Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Modeling and Predicting the Survival of Breast Cancer Patients via Deep Neural Networks and Bayesian Algorithm

Soheila Rezaei[1], Hossein Ghayoumi Zadeh[1*], Mohammad Hossein Gholizadeh[1], Ali Fayazi[1], Khosro Rezaee[2]

1. Department of Electrical Engineering, Faculty of Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran
2. Department of Biomedical Engineering, Meybod University, Meybod, Iran.

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Nov 29, 2022
Accepted: May 13, 2023

Keywords:
Breast Neoplasms
Deep learning
Health Information Systems
Medical Informatics
Prognosis
Survival Analysis

ABSTRACT

Introduction: Due to breast cancer's prevalence as the second leading cause of cancer death worldwide, accurate survival prediction models are crucial. This study aimed to use an optimized deep neural network to predict breast cancer patient survival.

Material and Methods: The present study is an analytical study. The information utilized in this research is derived from the METABRIC database, associated with the molecular categorization of breast cancer patients, from the International Federation of Breast Cancer's Molecular Taxonomy Data. The total number of patients studied is 1981. Of these, 888 patients were under care until their death, and the remaining patients withdrew from the study during its course. In this database, 22 clinical features of patients have been considered, which includes a total of 6 quantitative features and 16 qualitative features. A deep neural network model called the optimized DeepHit is used to predict survival. The optimal parameters for specific variables of the neural network are obtained by the Bayesian algorithm.

Results: The optimized model has achieved the criterion of c_index = 0.748, which is a criterion for measuring the capability of survival analysis models.

Conclusion: The proposed model was compared with previous models using real and synthetic datasets. The results show that the optimized DeepHit achieved significantly better performance and statistically significant improvements over previous methods.

►*Please cite this article as:*
Rezaei S, Ghayoumi Zadeh H, Gholizadeh MH, Fayazi A, Rezaee Kh. Modeling and Predicting the Survival of Breast Cancer Patients via Deep Neural Networks and Bayesian Algorithm. Iran J Med Phys 2024; 21: 203-210. 10.22038/IJMP.2023.69096.2217.

Introduction

Cancer is when some of the body's cells grow irregularly and affect other parts [1 ,2]. Breast cancer is the leading type of cancer among women worldwide, with about 1 million new diagnoses annually, accounting for 18% of all female cancers [3]. In the UK, it has the highest mortality rate for this disease, with an incidence of nearly two per thousand annually among women aged 50. For women aged 40 to 50, it is the leading cause of death, responsible for 20% of deaths in this age group [3].

With advancements in screening, diagnosis, and treatment methods, the breast cancer death rate is expected to significantly decrease in the near future, especially in developed countries, leading to increased survival rates [4]. Despite being the second leading cause of cancer deaths among women, breast cancer has a high survival rate. Early detection and specialized interventions can reduce late-stage diagnoses, improve treatments, enhance survival rates, lower mortality, and ultimately improve patients' quality of life. Currently, 97% of women survive at least five years [5].

Breast cancer has no specific cause but results from a combination of factors. Decreased survival is associated with advanced-stage disease, old age, more involved lymph nodes, aggressive tumor progression, negative hormone receptors, high Her2neu expression, various treatments (surgery, radiotherapy, chemotherapy), and low socioeconomic status, including low education, poverty, and poor health[6] . Extensive research aims to discover prevention and treatment methods and effective drugs.

In recent years, molecular biomarkers have gained importance due to their high sensitivity and specificity[7]. Tumor markers, a key type of molecular biomarker, are involved in cancer development and progression, generated by tumors or normal cells in response to tumors. Advances in molecular biology and cancer bioinformatics have enhanced the understanding of cancer biology and the development of new prognostic and predictive methods[7]. The morphological parameters for distinguishing breast cancer include tumor size, tumor grade, and the positive or negative status of immune histochemical

*Corresponding Author: Tel: +983431312398; Email: h.ghayoumizadeh@vru.ac.ir

markers such as estrogen receptors, progesterone receptors, HER2, and Ki67 (a marker for cellular proliferation) markers. From this point of view, breast tumors are classified into five main subtypes with distinct clinical characteristics, i.e., Luminal A, Luminal B, HER2 Overexpression, Normal-Like, and Basal-Like. Luminal groups show distinct Estrogen Receptor (ER), Progesterone Receptor(PR), and HER2 markers [8 ,9].

