.0853, and 0.0512, respectively, which had the best predictive effect in the prognostic model for breast cancer patients. Based on the SHAP method to analyze the important factors affecting the prognosis of breast cancer patients, the results showed that tumor stage, TNM stage, grade and age have a great impact on the prognosis of breast cancer patients.

Keywords: Prognostic Model; Influencing Factors; Breast Cancer.

1. Introduction

According to the latest analysis of global cancer burden data released by the International Agency for Research on Cancer (IARC) in 2020 [1], the number of breast cancer patients worldwide in 2020 alone was 2.26 million, accounting for 11.70% of the total number of cancer patients, making breast cancer the number one "killer" of women's health. Breast cancer is a common disease that affects women's life and health, and patient survival is affected by many factors, including tumor stage, histological grade, patient age and breast subtype [2]. Survival of breast cancer patients reflects the disease control and treatment effect of patients, which is an important reference value for assessing patients' prognosis and formulating appropriate treatments. The use of patient data to establish relevant models can help doctors to estimate breast cancer patients more accurately, so that they can take relevant measures to improve the treatment effect and survival quality of breast cancer patients in a targeted manner.

Prognostic modeling of breast cancer patients can predict the survival rate at any point in time, thus helping doctors and patients to understand the survival status of patients at different points in time, so that the treatment plan can be adjusted in time to provide better care for the patients. When studying the prognosis of breast cancer patients, there may be patients who are lost in the middle of the study or drop out of the study and do not record the last death event, which is called Censored Data. For these patients, it is not known whether they died, but it is known that they did not die at the last follow-up visit. This data, which does not record the occurrence of a patient's death, is called censored data. In cases where censored data exist, analysis using only data with observable death events without considering censored data can result in the loss of important information and inaccurate overall survival of breast cancer patients at a given point in time. Survival analysis methods can take into account censored data and incorporate this uncertainty by making full use of known data information to build models that predict patient survival at any given moment.

For the patient record, each patient will consist of a set of

characteristic variables $X \in R^d$ and the last observed survival time. For patients, only one of the censoring and death events will occur, so the indicative function $\sigma \in \{0; 1\}$ and the observed survival time $y > 0$ are used to represent the actual events and times that occur for the patient. Prediction of survival of breast cancer patients at any moment using survival analysis is achieved by using the data to construct a Survival Function. The survival function describes the probability of an individual surviving until a given time point T , as shown in Equation (1), and is usually expressed as $S(t)$, where t denotes time. $S(t)$ takes values between 0 and 1, and a larger $S(t)$ indicates a higher probability of an individual surviving until moment t , and vice versa.

$$S(t) = \Pr(T > t) \quad (1)$$

The problem of predicting the survival rate of breast cancer patients at any moment is a special problem related to both survival time and survival status, which contains a large amount of censored data, and it is necessary to construct a prediction model applicable to breast cancer patients based on their actual survival situation. In addition, previous prediction models using traditional survival analysis methods have significant shortcomings in prediction accuracy, and it is necessary to construct a prediction model for breast cancer patients' survival using more accurate machine learning survival analysis methods to address this problem feature.

2. Literature Review

As the most prevalent cancer worldwide, prediction of survival in breast cancer is an important research direction. Breast cancer survival prediction aims to predict the survival period and survival rate of breast cancer patients based on their clinical characteristics and treatment information, and to provide patients with personalized treatment plans and management recommendations. In recent years, breast cancer survival prediction has been extensively studied and investigated by many researchers using different methods and techniques.

The selection of breast cancer survival prediction research methods has gone through three stages: traditional survival

analysis methods, machine learning-based classification methods, and machine learning-based survival analysis methods. The traditional survival analysis methods focus on the exploration of the overall survival of breast cancer patients and are used to describe the survival of a group, based on which the individual influencing factors are analyzed in depth. Machine learning-based classification methods are used to predict and judge the survival time of individual patients and improve the prediction of individual patient survival by the model. The machine learning-based survival analysis method fully combines the advantages of survival analysis and machine learning, and is able to consider both the survival time and survival status of patients, as well as to accurately predict individual patients.

