

Machine learning methods based on mammogram images to estimate survival times for breast cancer patients

Noor Ayad Mohammed, Entsar Arebe Fadam

Department of statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

ABSTRACT

Estimating survival times based on medical images resulting from radiological imaging of some parts of the body affected by tumors and making a predictive system for them is considered one of the very important fields at the present time. This is because radiological imaging is one of the first and most important stages of medical diagnosis. Therefore the process of linking medical images including them within the work steps of the statistical analysis for estimating survival times is a modern and important topic. Its importance is helping doctors and medical specialists to determine the influencing factors and risk percentage associated with survival for each patient based on the medical image of the affected part.

In this paper, the medical images extracted from the mammogram device for breast cancer patients in Iraq. These images were included in the machine learning method for estimating survival of patients. Based on two methods to extract features, The first one is the Fast independent component algorithm (Fast ICA algorithm) and the second one is Nonnegative Matrix Factorization (NMF algorithm). With two machine learning algorithms, The first method is Random survival forests algorithm and the second method is support vector machine algorithm SVM (SVM).

Through the application of the supervised machine learning method on mammogram images of patients with breast cancer, it was found that the best model for estimating survival according to the mean square error (MSE) and concordance index (C-Index) criterion is the model resulting from the use of the Fast ICA algorithm with the random survival forest algorithm compared with the other three models. Accordingly, it is recommended that interested medical agencies and institutions to adopt this model.

Keywords: Survival function, Feature extraction, Machine learning, Mammogram medical images, Breast cancer.

Corresponding Author:

Noor A. Mohammed

Department of statistics

College of Administration and Economics, University of Baghdad.

Baghdad, Iraq

E-mail: noor.mohammed1101@coadec.uobaghdad.edu.iq

1. Introduction

The process of building a predictive system for patient survival based on modern artificial intelligence methods represented by the machine learning method. This is one of the topics of great and increasing importance at the present time. Due to the advantages that this method possesses that made it superior to the common statistical methods. In estimating survival times, the learning method is accurate, easy to implement, and the ability to deal with big data represented by images. The images can be medically related to patients with tumors or other medical conditions that are diagnosed through radiography, such as scanners, mammograms, etc. Also it is possible that the images used are of space and satellite images. This method has the ability to deal with related variables that are difficult to deal with in traditional statistical models used in estimating survival.

The process of building the predictive system is based on selecting the best predictive model has been used to estimate survival time. Iraqi data set is a group of breast cancer patients who were diagnosed by oncologists at



the Medical City Hospital in Baghdad. Mammogram images for all of these patients. They are numbered 100 patients included within the method of machine learning to estimate their survival.

The application of the machine learning method to estimate survival involves several stages. The first important stage is the availability of sufficient data on the subject of the study. which is represented here by mammogram images for breast cancer patients in Iraq. The second stage includes extracting important features from mammogram images for patients. Which was done based on two algorithms prepared for this Purpose namely faster ICA algorithm and MNF algorithm. The third stage includes the introduction of the features extracted from the second stage into the machine learning algorithms, which are the random survival forest algorithm and the SVM algorithm.

As for the fourth and final stage, it includes the comparison criteria that are used to compare the models used in this research paper.

The topic of using the machine learning method to estimate survival started around the year 2000 when Zupan and et al. [21] presented a research paper dealt with an estimate of survival of the disease returning to prostate cancer patients. The advantages of this modern method were to overcome the normal statistical methods in dealing with the nonlinear and censored data, and dealing with such problems without the need to know the statistical distribution of the data. The research also dealt with the use of the decision-making tree algorithm, and Bayes Classifier showed much better results compared to the cox model for survival.

In 2017, Leger, et al. [15] dealt with a comparison of machine learning algorithms with downsizing algorithms for quantitative imaging data modeling to characterize tumor phenotype and predict clinical outcomes. Here 11 machine learning algorithms with 12 feature extraction algorithms to predict tumor size for affected patients squamous cell carcinoma of the head and neck, thus predicting survival.

Lee and Lim [14] presented in 2019 a research on statistical methods for analyzing survival function using genomic data. Many traditional statistical methods were used to analyze survival for controlled observations and because the genomic data is high-dimensional, more advanced methods were addressed to build a predictive survival model with high data dimensions. In addition, machine learning methods have been adapted to analyze survival in a way that suits the effects of the non-linear and complex interaction between predictive variables and to ensure an accurate prediction of the probability of survival.