Data analysis evaluates statistical data to uncover useful information, aiding in decision-making and predictions. Survival analysis, a branch of statistical methods, focuses on the expected duration until an event occurs. It aims to model the relationship between patient survival time and clinical characteristics to improve survival through effective treatments [10 ,11].

To investigate survival analysis, statistical models are categorized into three types: parametric, non-parametric, and quasi-parametric. Parametric models use distributions like log-logistic, Weibull, and exponential for survival time [12]. Non-parametric models, such as the Kaplan-Meier model, do not assume any distribution and estimate the survival function directly. The non-parametric life table model, an extension of Kaplan-Meier, is used for large datasets[13]. The Cox proportional hazard regression (CPH) model, proposed by Cox , assumes proportional hazards without specifying survival time distribution, using partial probabilities for parameter estimation[14,15] .

For complex, high-dimensional data with nonlinear interactions, more sophisticated models are needed. Machine learning models like the random survival forest (RSF) model handle such complexity by selecting influential variables on survival and managing high correlation features [16]. Artificial neural networks (ANNs) are also used for survival analysis due to their ability to learn nonlinear interactions between features [17]. Early ANN models were simple, with no hidden layers, focusing on predicting hazard functions [18, 19]. More advanced models, such as those by Street (1998) [20] and Yu et al. (2011) [21], incorporated multiple output nodes and combined regression models, respectively [22].

Recent developments include the DeepSurv model by Katzman et al. (2018) [23,24], which uses deep learning to handle nonlinear interactions, and the DeepHit model by Lee et al. (2018) [25], which learns survival time distributions without assumptions about the underlying stochastic process. This study uses an optimized version of DeepHit, incorporating structural layer changes and Bayesian optimization, resulting in significantly improved outcomes compared to previous versions.

Materials and Methods

The data used is from the METABRIC database related to molecular classification of breast cancer patients in the international consortium [26], which includes gene profiles and clinical features of patients (access number: https://ega-archive.org/studies/EGAS00000000083)[27]. The total number of patients studied is 1,981. Of these, 44.8% (888 people) were under care until death, and the remaining patients (1,093 people) were censored during the study and used as right-censored data. The information of this database are shown in Table 1 and 2. This dataset includes 22 clinical features of patients, consisting of 6 quantitative features such as patient age, tumor size, and the number of positive and removed lymph nodes, etc., and 16 qualitative features such as grade and site, etc. The sample distribution for each qualitative index related to each feature is shown in Table 2. For example, the grade feature, used to compare the shape and amount of difference between cancerous tissue cells and normal, healthy tissue, includes grades 0, 1, and 2. Of the 1981 available samples, 170 patients were classified as grade 0, 767 as grade 1, and 1044 as grade 2. The qualitative features were one-hot encoded for using this dataset, and missing data for quantitative features were replaced with the mean of the actual data. Additionally, the time of death occurrence and censoring time were recorded for patients in this database.

Table 1. Quantitative Characteristics of METABRIC data related to molecular classification of breast cancer patients

Data	Minimum	maximum	average
Age at diagnosis	21	92	60.64
size	0	182	25.39
Lymph nodes positive	0	45	1.8
Lymph nodes removed	0	48	12.82
NPI: Nottingham Prognostic Index	1	6.3	3.99
stage	0	4	1.19

Table 2. Qualitative Characteristics of METABRIC data related to molecular classification of breast cancer patients