Traditional survival analysis methods have focused on investigating the factors influencing the survival of breast cancer patients using the CPH method and the KM method. Existing studies have found that the factors affecting survival of breast cancer patients are multifaceted and can be divided into four main areas of influence: patient demographic characteristics, tumor characteristics and subtypes, the impact of other tumors, and the impact of treatment modalities. In terms of patient demographic characteristics, race [3], health care and social factors associated with racial differences [4] were significantly associated with survival of breast cancer patients. In terms of tumor characteristics and subtypes, Liu et al [5] found that variables such as tumor location, grade, regional lymph node status, and tumor size were significantly associated with survival in young women with early-stage breast cancer. Using the CPH approach, Han et al [6] found that estrogen receptor or progesterone receptor status was a risk factor for death in breast cancer patients. Zhang et al [7] noted through their study that in good subtypes in which patients did not need systemic chemotherapy to achieve a high survival rate. Regarding the effect of other tumors, Kim et al [8] found that two or more primary cancers and short intervals between the occurrence of multiple primary cancers were adverse factors for lower survival in breast cancer patients. Pruitt et al [9] studied 138,576 women diagnosed with breast cancer and found that the overall survival of patients was influenced by the type, timing and stage of previous cancer. Regarding the effect of treatment modality, existing studies have focused on the effect of surgical modality and radiotherapy on patient survival. Several studies have found that breast-conserving mastectomy is an appropriate approach to achieve long-term survival in breast cancer patients than total mastectomy [10-12]. Wang et al [13] found that chemotherapy was not an important factor in survival, but patients with hormone receptor HR-positive high-risk breast cancer may benefit from chemotherapy. Li et al [14] noted that the effect of radiotherapy on secondary malignancy incidence was different in different types of cancers with different effects. The factors affecting the survival of breast cancer patients are numerous and complex, and the results of the above studies can serve as an important basis for the initial selection of influencing factors.

Machine learning-based classification methods and machine learning-based survival analysis methods focus on accurate prediction of survival time and survival of individual breast cancer patients. Kate and Nadig [15] used three different machine learning methods to build models for predicting breast cancer survival for each stage and compared them with a traditional joint model built for all stages. Salehi et al [16] improved for MLP and developed two machine

learning techniques, stacked MLP and hybrid MLP, to predict survival of breast cancer patients with prediction accuracy of 84.32% and 83.86%, respectively. Sedighi-Maman and Mondello [17] combined sampling, feature selection, and machine learning methods to predict different stages of breast cancer patients' survival status and specific survival months for deceased patients. Kaur et al [18] proposed a stacked integrated model with parallel Bayesian parameter optimization technique to adjust hyperparameters to predict survival of breast cancer patients. Han et al [6] predicted 3-, 5- and 10-year survival of small breast cancer patients using single factor analysis and multifactor analysis methods.

At this stage, research on breast cancer survival prediction is still mostly based on traditional survival analysis methods, and the application of machine learning-based classification methods and machine learning-based survival analysis methods in breast cancer survival prediction is still insufficient. In addition, less attention has been paid to the imbalance of breast cancer patient survival in machine learning-based classification methods. The application of existing machine learning models has only improved the prediction accuracy of cancer patients' survival, but few studies have further interpreted their prediction results, i.e., most of the existing machine learning models are not explanatory and do not play a good role in understanding breast cancer patients' survival.

3. Experimental Analysis

In this study, a survival prediction model based on the machine learning survival analysis method was established to predict the survival rate of breast cancer patients at any moment and analyze the related influencing factors by gaining an in-depth understanding of the survival of breast cancer patients. The machine learning survival analysis method not only effectively handles the large amount of censored data in breast cancer patient survival data, but also has significant advantages in terms of prediction accuracy. The process of model construction is shown below.