In the year 2020, Spooner, et al. [20] compared a number of machine learning methods for survival analysis of high-dimensional medical data for the purpose of predicting dementia. Ten machine learning algorithms were used with eight methods to select important variables in order to overcome the difficulties associated with the analysis Survival of a patient until he entered the state of dementia. Generally, data suffer from high dimension, censored data, and the state of heterogeneity. The use of machine learning models to predict the patient's entry into the state of dementia is essential.

In the same year, Kantidakis et al. [11] presented a research paper related to the topic of predicting survival in liver transplantation based on statistical models, Cox models, and machine learning techniques. Here random survival forests and two types of neural networks were used and applied to a large sample size, and large variables affecting survival were studied. The results of machine learning techniques were compared with three different Cox models, and the results showed the superiority of neural networks over random survival forests and all Cox models according to the comparison criteria used.

In 2021, Torres et al. [17] presented a paper showing the superiority of interpretable machine learning techniques over Cox regression model for predicting survival for breast cancer patients. Torres et al. proved this adapted three methods of machine learning techniques: Random Survival Forests, The Support Vector Machines, and the Extreme Gradient Boosting [XGB].

In 2022, Huang et al. [7] dealt with research on machine learning models for predicting survival for patients with ampullary adenocarcinoma (ampuloid adenocarcinoma), as the data set for the research included 2935 patients, and the data was divided into 70% to be the training data and 30% to be Test data. Six methods of estimating survival were queried : The American Joint Committee on Cancer TNM stage, Cox Proportional Hazards regression, CoxTime, DeepSurv, XGBoost Survival Embeddings, and Random Survival Forests.

Through the use of the compatibility index, it became clear that the DeepSurv method was the best in estimating survival, as it gave the highest value to it.

In 2023, Hatano et al. [5] presented a paper dealing with machine learning methods for predicting the probability of age to spinocerebellar ataxia (SCA)3 and dentatorubral-pallidoluysian atrophy (DRPLA) by survival curve analysis. The results of two machine learning methods were compared, which are random survival forest [RSF] and Deep Surv. With six parametric methods of survival analysis, which are (Weibull, exponential, Gaussian, logistic, loglogistic, and log Gaussian). Through the use of three criteria for comparison between the methods, namely root mean squared error (RMSE), mean absolute error (MAE), and the integrated Brier score. It turned out that the machine learning methods were the best in estimating survival, especially the method Deep Surv.

2. Methods

2.1. The research dataset

Survival analysis is concerned with the analysis of survival data [18], which means the time between the beginning of the event until the occurrence of the end event represented by death or any other end event, which is subject to a set of strict conditions, including that it has a specific statistical distribution, or some cases do not reach the end event and are called censored data, or The survival times are dependent on a large number of variables, so the machine learning techniques has the ability to deal with these problems [19].

Before starting to clarify the theoretical aspect of the machine learning method, it is necessary to address the first and most important steps of this method, which is the data of the studied case.

As the data for this research paper represent the medical images extracted from the mammography device and touched by the medical mammogram device for patients with breast cancer.

The images resulting from imaging with this device include four images (anterior and lateral image of the right breast and anterior and lateral image of the left breast).With regard to the subject of the research, we have dealt with only two images, the front and the side, of the breast affected by tumors.

A group of 100 patients diagnosed with this disease was monitored in the Teaching Oncology Hospital in the Medical City, Baghdad, during a period of one year and eight months. During this period, 12 patients died, while 88 remained within the control condition.

Note that the resulting images are gray scale images, which are a type of digital image whose elements take on gray gradations [2].

2.2 Feature extraction algorithms (extract important features)

Many real data at the moment have many problems. Among them is the problem of big data, which must have methods to deal with these, including methods of extracting features .Here, some of these methods have been used [1].

2.2.1. The fast independent component analysis algorithm (Fast ICA)

The Fast independent component analysis algorithm (Fast ICA) is known as one of the most important and used ICA algorithms, It is first presented by Hyvärinen in 2000. before starting to explain it must be explained and clarified the ICA algorithm or what is called the algorithm of the analysis of independent component, which is one of the statistical methods and its idea depends on making the variables involved in the study studied independent of each other, i.e. converting a complex group and as large of correlated data to partial statistically independent component, ICA has multiple and wide applications in several fields and particular in digital signal processing, extracting important features, analysis of financial time series, digital image processing and other applications [3].

The model of the ICA algorithm can be expressed as follows:

$$X = AS + n \tag{1}$$

Where:

X : Represent a vector of observation.