Features	Number Of Samples based on Quality index											
	0	1	2	3	4	5	6	7	8	9	10	11
Her2 Expr	234	1747	-	-	-	-	-	-	-	-	-	-
ER Expr	1545	436	-	-	-	-	-	-	-	-	-	-
PR Expz	1091	890	-	-	-	-	-	-	-	-	-	-
ER IHC status	404	1577	-	-	-	-	-	-	-	-	-	
Inf men status	1502	479	-	-	-	-	-	-	-	-	-	-
cellularity	1052	216	713	-	-	-	-	-	-	-	-	
grade	170	767	1044	-	-	-	-	-	-	-	-	-
HER2 IHC status	1840	30	111	-	-	-	-	-	-	-	-	-
HER2 SNP6 state	418	97	1466	-	-	-	-	-	-	-	-	-
Genefu	1395	363	127	-	-	-	-	-	-	-	-	-
Treatment	48	31	152	160	407	660	299	224	-	-	-	-
group	773	433	237	531	7	-	-	-	-	-	-	-
site	570	270	733	237	171	-	-	-	-	-	-	-
Int clust memb	69	42	152	1212	82	42	105	139	61	77	-	-
Pam50 Subtype	303	228	776	470	6	198	-	-	-	-	-	-
histological	10	1559	89	28	46	65	144	9	2	12	12	5

- Her2 Expr: Human Epidermal Growth Factor Receptor 2 Expression
- ER Expr: Estrogen Receptor Expression
- PR Expz: Progesterone Receptor Expression
- ER IHC status: Estrogen Receptor Immunohistochemistry Status
- Inf men status: Information on Menopause Status
- Cellularity: Degree of Cellularity
- HER2 IHC status: Human Epidermal Growth Factor Receptor 2 Immunohistochemistry Status
- HER2 SNP6 state: Human Epidermal Growth Factor Receptor 2 Single Nucleotide Polymorphism 6 State
- Genefu: Gene Functional Status
- Treatment: Types of Treatments Received
- Group: Patient Group Classification
- Site: Site of Tumor
- Int clust memb: Integrated Cluster Membership
- Pam50 Subtype: PAM50 Subtype Classification
- Histological: Histological Type of Tumor

To use this database, changes must first be made to it. In this regard, normalization operations have been performed on the data associated with the features and the data with a correlation greater than 0.95 have been removed. Moreover, to improve the performance of the model, kernel principal component analysis (KPCA) method has been applied to the data associated with the features and as a result the numbers of dimensions of the features have been reduced to 15. The KPCA overcomes many of the limitations of the PCA linear method through nonlinear mapping of the input space to a higher-dimensional feature space. The linearity of KPCA in the feature space but its nonlinearity in the input space enables it to extract the low-dimensional features contained in high-dimensional statistical information [28].

The DeepHit model is a deep learning neural network based on the back propagation algorithm and is designed based on multi-task learning. The DeepHit makes no assumptions about the model and data. The model architecture is shown in Figure 1. As it can be seen, the network consists of two parts: The first part is associated with the shared sub network, which consists of a fully connected layer followed by a dropout layer. The second part is related to a group of cause-specific sub-networks, which consists of a fully connected layer for each event, followed by a dropout layer. Due to this structure, this model can easily be used for datasets with one or more competitive risks. This model has a residual connection between features (main input) and cause-specific sub-networks input. This implies that the input of cause-specific sub-networks includes the main input in addition to the output of the shared sub-network. This additional input allows cause-specific sub-networks to better learn the non-common features of the multiple causes. To assure the model learn the joint distribution of competitive events instead of the marginal distribution, a single softmax layer is used as the output layer of the model. Therefore, the model's output is a vector y for each instance in the data set with the feature x, which indicates the probability of experiencing the event k at time t [25].