First, the raw data were input and preprocessed. In this study, we used the breast cancer patient data extracted from the SEER database, taking into account the censored data from missed visits or withdrawal from the study, and each entry contained various factors affecting the survival of breast cancer patients as well as the survival time and survival status recorded at the last follow-up visit. The data of breast cancer patients were sequentially processed for missing values, feature coding and normalization.

Next, feature screening. To prevent the survival analysis method from losing its good nature in prediction if the number of features is too large, this chapter uses KM method, Log-rank method and one-way CPH method to perform significance analysis on all features to get a preliminary understanding of the effect of each feature on the survival time of breast cancer patients. Feature screening was performed based on the relevant test results to lay the foundation for building an accurate survival prediction model for breast cancer patients.

Again, the survival analysis model was constructed and evaluated. The data after feature screening were divided into training and test sets, and the training set was used for model construction. In this paper, four survival analysis methods were used in model construction: CPH, Cox-ElasticNet, DeepSurv, and RSF, followed by using the constructed models to predict the samples in the test set and evaluating the

effect of each model.

Finally, the influencing factors are analyzed. The best-performing survival analysis prediction model was selected for prediction and the SHAP explanatory model was used to explain the influencing factors.

3.1. Data Description

This study used breast cancer patient data from the SEER repository, and the sample of breast cancer data was appropriately screened for complete information on characteristics, patient age, and time period of collection. Considering the prediction of survival of breast cancer patients at any point in time, the data of breast cancer patients who were lost to follow-up and withdrew from the study were also included in this chapter, and they were modeled using survival analysis methods. The total amount of data used in this chapter was 36230 breast cancer patient sample data, of which 3864 patient samples were observed to have complete survival time and 32,366 patient samples were censored, and the censored percentage of this breast cancer patient sample data was approximately 89.33%.

Preprocessing of the extracted data was performed. The data containing missing values were processed first. Since the data samples used in this study were very sufficient, deleting some of the samples containing missing values did not have a significant impact on the overall data. Therefore, the samples containing missing values are removed. Then, the serial type data were converted into numerical type, and the subtype data

were one-hot coded. Finally, all the data are normalized to prevent the model prediction performance from being affected by different factors taking different ranges of values.

In order for the data to be applied to the survival analysis method modeling, a distinction needs to be made between censored samples and uncensored samples. The variable δ_i is introduced according to the last recorded survival status in the original data of breast cancer patients, where $\delta_i=1$ means uncensored, i.e., the breast cancer patient eventually died and the patient's survival time O_i was observed; $\delta_i=0$ means censored, the breast cancer patient is still alive, but the recorded survival time is less than the study duration, and the recorded time C_i is the censored time at this time. A uniform metric was defined for survival time, i.e., observation time $T_i = \begin{cases} O_i & \delta_i = 1 \\ C_i & \delta_i = 0 \end{cases}$.

Before predicting survival from the breast cancer patient data, the KM method was used to get an overall picture of the survival of the breast cancer patient population, as shown in Figure 1.

Figure 1 shows the overall survival curves for breast cancer patients, with the overall survival rate of patients becoming smaller as they survive longer. Overall, the steepness of the survival function for breast cancer patients is relatively consistent, with a slight slowdown in the survival curve after 5 years.