A :That represents the mixing matrix.

S: Independent component vector.

n: Random errors vector.

The ICA algorithm includes many algorithms to achieve this goal and the Fast ICA algorithm is one of the most common and used algorithms for the estimation of independent component, as the idea of this algorithm depends on measure of the non-Gaussian, which is a measure or standard for independence, this algorithm is characterized by speed the high implementation of the typical function is a high stability [8]. Therefore, the steps of making a fast ICA algorithm can be explained as follows:

Step1: Enter the matrix of the image X , which has the dimension $N \times M$.

Step2: Starting with the steps of pre-processing the input matrix X , which is the centering, It is done by subtracting the average of each row of the matrix X from each value of this row according to the following formula

$$X_{ij} = X_{ij} - \frac{1}{M} \sum_j X_{ij} \quad (2)$$

Where $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$

Through this step, the expected value to each row will be equal to zero.

Step3: Whitening the data which is made each vector of the centering matrix are uncorrelated and have variance equal to 1. The whitening process takes place in several ways; One of them is including Eigen values decomposition on the variance and covariance matrix into the central data matrix, which is done according to the following formula:

$$\tilde{X} = E D^{-1/2} E^T X \quad (3)$$

Where:

\tilde{X} : Whitening matrix.

E : Orthogonal matrix of the Eigen vectors of $E(XX^T)$.

D : Diagonal matrix of the Eigen values of $E(XX^T)$.

Step4: Determine the value of C , which represents the number of independent components required, and in the event that it is one component, one weight vector is determined, as will be explained in the next step, and if it is more than one component, more than one weight vector is determined.

Step5: Randomly determine the weight vector W that maximizes the non-Gaussian scale to the projection $W^T X$, Which is updated according to the following formula

$$W^+ = W - [E\{Xg(W^T X)\} - \beta W] / [E\{g'(W^T X)\} - \beta] \quad (4)$$

Where:

$$\beta = E\{W^T X g(W^T X)\}$$

Step6: After determining the allowable convergence value, let it be ϵ . Then make the normalization to the weight vector as follow

$$|W_{K+1} - W_K| < \epsilon$$

If the algorithm achieves a value less than or equal to the amount of convergence ϵ , the last weights vector is adopted, otherwise we return to step 5.

Step7: Repeat the above steps C times, which represents the number of independent components required. It is usually less than the number of rows in the original matrix as follow

$$C \leq N.$$

2.2.2. Nonnegative matrix factorization (NMF) algorithm

The NMF algorithm is considered one of the important and useful algorithms in multivariate statistical analysis because it is used to reduce high dimensions and extract important features[13]. It was used for the first time by (Seung & Lee) in 1999. The basis for the work of this algorithm begins with the availability of data represented by non-negative matrices, as usually most of the real cases, including images, have the original matrix as a non-

negative matrix [20] and if the matrix includes negative values, there are many methods through which the matrix is transformed negative to a non-negative matrix.

The working steps of this algorithm can be explained according to the following steps[12]:

Step1: Provide the non-negative matrix for the studied case, let it be the V matrix with dimension $m \times n$ for the raw medical images after converting them into a digital matrix.

Step2: The existence of a constant called r so that it is less than the number of rows and less than the number of columns in the original matrix V, or that

$$(n + m)r < nm \quad (5)$$

Step3: Using the matrix V, a non-negative matrix W of dimension $m \times r$ is found, which represents the non-negative matrix H of dimension $r \times n$ so that

$$V \approx WH \quad (6)$$

Therefore, the dimensions of the factor matrices are much less than the dimensions of the original matrix, and this is the basis of the principle of the MNF algorithm.

Step4: To solve the convergence problem in the previous step, cost functions should be used, which aim to determine the quality of convergence, which is usually calculated according to some measures of the distance between the non-negative matrices of the original data V and the matrix resulting from multiplying the two non-negative matrices W and H. Among these measures:

1- The square of the Euclidean distance, which is expressed by the following formula

$$\|V - WH\|^2 \quad (7)$$

2- The measure of convergence, which can be called the Kullback-Leibler convergence, whose formula is given below

$$D(V \setminus WH) \quad (8)$$

In order to calculate the square of the Euclidean distance shown in formula 9, it is possible to use the method proven by Seung & Lee, as it used the rules of multiplicative updating to find W and H according to the iterative formulas below:

$$H^{n+1}_{[i,j]} \leftarrow H^n_{[i,j]} \frac{((W^n)^T V)_{[i,j]}}{((W^n)^T W^n H^n)_{[i,j]}} \quad (9)$$

$$W^{n+1}_{[i,j]} \leftarrow W^n_{[i,j]} \frac{(V(H^{n+1})^T)_{[i,j]}}{(W^n H^{n+1} (H^{n+1})^T)_{[i,j]}} \quad (10)$$

n and n+1 in the above two formulas refer to the previous and subsequent recurrence index and that this repetition stops when we reach the stable state to W and H.