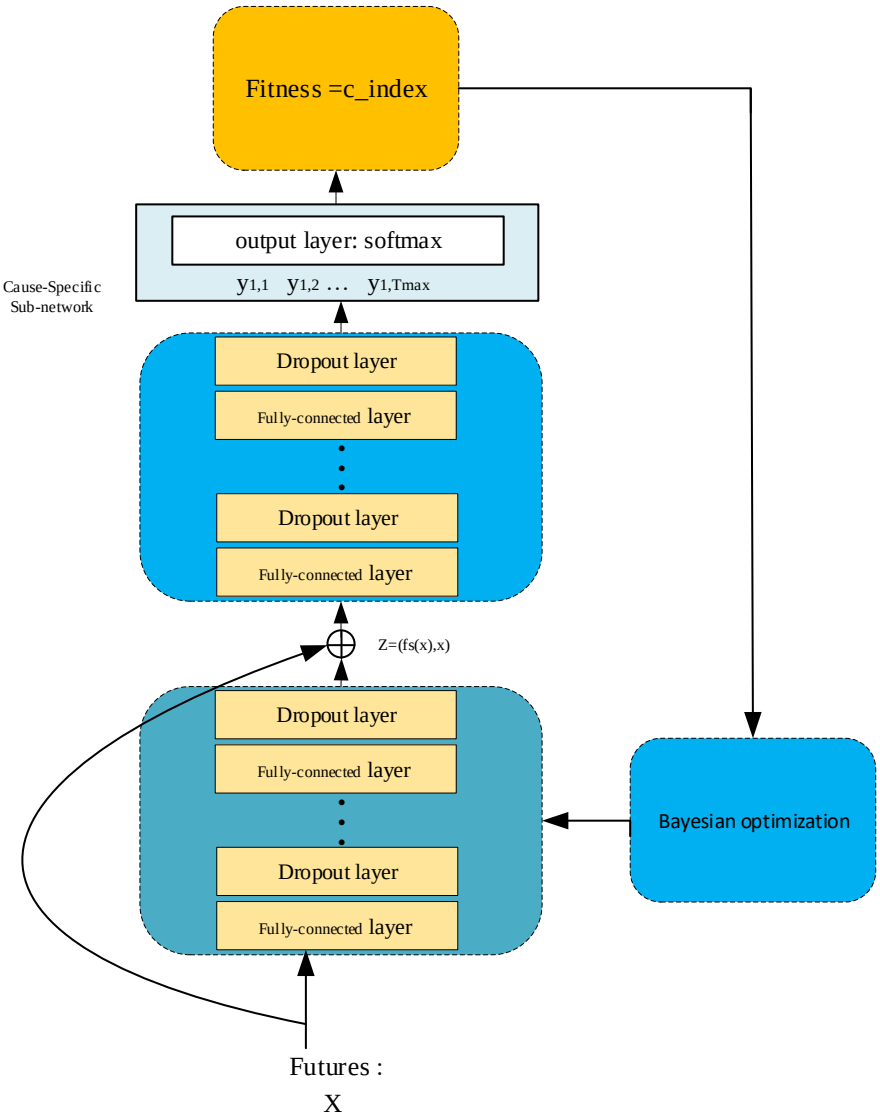

Figure 1. DeepHit deep neural network algorithm architecture

Each cause-specific sub-network utilizes the context vector (a fixed-length vector) along with the latest measurements, represented as a vector $z = (f_s(x), x))$, as inputs. The output generated is a vector $f_{ck}(z)$, which corresponds to the joint distribution of the initial occurrence time for a specific cause K. The context vector from the shared subnetwork serves as the input for these sub-networks, allowing them access to the learned common representation of the longitudinal history $f_s(x)$ while also enabling them to learn unique, non-common parts of the representation. If the sub-networks only use the common representation as input, the unique aspects of the representation will be neglected.

These outputs form a shared probability distribution for the initial collision and event. As a result, the cause-specific sub-networks concurrently learn the distribution of the first hitting time for each distinct cause.

The output from the softmax layer represents a probability distribution $y = [y_{1,1}, \cdots, y_{1,Tmax}, y_{K,1}, \cdots, y_{K,Tmax}]$. For a patient with covariates x, an output element $y_{K,s}$ denotes the probability $\hat{P}(s, k|x)$ that the patient will encounter event k at time s. This architecture enables the network to capture nonlinear and even disproportionate relationships between covariates and risks [25].