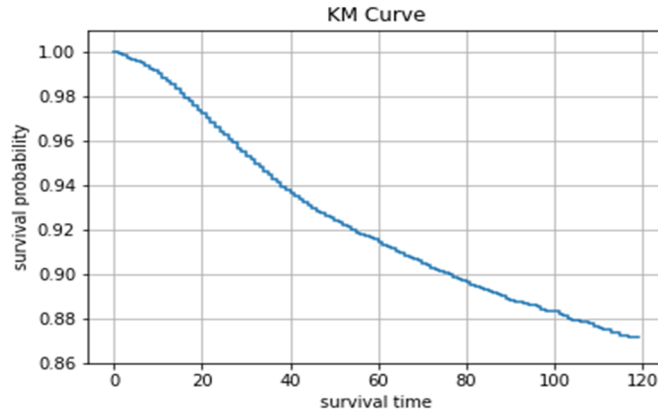


Fig 1. Survival KM curve of breast cancer patients

3.2. Feature Filtering

Performing feature screening can reduce the number of unnecessary factors and improve model interpretation and computational efficiency while improving model prediction performance and reducing the risk of overfitting. In practical applications, the dataset may contain a large number of redundant or irrelevant factors, which may affect the model performance and even lead to overfitting problems. By performing feature filtering, these unnecessary features can be eliminated and the efficiency and accuracy of the model can be improved. Secondly, feature filtering can also help understand the degree of influence of each factor in the dataset on the model and further dig deeper into the inherent patterns and characteristics of the dataset. In addition, most of the traditional survival analysis methods are based on certain assumptions, which require that there is no excessive sparsity among the factors of the data. If all factors of all data are introduced into the model, thus leading to instability of the traditional survival analysis model. Therefore, the features

need to be filtered to ensure the stability and computational efficiency of the model. Therefore, this experiment will use Log-rank test, KM method and one-way CPH analysis to test the significance of the features before modeling, and the features will be screened according to the combined results.

In clinical statistics, Log-rank, KM method and CPH method are the most widely used survival analysis methods, which can provide a more comprehensive understanding of patient survival data. With the application of machine learning survival analysis methods in model building, these methods are increasingly applied in the feature screening stage before model building, which in turn lays the foundation for building survival prediction models.

The original dataset of breast cancer patients contains 22 clinical influences, and the total number of features after one-hot coding is 50. Because the efficiency of the survival analysis model becomes low when the number of features is too large, and in order to identify the important influencing factors, the features were screened using the log-rank method, KM method, and one-way CPH method for feature

significance testing.

First, the significance of serial and categorical characteristics was tested using the log-rank method. A significance level of 0.05 was chosen as the threshold. When $p \leq 0.05$, the feature is retained, otherwise, the feature is removed.

Next, KM curves were drawn based on the effects of serial-type features that passed the Log-rank test and category-type

features that had only two cases of taking values on the survival time of patients, as shown in Figure 2, which can visually demonstrate the effects of each feature on the survival of patients when taking different values. Through Figure 4.3, it is obvious that the greater the extent of tumor stage spread of breast cancer, the lower the overall survival rate of patients; the more advanced the tumor stage of breast cancer, the less optimistic the overall survival of patients.