2.3. The machine learning algorithms used to estimate the survival

2.3.1. Random survival forests algorithm (RSF)

It is one of the machine learning algorithms that are used for classification and regression. It is considered an expansion of the principle of survival trees. Instead of making a single survival tree according to the steps that were presented in the regression tree in the previous paragraph, several trees will be made, as the idea of random forests depends on finding more than one sample For the entered data group by using the Bootstrapping sampling and then making the steps of the regression tree for each group, and therefore the predictive value will be the computational mean of the predictive values resulting from each tree, the goal of this step is to refuse the random state and reduce the variance without increasing the bias in the data group and thus increasing Accuracy to predict survival times [9].

The steps of this algorithm can be displayed as follows:

Step1: Enter the data set for the study of survival times.

Step2: Use Bootstrap sampling to draw samples with a number of B from the collection of data entered according to the above step, as each drawn sample includes n observations, noting that the samples are withdrawn with the return randomly.

Step3: Making the regression tree for each sample that was found by using a bootstrap sampling according to the steps of the regression tree shown in the previous paragraph, and since each tree will give prediction to the survival time, therefore we will get B from predictions.

Step4: Calculation The mean of the predictions obtained according to the third step to be the predictive value of the new singular or new observation.

2.3.2. Support vector regression (SVR)

It is considered one of the common and frequently used supervised machine learning algorithms. It is used to solve classification problems. It is called the automatic support vector and is denoted by the symbol SVM, which was presented for the first time by Vapnik in 1995, and then he modified this algorithm in 1998 to be used for prediction and to be parallel with regression models, so it is called the regression vector [4]. Automatic and symbolized by the symbol SVR, the basic idea of the automatic support regression depends on finding a function that has the most deviation by a certain amount and let it be ϵ from the real and realistic goal Y for all training data and since the subject of research is related to predicting survival, so we will resort to clarifying the work of the automatic regression vector method To predict survival in the form of algorithmic steps, as shown below [6]:

Step1: Enter the data related to the subject of the study, which here represents the medical images after converting them into a digital matrix and extracting the important features from them, and then entering the output variable, which survival times. Thus, the data will be pairs to the X variable ($X \in R^d$) and the Y variable that represents the survival times.

Step2: converting data from the original two-dimensional space to another three-dimensional space for the purpose of getting rid of the problem that the data is non-linear, as most of the real applied data is not linear, and to solve this problem, resorting to the use of kernel functions to convert data to three-dimensional space Dimensions, and there are many kernel functions that can be used for this purpose, such as the linear, Gaussian, or polynomial function, noting that this step takes place in the case of classification or regression, and thus a linear model is built within the feature space.

Step3: Finding the so-called hyper level, whose aim in the case of classification is to separate the two sets of data, and in the case of regression, it is the line that is used to predict continuous results. Therefore, the linear model is given according to the following formula

$$f(x, w) = \sum_{i=1}^m w_i \phi_i(x) + b \quad (11)$$

Where b is the bias and W are the coefficients of the function f

$\phi_i(x)$: It represents a set of nonlinear transformations (one of the kernel functions) that transforms the original space into a three-dimensional space.

Step4: finding the decision limits, the aim of which in the case of classification is to place the values of the first group within one side and the values of the other group within another side. This is the same goal in the case of regression, and its mathematical formulas are given as follows:

$$\sum_{i=1}^m w_i \phi_i(x) + b = +a \quad (12)$$

$$\sum_{i=1}^m w_i \phi_i(x) + b = -a \quad (13)$$

Step5: The values of the training data will be distributed according to the hyper level line and the two decision boundary lines. Therefore, the values that will be closer to the hyper level and between the two decision

boundary lines are the best, which will give us the best model for estimating survival. Therefore, the estimated value to the new value \hat{Y} can be found by applying the neighbor algorithm the closest.