In this model, to train DeepHit, we have two loss functions, loss log likelihood and loss ranking, which are calculated according to the following equation [25]:

$$\mathcal{L}_{total} = \beta\mathcal{L}_1 + \mathcal{L}_2 \quad (1)$$

Where β is a weight constant, $\mathcal{L}_1$is the negative log-likelihood function which describes the joint distribution of the first hitting time and events, and $\mathcal{L}_2$ includes a combination of cause-specific ranking loss functions. The $\mathcal{L}_1$ function is modified to include censored data and competing risks. The log-likelihood function consists of two terms. The first part considers both the occurrence of an event and the occurrence time of an event for a subject who is not censored. The second part considers the time at which the patient is censored for a

subject who is censored, which is calculated as follows [25]:

$$\mathcal{L}_1 = -\sum_{i=1}^{N}[\mathbb{1}(k^{(i)} \neq \emptyset).log(y^{(i)}_{k^{(i)},s^{(i)}}) + \mathbb{1}(k^{(i)} = \emptyset).log(1 - \sum_{k=1}^{K}\hat{F}_k(s^{(i)}|X^{(i)}))] \quad (2)$$

In this equation, $\mathcal{L}_1$ represents the negative log-likelihood function for the joint distribution of the first hitting time and events. N denotes the total number of patients or instances in the dataset. The function $\mathbb{1}(\cdot)$ is an indicator function that returns 1 if the condition inside the parentheses is true and 0 otherwise. *k(i)* refers to the event type for the *i-th* patient (if the event is not censoring), while *s(i)* indicates the time at which the event (or censoring) occurred for the *i-th* patient. The symbol ∅ represents the censoring event. [25].

As aforementioned, $\mathcal{L}_2$ incorporates a combination of cause-specific ranking loss functions. Because this multi-task learning model requires cause-specific loss functions, which is the cumulative incidence function (CIF). This function expresses the probability that a particular event k occurs before or exactly at time t conditional on covariates x. To estimate CIF, the sum of the probabilities from the time of the first observation to the time of event k is calculated, which is obtained according to Eq. 3 [25].

$$\mathcal{L}_2 = \sum_{k=1}^{K}\alpha_k.\sum_{i\neq j}A_{k,i,j}.\eta(\hat{F}_k(s^{(i)}|X^{(i)}),\hat{F}_k(s^{(i)}|X^{(j)}) \quad (3)$$

Where the coefficients α_k are selected to adjust the balance of the ranking losses of the *k*-th competing event. $\eta(x,y)$ represents a convex loss function. It should be noted that for convenience, here the coefficients α are all assumed to be equal (i.e. $\alpha_k = \alpha$ for $k = 1,\dots,K$) and the function $\eta(x,y) = exp\left(\frac{-(x-y)}{\sigma}\right)$ is used as a loss function [25]. A ranking loss function ($A_{k,i,j}$) is also utilized which complies with the idea of concordance: a patient who dies at time s must have a higher risk at time s than a patient who survived longer than s, which is defined as follows[25]:

$$A_{k,i,j} \triangleq \mathbb{1}(k^{(i)} = k, s^{(i)} < s^{(j)}) \quad (4)$$

Although in the previous version of DeepHit, the authors tried to put the most optimal parameters in the model and relationships, it certainly was not the most optimal [25].

Efficient hyper-parameter optimization algorithms are crucial for optimizing machine learning and deep learning methods. This work uses the Bayesian Optimization (BO) algorithm, which directs the search to find the maximum or minimum of an objective function using Bayesian theorem[26]. The BO approach is widely used for hyper-parameter tuning in complex objective functions of machine learning and deep learning models.

The BO algorithm has two key components: a surrogate statistical model and an acquisition function. The surrogate model approximates the objective function, while the acquisition function proposes sampling points in the search space and scores the utility of evaluating candidate inputs using the surrogate model, such as expected improvement [29].