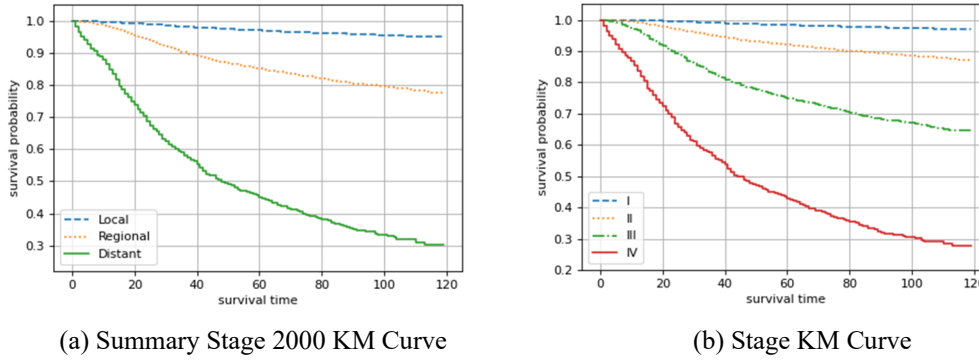


Fig 2. Survival KM curve of breast cancer influencing factors

Table 1. Results of feature selection

features	p value of Log-rank	p value of CPH	features	p value of Log-rank	p value of CPH
Age	—	0.0000	Histologic Type ICD-O-3_ILC	0.6706	—
T	0.0000	0.0000	Histologic Type ICD-O-3_other	0.5571	—
N	0.0000	0.0000	Laterality_Right	0.0703	—
M	0.0000	0.0000	Surg Prim Site_Modified	0.0000	0.0000
Stage	0.0000	0.0000	Surg Prim Site_Partial	0.0000	0.0000
Summary stage 2000	0.0000	0.0000	Surg Prim Site_Radical	0.0000	0.0001
Income	0.0000	0.0000	Surg Prim Site_Subcut	0.6532	—
Grade	0.0000	0.0000	Surg Prim Site_Total	0.5776	—
Radiation recodes	0.0000	0.0000	Breast Subtype_HR+HER2-	0.0000	0.0000
Chemotherapy recodes	0.0000	0.0000	Breast Subtype_HR-HER2+	0.0000	0.0000
ER	0.0000	0.0000	Breast Subtype_HR-HER2-	0.0000	0.0000
PR	0.0000	0.0000	RX Summ--Systemic/Sur Seq_after	0.0000	0.0000
HER2	0.0006	0.0142	RX Summ--Systemic/Sur Seq_before	0.0000	0.0000
Race_Asian	0.0000	0.0003	RX Summ--Systemic/Sur Seq_between	0.0000	0.0000
Race_Black	0.0000	0.0000	RX Summ--Systemic/Sur Seq_both	0.0000	0.0000
Race_White	0.0000	0.0000	RX Summ--Systemic/Sur Seq_none	0.0001	0.0042
Primary Site_501	0.0018	0.0253	Marital status_Married	0.0000	0.0000
Primary Site_502	0.0000	0.0000	Marital status_Partner	0.3395	—
Primary Site_503	0.5225	—	Marital status_Separate	0.0003	0.0105
Primary Site_504	0.0000	0.0002	Marital status_Single	0.0000	0.0000
Primary Site_505	0.2147	—	Marital status_Widowed	0.0000	0.0000
Primary Site_506	0.8832	—	Area_city	0.5160	—
Primary Site_508	0.0894	—	Area_medium-sized	0.0070	0.0525
Primary Site_509	0.0000	0.0000	Area_metropolitan	0.5677	—
Histologic Type ICD-O-3_IDC+ILC	0.1981	—	Area_not adjacent	0.0000	0.0013

Finally, a one-factor CPH model was established to analyze whether each feature had a significant effect on patient survival, and the insignificant features were eliminated, and the screened features were used as the final features for establishing the breast cancer survival model. One-factor

CPH models were built and p-values were checked according to each feature separately. To screen the features related to survival time to build the survival analysis model, a significance level of 0.05 was selected as the threshold in this paper. When $p \leq 0.05$, the feature was retained, otherwise, the

feature was removed.

Table 1 shows the p-values of each feature tested by Log-rank test and one-way CPH method. p-values greater than 0.05 were not significant and were removed, and the remaining features were 36 after removal.

3.3. Model Training and Validation

Training and validation of the arbitrary moment survival prediction model for breast cancer patients using processed data are done in python language for all experiments in this chapter. The processed breast cancer data were used to train the survival analysis model, and the breast cancer patient data were divided into training and testing sets for ten-fold cross-validation to fully ensure the training and testing effects of the model and prevent the overfitting phenomenon. The data were divided into ten equal parts, and nine of them were

selected each time as the training set to train the model and one as the test set for model validation.

After dividing the training set and test set data, survival analysis was used to build the model on the training set, and the prediction performance of the model was evaluated using the test set. In this study, four survival analysis methods, CPH, Cox-ElasticNet, DeepSurv, and RSF, were used to predict the survival rate and relative risk of breast cancer patients at any time, and the prediction performance of each model was evaluated comprehensively from all aspects to determine an optimal prediction model. In the process of training the model, the setting of model parameters has a crucial influence on the training effect and the final prediction ability of the model. In this study, the grid tuning method was used to search the important parameters of the four models, and some relevant parameters were obtained as shown in Table 2.