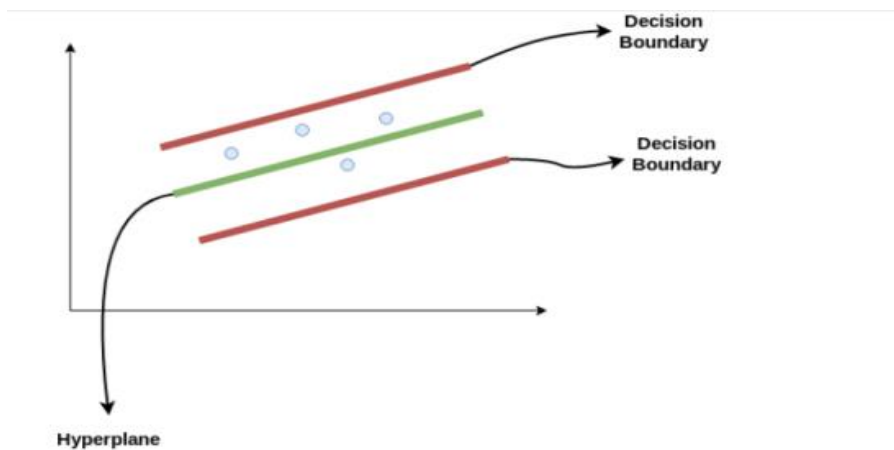


Figure1. Explain how the SVR algorithm works

3. Results and discussion

3.1. Stages of estimating survival time using machine learning algorithms

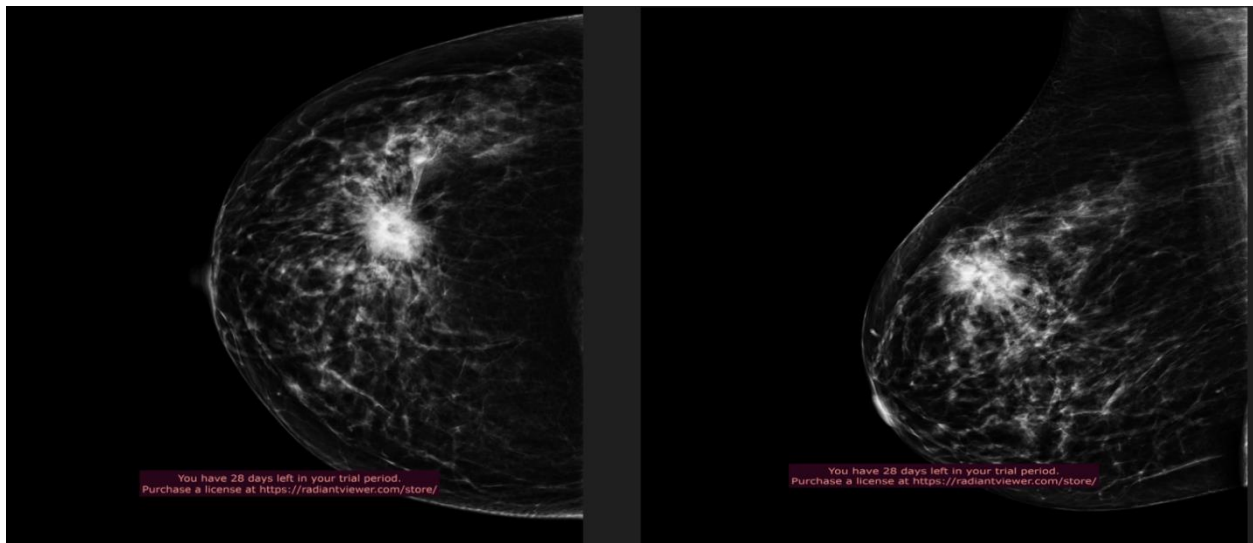
Survival estimation using machine learning methods includes many stages that can be explained as below. The first stage is preparation of data on the subject of the research. This stage includes many steps:

1- Since the machine learning method used in this research paper is supervised learning, therefore, the required final output must be available from the predicted model, and the final output here represents the survival times for 100 breast cancer patients whose survival times were monitored and were as in the table below.