The surrogate statistical model provides a prior distribution over the objective function at any candidate point which is chosen uniformly at random and updates the prior distribution with samples taken from the objective function to attain a suitable posterior that better approximates the objective function. A popular surrogate model for Bayesian optimization algorithm is Gaussian processes prior which is used in Table 3[29].

As mentioned, fine-tuning the parameters of deep learning models can greatly impact performance. The proposed model has 4 important parameters to adjust, shown in Table 4.

The c_index criterion is used to evaluate the obtained parameters by Bayesian algorithm. This criterion demonstrates the model's ability to provide a reliable rating of survival times based on sample risk scores. In this evaluation, the model identifies the total number of pairs and then concordant pairs (pairs for which the actual time of their event and the time predicted by the model are the same). The probability of correctly predicting the model is investigated by dividing the concordant pairs by the total number of pairs [25].

Table 3. Basic pseudo-code for Bayesian optimization

Initialization:
- Put a Gaussian process as a prior distribution over function *f*

Process:
- Evaluate *f* at *n*0 points in accordance with an initial space-filling experimental scheme
- Let n = n0.

while $n \leq N$ do
- Calculate the posterior probability distribution on *f* by updating the prior probability
- Identify x_n as a maximizer of the acquisition function (the current posterior distribution calculates the acquisition function)
- Evaluate $y_n = f(x_n)$.
- Increment *n*

end while

Return:
- Return a solution based on either the point corresponding to the largest f(x) or the point corresponding to the largest posterior mean.

Table 4. Parameters in Bayesian optimization algorithm

Parameters description	variation range
Training learning rate	Lr ϵ [0.0001,0.001]
number of fully connected layers	N_{fc} ∈ [40,120]
Alpha coefficient in loss function	α ∈ [1,4]
Beta coefficient in loss function	β ∈ [1,4]

As mentioned, the time-dependent coordination index (C^{td}-index) was used as a performance criterion [24]. It should be noted that the typical C-index is a distinct index which is mainly based on the assumption that patients who live longer should have a lower risk than patients who live shorter. The C^{td}-index for the event k is defined as follows [25]:

$$C^{td} = p\left(\hat{F}_k\left(s^{(i)}|X^{(i)}\right) > \hat{F}_k\left(s^{(i)}|X^{(j)}\right)|s^{(i)}|s^{(j)}\right) \approx \frac{\sum_{i\neq j} A_{k,i,j}.\mathbb{1}\left(\hat{F}_k\left(s^{(i)}|X^{(i)}\right) > \hat{F}_k\left(s^{(i)}|X^{(j)}\right)\right)}{\sum_{i\neq j} A_{k,i,j}} \quad (5)$$

In this context, $A_{k,i,j}$ serves as an index function for a pair (i,j) deemed appropriate for event K, with its approximation derived from experimental definitions. Consequently, the C^{td}-index for event K is calculated by comparing pairs where one patient experienced event K at a specific time, whereas the other patient did not experience any event and was not censored at that particular time.

The programming of the proposed model was carried out using TensorFlow 2 in the Google Colab environment. To run the algorithm, the data set is divided into two categories: the training and test data sets, with a ratio of 80/20. It should be noted that by default 15% of 80% of the training data is considered as validation data in programming.

Results

The parameters tuned by means of the Bayesian algorithm are given in Table 5. The results for the C^{td}-index are shown in Table 6, which is compared to the DeepHit neural network in past works and other machine learning algorithms.

Because the early stopping method is used to avoid overfitting, the learning process is stopped at the epoch to 120. The loss curve is shown in figure 2.