Table 2. Models' Parameters

Model	Parameter	Parameter value
CPH	alpha	0.1
	n_iter	200
Cox-Elastic Net	l1_ratio	0.1
	alpha_min_ratio	0.01
DeepSurv	lr	0.001
	init_method	xav_uniform
RSF	estimator__max_depth	10
	estimator__min_samples_split	5
	estimator__min_samples_leaf	3

The magnitudes of the C-Index, BS and IBS metrics for each trained model in predicting the survival of breast cancer patients at any moment were obtained through experimental validation and the experimental results are shown in Table 3.

Table 3. Performance comparison of survival prediction models for breast cancer patients

Model	C-Index	BS	IBS
CPH	0.8472	0.0930	0.0530
Cox-Elastic Net	0.8446	0.0939	0.0536
DeepSurv	0.6582	0.1108	0.0678
RSF	0.8535	0.0853	0.0512

From the experimental results in Table 3, it can be seen that the RSF model has a C-Index of 0.8535, which is higher than other survival analysis methods, and BS and IBS of 0.0853 and 0.0512, respectively, which are the smallest values among all models. All metrics indicated the superiority of the RSF model in overall performance. For the four models constructed in this chapter, the model with the best discrimination in predicting the survival of breast cancer patients at any moment in time was the RSF model, followed by the CPH model. It can be seen that although CPH is a traditional statistical survival analysis model, it still has

considerable value in practical applications despite its many shortcomings. When the requirement for model prediction accuracy is not strict, the CPH method can still be used for patient survival exploration.

3.4. Analysis of Influencing Factors

Predicting the survival rate of breast cancer patients at any given moment is essentially a matter of building a survival function by learning known data on breast cancer patients, and then using the survival function to make predictions about the survival rate. Therefore, to study the factors influencing the survival rate of breast cancer patients at any moment is to study which factors play an important role in the construction of the survival function. In order to visualize the factors influencing the survival rate of breast cancer patients at any moment, the RSF model, which performs best in the survival rate prediction problem of breast cancer patients, is selected to calculate the SHAP values of the relevant features affecting the survival rate of breast cancer patients at any moment and analyze the influence of each feature on the survival rate of patients. Among them, the 16 features with the most important impact on the survival rate of breast cancer patients and their SHAP values are shown in Figure 3 and Figure 4.

According to Figure 3, overall, the tumor information of breast cancer patients remains the most important aspect that affects the survival rate of breast cancer patients. Specifically, tumor stage, tumor stage, TNM stage, histological grading, and relevant hormone receptor status have a greater impact on the survival rate of breast cancer patients. All of these relevant

characteristics symbolize the disease progression and severity of breast cancer patients, and are factors that are closely monitored in breast cancer treatment. In terms of treatment information, whether a breast cancer patient is treated with chemotherapy versus radiation therapy and whether the patient's surgical procedure is modified radical surgery play an important role in predicting the survival of breast cancer patients. Among the demographic influencing factors, the age of the patient and whether or not the patient is married have a greater impact on the survival rate of breast cancer patients.