Table1. The table shows the observed survival times (in months) for 100 patients over 18 months

	Survival time		Survival time		Survival time		Survival time
1	4	26	7	51	5	76	4
2	9	27	5	52	9	77	5
3	6	28	4	53	15	78	4
4	5	29	8	54	5	79	4
5	4	30	3	55	8	80	6
6	5	31	4	56	9	81	5
7	5	32	3	57	8	82	5
8	5	33	2	58	2	83	5
9	5	34	3	59	20	84	5
10	8	35	7	60	6	85	6
11	4	36	3	61	4	86	5
12	4	37	4	62	5	87	5
13	6	38	10	63	6	88	5
14	4	39	3	64	3	89	5
15	4	40	11	65	1	90	6
16	6	41	4	66	4	91	5
17	5	42	10	67	10	92	5
18	1	43	5	68	6	93	5
19	6	44	6	69	5	94	5
20	5	45	10	70	6	95	5
21	3	46	8	71	5	96	5
22	8	47	3	72	5	97	5
23	2	48	6	73	6	98	5
24	4	49	4	74	5	99	5
25	2	50	1	75	3	100	6

2- Read the file for the images or the so -called Path, any place to store the mammogram images in the JPG in the computer, which is being found by using the code called TNCER Flow, which is located within the Python program code, knowing that the images extracted for each patient have breast cancer are as in the form below.



(a)

(b)

Figure 2. The raw image of a breast cancer patient, as the image (a) represents the front image of the affected breast and the image (b) represents the side image of the affected breast.

- 3- Make a step to change the image size (Resize) for each image to be 1024 x 1024 for all algorithms within the work steps and the purpose of this step is to remove all the signs in the image and the unnecessary that are represented in the image above the writing at the bottom of the image.
- 4- Change the size of the image (RIZE) again for the necessary dimensions for each algorithm, as each algorithm of the features extraction (variables extraction) method works on the size of a different image from the other, and therefore this step is a kind of pre preparation for each image before inserting them into the algorithms to reduce dimensions or choose features.
- 5- Performing data Agmantation for each image, which means rotating the image at an angle of 90, 180, 270, so that for each image we have four images, and therefore each patient will have 8 images (because we dealt with two images for each patient, a front image and a side image of the tumour-infected breast).
- 6- Performing the Faltn for the image, which is means converting the image from a matrix to a vector consisting of n values, that is because of the algorithms for dimensions reduction or features extraction deal with data in the form of a vector, not a matrix.
- 7- Reshaping the two images (Reshape), the aim of which is to combine the values of the first vector resulting from the above step for the first image with the values of the second vector resulting from the above step for the second image, as the resulting vector has a dimension of 1×280 for each patient, but when it is approved for all patients It will be 280×70 in dimension.

The second stage/ (Feature extraction algorithms) extracting the important features from the image of the first stage, introducing vectors and matrices resulting from the above step into the dimension reduction algorithms (features extraction), as three algorithms were relied upon to extract features, namely:

- Fast ICA Algorithm
- NMF Algorithm

The third stage is Machine Learning algorithms. After extracting the important features according to the algorithms of the second step, these features are entered within the following machine learning algorithms

- Random forests Algorithm
- SVM Algorithm

Stage forth stage/ (comparison criteria for the models)

In this stage, a comparison is made between the combinations of feature extraction algorithms and machine learning algorithms to determine the best model in estimating survival according to two criteria.

- MSE
- C-Index

The results are shown according to the table below:

Table 2. Results of applying machine learning algorithms to real data about survival times based on MSE, C-Index

	MSE		C-Index	
	Fast ICA	NMF	Fast ICA	NMF
Random forests	0.006375	0.02021	0.85506	0.70137
SVM	0.011192	0.02862	0.78187	0.65991

Table 3. Results of applying machine learning algorithms to data using Kaplan-Meier survival method based on MSE, C-Index

	MSE		C-Index	
	Fast ICA	NMF	Fast ICA	NMF
Random forests	0.0236	0.07707	0.8928	0.7255
SVM	0.05694	0.10018	0.78187	0.6499