Table 5. Specifications of tuned parameters to run the optimized Deep Hit deep neural network

Parameter	Size Batch	Learning Rate	epochs	**α**	**β**	Number of Fully connection layers
Value	1024	0.000659	500	1.312	2.716	67

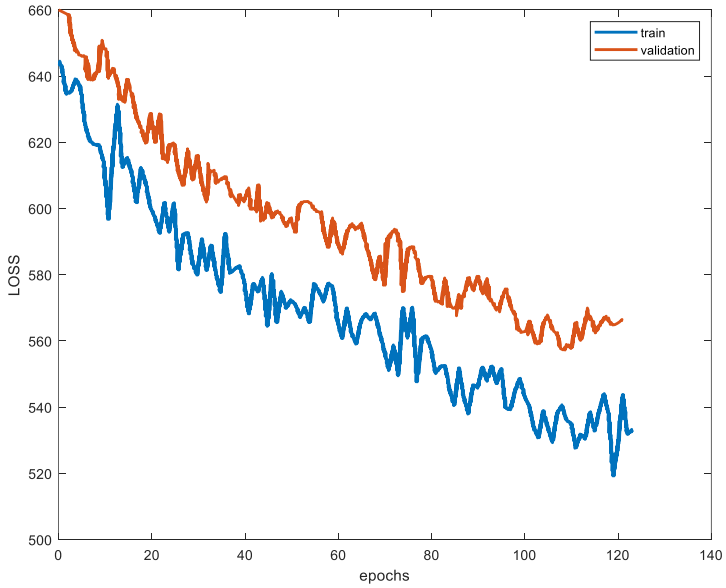

Figure 2. The value of the loss function in the learning process

Table 6. Comparison of the results corresponds to the C^{td}-index in the proposed optimized model

Algorithm	Cox	RSF	ThresReg	MP-RForest	MP-AdaBoost	MP-LogitR	DeepSurv	DeepHit [25]	This work
C^{td}-index values	0.648	0.672	0.649	0.650	0.633	0.661	0.648	0.691	0.748

Discussion

Breast cancer is an important epidemiological issue with global spread and is one of the most important causes of death and, in fact, the second leading cause of death due to cancer in women. The use of novel methods based on artificial intelligence has increased significantly in recent studies and research due to the ability to reduce the complexity of model relationships and increase model learning with better detection and high accuracy in prediction. The present study proposed an optimized DeepHit deep neural network model. As can be seen from previous studies [25], the survival rate is one of the most important indicators that help the medical community provide an appropriate diagnostic and therapeutic method through prognostic estimation of the disease.

Data mining is able to discover and extract new valuable knowledge from retrospective data. The method of data processing and selected variables significantly affect knowledge discovery that KPCA was used in the proposed model. The proposed optimized model has changes such as increasing the number of fully connected layers to size 67, changing the batch size to 1024, using the KPCA model and setting the optimal parameters α and β, which has significantly increased the C^{td} results as compared to previous works. Since the C^{td} discriminative index is not reliant on a single fixed time point, it offers a suitable evaluation for scenarios where the impact of covariates on survival varies over time, indicating non-proportional hazards across different time periods [25].

In the context of METABRIC datasets, characterized by a single event (risk), the performance of the proposed optimized DeepHit model was distinctly compared with two other survival models, DeepHit and DeepSurv. Various families of survival models, developed for predicting mortality using machine learning algorithms, were evaluated. These models include random forests (MP-RForest), logistic regression (MP-LogitR), AdaBoost algorithm (MP-AdaBoost), and a deep neural network (DeepSurv) that is based on the Cox proportional hazards assumption [30]. As demonstrated in Table 5, the results indicate that the enhanced DeepHit model outperforms the other models.

The significant improvement in performance can be attributed to the advanced techniques employed, such as KPCA for feature reduction and Bayesian optimization for parameter tuning. These enhancements allowed the model to capture more relevant patterns and relationships within the data, leading to more accurate survival predictions.

Conclusion

Survival analysis is a valuable tool in clinical research for evaluating treatments, disease control, and prognosis. Artificial neural networks, effective for pattern recognition and clinical prediction, are used to investigate nonlinear relationships and complex interactions. This article presents an optimized DeepHit method for analyzing survival data. The neural network, trained by DeepHit, accurately learns the joint distribution of time and events, improving performance over previous DeepHit models.