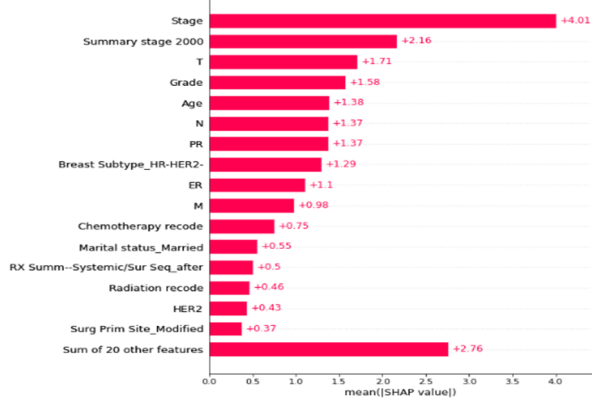


Fig 3. Important factors influencing on breast cancer patients

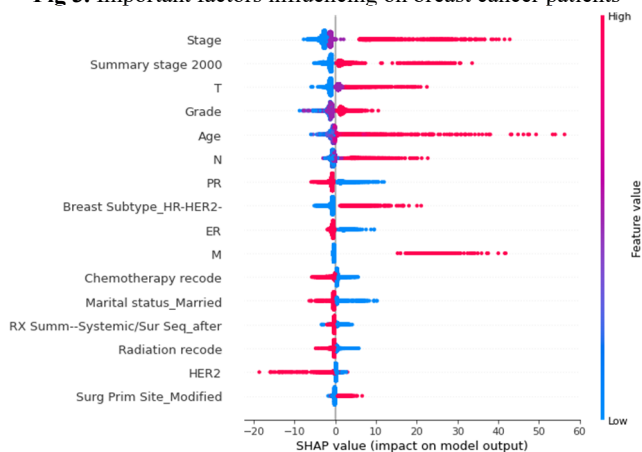


Fig 4. SHAP of influencing factors for breast cancer patients

The model established by the survival analysis method is a survival function about the survival rate of patients at any moment. According to the survival function, the survival risk of breast cancer patients during the survival period can be calculated, and the survival risk is inversely proportional to the survival time of patients, which can be used to compare the relative length of survival time between different patients. Figure 4 reflects the positive and negative influence of important factors on the survival risk of breast cancer patients, and the larger the SHAP value indicates the greater the contribution of the factor to the survival risk, i.e., the larger the SHAP value, the lower the survival rate. As shown in Figure 4, the more advanced the tumor stage and TNM stage and the higher the histological grade of breast cancer patients, the higher the survival risk of breast cancer patients. Patients with negative ER, PR, and HER2 for the relevant hormone receptors had lower survival rates. Patients who take chemotherapy are less likely to die and it is evident that chemotherapy is an aggressive measure for breast cancer patients to prolong their lives. Patients who do not undergo radiotherapy have a higher risk of survival because radiotherapy effectively kills tumor cells and reduces the risk of cancer recurrence. Survival risk is higher when the surgical

procedure taken for breast cancer patients is modified radical surgery, which is seen to be less effective. Among the demographic influences, the higher the age, the higher the risk of survival; the risk of survival is higher in unmarried breast cancer patients.

4. Summary

In this paper, survival prediction is performed for the entire survival cycle of breast cancer patients, and the data used are considering the survival problem of censored data at any given time point. First, survival analysis methods were adopted for feature screening. Feature significance tests were performed according to Log-rank and one-way CPH methods, and the survival curves of some features were depicted using the KM method, and the significant features were selected as the final features of the input model. Subsequently, four survival analysis methods were used to establish the survival rate prediction model, and through the comparative analysis of the experimental results of the four methods, it was found that the RSF model had significant advantages in predicting the survival rate of breast cancer patients at any moment. Finally, the factors influencing the survival rate of breast cancer patients based on the RSF model were explained and analyzed using the SHAP method. The results showed that tumor stage, TNM stage, grade and age have a great impact on the prognosis of breast cancer patients. The prediction model for survival rate of breast cancer patients at any moment constructed in this study has good performance and can assist in determining the prognostic survival of breast cancer patients and provide support for clinical prognostic evaluation.