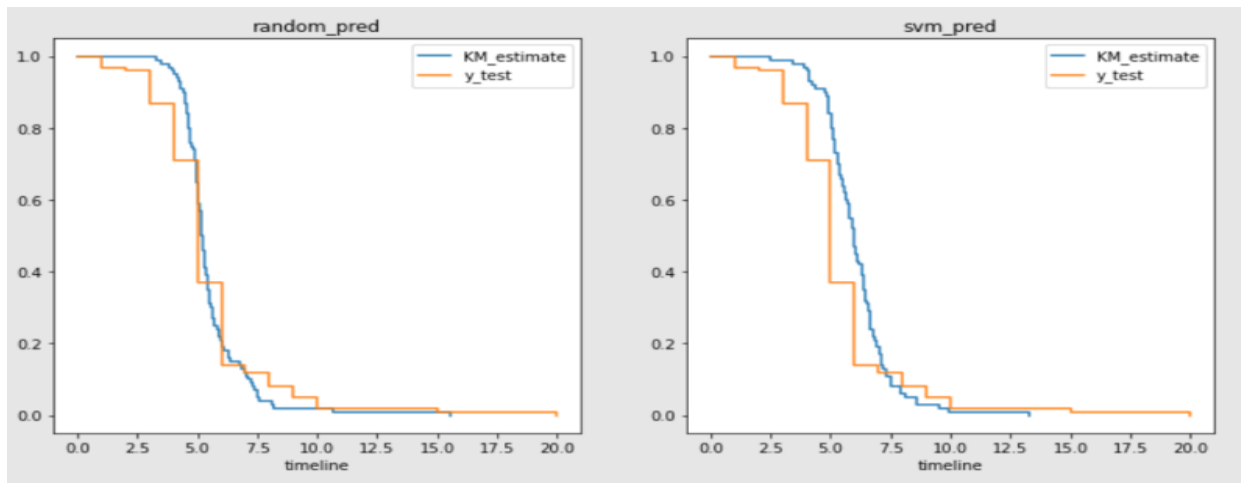


Figure 3. Fast ICA algorithm with random forests and SVM algorithms

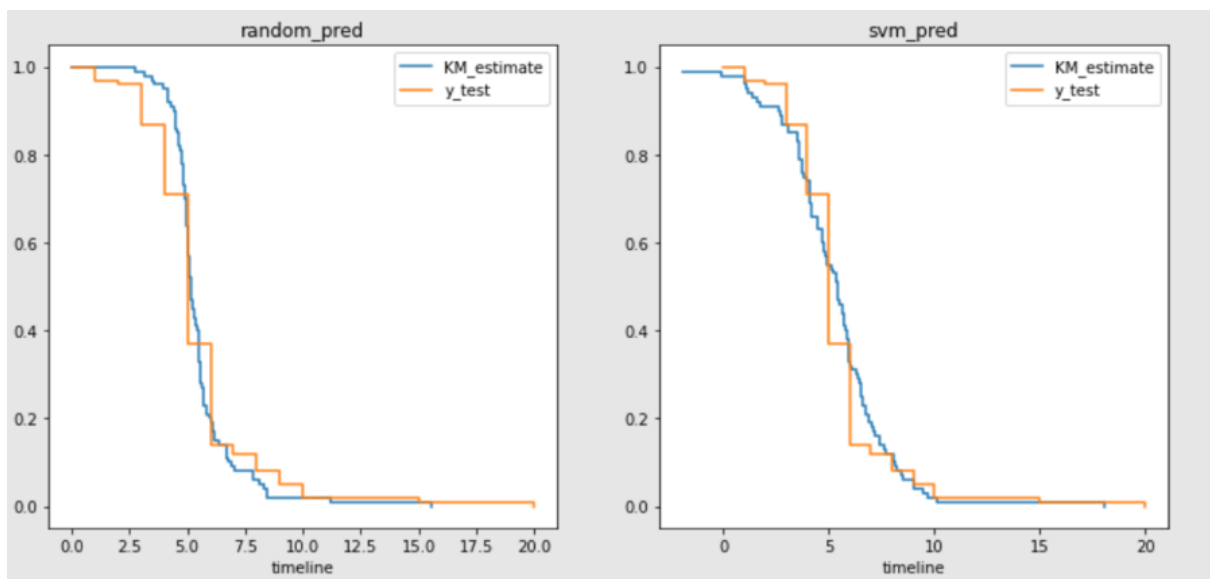


Figure 4. NMF algorithm with random forests and SVM algorithms

4. Conclusion

After completing the stages of estimating survival times and building a predictive system for them using machine learning methods based on real Iraqi data on mammogram images for breast cancer patients in Baghdad, it was shown through the results of Table 2 that the best model for estimating survival times is the model resulting from using the Fast ICA method with Random Forests algorithm because it gave the lowest value to the MSE criterion and the highest value to the C-index criterion compared to other models. As for estimating the survival probability of Kaplan-Meier, the best model for its estimation was the same as the previous model according to the two comparison criteria and as shown in Table 3 and Figures 3 and 4. Thus, we recommend adopting the model resulting from the application of the Fast ICA algorithm to extract features with the machine learning algorithm Random Forests to estimate survival based on medical images by doctors and medical professionals, Thus, to increase the accuracy of the estimation results, it is possible to adopt the same algorithms that were used in this research paper, but with larger sample sizes, as it is known that the greater the sample size, the greater the accuracy of prediction and estimation.