References

1. El-Bendary N, Belal NA. A feature-fusion framework of clinical, genomics, and histopathological data for METABRIC breast cancer subtype classification. Applied Soft Computing. 2020 Jun 1;91:106238.
2. Singh D, Singh AK. Role of image thermography in early breast cancer detection-Past, present and future. Computer methods and programs in biomedicine. 2020 Jan 1;183:105074.
3. McPherson K, Steel C, Dixon JM. Breast cancer-epidemiology, risk factors, and genetics. BMJ. 2000;321(7261):624-8.
4. Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. The lancet. 2011; 378(9801):1461-84.
5. Delen, Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine. 2005;34(2):113-27
6. Campone M, Fumoleau P, Bourbouloux E, Kerbrat P, Roché H. Taxanes in adjuvant breast cancer setting: which standard in Europe? Critical reviews in oncology/hematology. 2005; 55(3):167-75.
7. hammadpour A, Jahangirian E, Moharrami T, Rad GG, Sheikhani LJ, Taghizadeh S. Breast Cancer, Genetic Factors and Methods of Diagnosis. Sarem Journal of Reproductive Medicine. 2020;4(4):198-207.
8. Mukherjee A, Russell R, Chin SF, Liu B, Rueda OM, Ali HR, et al. Associations between genomic stratification of breast cancer and centrally reviewed tumour pathology in the METABRIC cohort. NPJ breast cancer. 2018;4(1):1-9.
9. Rakha EA, Green AR. Molecular classification of breast cancer: what the pathologist needs to know. Pathology. 2017;49(2):111-9.
10. Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M. Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. BMC medical genomics. 2019; 12(10):1-13.
11. Collett D. Modelling survival data in medical research. 2015: CRC press.
12. Stevenson M, EpiCentre IV. An introduction to survival analysis. EpiCentre, IVABS, Massey University. 2009 Jun 4.
13. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. International journal of Ayurveda research. 2010;1(4):274.
14. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annual review of public health. 1999; 20(1):145-57.
15. Therneau TM, Grambsch PM, Therneau TM, Grambsch PM. The cox model. Springer New York; 2000.
16. O'Brien RC, Ishwaran H, Szczotka-Flynn LB, Lass JH. Random survival forests analysis of intraoperative complications as predictors of

descemet stripping automated endothelial keratoplasty graft failure in the cornea preservation time study. JAMA ophthalmology. 2021 Feb 1;139(2):191-7.
17. Pickett KL, Suresh K, Campbell KR, Davis S, Juarez-Colunga E. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. BMC medical research methodology. 2021 Dec;21:1-4.
18. Liestbl K, Andersen PK, Andersen U.Survival analysis and neural nets. Statistics in medicine. 1994; 13(12):1189-200.
19. Faraggi D, Simon R. A neural network model for survival data. Statistics in medicine. 1995; 14(1): 73-82.
20. Street WN. A Neural Network Model for Prognostic Prediction. InICML.1998; 540-6.
21. Yu CN, Greiner R, Lin HC, Baracos V. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. Advances in neural information processing systems. 2011.;24:1845-53.
22. Fotso S. Deep neural networks for survival analysis based on a multi-task framework. arXiv preprint arXiv:1801.05512, 2018.
23. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology. 2018;18(1):1-12.
24. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine. 1996;(4)15:361-87.
25. Lee C, Zame W, Yoon J, Van Der Schaar M. Deephit: A deep learning approach to survival analysis with competing risks. in Thirty-second AAAI conference on artificial intelligence. 2018.
26. A Tour of Survival Analysis, from Classical to Modern, Some Software Packages. Available from:: https://sites.google.com/view/chil-survival.
27. Nagpal C, Li X, Dubrawski A. Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. IEEE Journal of Biomedical and Health Informatics. 2021;25(8):3163-75.
28. Cuevas-Delgado P, Dudzik D, Miguel V, Lamas S, Barbas C. Data-dependent normalization strategies for untargeted metabolomics—A case study. Analytical and Bioanalytical Chemistry. 2020 Sep;412:6391-405.
29. Frazier PI. A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811. 2018.
30. Lee ML, Whitmore GA. Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. Statistical Science. 2006; 21(4):501-13.