References

- [1] World Health Organization. Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020. [EB/OL]. [2021-12-07]. <https://www.iarc.who.int/featured-news/latest-global-cancer-data-cancer-burden-rises-to-19-3-million-new-cases-and-10-0-million-cancer-deaths-in-2020/>.
- [2] Han Y, Wang J, Xu B. Clinicopathological characteristics and prognosis of breast cancer with special histological types: A surveillance, epidemiology, and end results database analysis[J]. Breast, 2020, 54(12): 114-120.
- [3] Lorona N C, Malone K E, Li C I. Racial/ethnic disparities in risk of breast cancer mortality by molecular subtype and stage at diagnosis[J]. Breast Cancer Research and Treatment, 2021, 190(3): 549-558.
- [4] Thio Q C B S, Karhade A V, Ogink P T, et al. Can Machine-learning Techniques Be Used for 5-year Survival Prediction of Patients With Chondrosarcoma? [J]. Clinical Orthopaedics and Related Research, 2018, 476(10): 2040-2048.
- [5] Liu R, Xiao Z S, Hu D X, et al. Cancer-Specific Survival Outcome in Early-Stage Young Breast Cancer: Evidence from the SEER Database Analysis[J]. Frontiers in Endocrinology, 2022, 12(1): 811878.
- [6] Han Y, Wang J, Sun Y, et al. Prognostic Model and Nomogram for Estimating Survival of Small Breast Cancer: A SEER-based Analysis[J]. Clin Breast Cancer, 2021, 21(5): e497-e505.
- [7] Zhang H W, Zhang N, Moran M S, et al. Special subtypes with favorable prognosis in breast cancer: A registry-based cohort study and network meta-analysis[J]. Cancer Treatment Reviews, 2020, 91(10): 1-10.

- [8] Kim B K, Oh S J, Song J Y, et al. Clinical Characteristics and Prognosis Associated with Multiple Primary Cancers in Breast Cancer Patients[J]. *J Breast Cancer*, 2018, 21(1): 62-69.
- [9] Pruitt S L, Zhu H, Heitjan D F, et al. Survival of women diagnosed with breast cancer and who have survived a previous cancer[J]. *Breast Cancer Research and Treatment*, 2021, 187(3): 853-865.
- [10] Fu M D, Chen Q T, Zeng L Y, et al. Prognosis Comparison Between Nipple-Sparing Mastectomy and Total Mastectomy in Breast Cancer: A Case-Control Study After Propensity Score Matching[J]. *Annals of Surgical Oncology*, 2021, 28(11): 1-10.
- [11] Wang S, Zhang Y Y, Yin F X, et al. Survival Outcomes After Breast-Conserving Therapy Compared with Mastectomy for Patients With Early-Stage Invasive Micropapillary Carcinoma of the Breast: A SEER Population-Based Study[J]. *Frontiers in Oncology*, 2021, 11(11): 1-11.
- [12] Zhang J, Yang C, Lei C, et al. Survival outcomes after breast-conserving therapy compared with mastectomy for patients with early-stage metaplastic breast cancer: a population-based study of 2412 patients[J]. *Breast*, 2021, 58(8): 10-17.
- [13] Wang S, Hu J, Zhang Y, et al. Presentation and survival by hormonal receptor status in metaplastic breast cancer: A propensity score-matched analysis[J]. *Breast*, 2021, 60(12): 168-176.
- [14] Li W, Xiao H, Xu X, et al. The Impact of Radiotherapy on the Incidence of Secondary Malignancies: A Pan-Cancer Study in the US SEER Cancer Registries[J]. *Curr Oncol*, 2021, 28(1): 301-316.
- [15] Kate R J, Nadig R. Stage-specific predictive models for breast cancer survivability[J]. *International Journal of Medical Informatics*, 2017, 97(1): 304-311.
- [16] Salehi M, Razmara J, Lotfi S. A Novel Data Mining on Breast Cancer Survivability Using MLP Ensemble Learners[J]. *Computer Journal*, 2020, 63(3): 435-447.
- [17] Sedighi- Maman Z, Mondello A. A two-stage modeling approach for breast cancer survivability prediction[J]. *International Journal of Medical Informatics*, 2021, 149(5): 104-111.
- [18] Kaur P, Singh A, Chana I. BSense: A parallel Bayesian hyperparameter optimized Stacked ensemble model for breast cancer survival prediction[J]. *Journal of Computational Science*, 2022, 60: 101570.