Future studies can also adopt the same algorithms, but for other types of diseases.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

- [1] O. A. Abdulwahhb and M. S. Abd Alrazak, "Using nonlinear dimensionality reduction techniques in big data analysis", *Periodicals of Engineering and Natural Sciences*, vol. 8, no. 1, pp. 142–155, 2020.
- [2] A. G. Al-Rawi and A. M. Mohammed, "Split and Merge Regions of Satellite Images using the Non-Hierarchical Algorithm of Cluster Analysis," *Journal of Economics and Administrative Sciences*, vol. 25, no. 111, pp. 466–484, 2019.
- [3] X. Benlin, L. Fangfang, M. Xingliang, and J. Huazhong, "Study on independent component analysis' application in classification and change detection of multispectral images", *The International Archives of the Photogrammetry*, *Remote Sensing and Spatial Information Sciences*, 2008.
- [4] R. K. Hasan and M. M. Rene, "Prediction using Gaussian Process Regression and Support Vector Regression", *Journal of Hama University*, vol. 2, no. 12, pp. 17–31, 2019.
- [5] Y. Hatano, T. Ishihara, S. Hirokawa, and O. Onodera, "Machine learning approach for the prediction of age-specific probability of SCA3 and DRPLA by survival curve analysis," *Neurol. Genet.*, vol. 9, no. 3, p. e200075, 2023.
- [6] X. Hu, *Support vector machine and its application to regression and classification*, *A Masters Thesis Presented to The Graduate College of Missouri State University In Partial Fulfillment Of the Requirements for the Degree*. 2019.
- [7] T. Huang *et al.*, "Machine learning models for predicting survival in patients with ampullary adenocarcinoma", *Asia-Pacific Journal of Oncology Nursing*, vol. 9, pp. 2347–5625, 2022.
- [8] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications" , *Neural Networks Research Centre Helsinki University of Technology P.*, *Neural Networks Research Centre Helsinki University of Technology P.O. Box*, vol. 5400, no. 4–5, pp. 411–430, 2000.
- [9] H. Isltwaran, U. B. Bogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests", *The Annals of Applied Statistics*, vol. 2, no. 3, pp. 841–860, 2008.
- [10] A. G. Jabir and M. A. Mohammed, "Using multidimensional scaling technique in image dimension reduction for satellite image," *Periodicals of Engineering and Natural Sciences*, vol. 8, pp. 447–454, 2020.
- [11] G. Kantidakis, H. Putter, C. Lancia, J. de Boer, A. E. Braat, and M. Fiocco, "Survival prediction models since liver transplantation - comparisons between Cox models and machine learning techniques," *BMC Med. Res. Methodol.*, vol. 20, no. 1, 2020.

-
- [12] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [13] P. Li *et al.*, "Guided Semi-Supervised Non-negative Matrix Factorization," *Algorithms*, vol. 15, no. 5, p. 136, 2022.
- [14] S. Lee and H. Lim, "Review of statistical methods for survival analysis using genomic data," *Genomics Inform.*, vol. 17, no. 4, p. e41, 2019.
- [15] S. Leger *et al.*, "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling," *Sci. Rep.*, vol. 7, no. 1, 2017.
- [16] A. Spooner *et al.*, "A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction," *Sci. Rep.*, vol. 10, no. 1, 2020.
- [17] A. Moncada-Torres, M. C. van Maaren, M. P. Hendriks, S. Siesling, and G. Geleijnse, "Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival," *Sci. Rep.*, vol. 11, no. 1, 2021.
- [18] A. A. Yaqoob and O. A. Ali, "Dynamic modeling for discrete survival data by using artificial neural networks and iteratively weighted Kalman filter smoothing with comparison," *International Journal of Agricultural and Statistical Sciences*, vol. 17, no. 1265–1274, 2021.
- [19] A. A. Yaqoob and O. A. Ali, "Dynamic modeling of time-varying estimation for discrete survival analysis for dialysis patients in Badrah," *International Journal of Agricultural and Statistical Sciences*, vol. 17, pp. 1323–1332, 2021.
- [20] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Image Analysis*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 333–342.
- [21] B. Zupan, J. Demsar, M. W. Kattan, J. R. Beck, and I. Bratko, "Machine learning for survival analysis: a case study on recurrence of prostate cancer," *Artif. Intell. Med.*, vol. 20, no. 1, pp. 59–75, 2000.