A machine learning approach to predict self-efficacy in breast cancer survivors

İsmail Toygar [1], Su Özgür [2], Gülcan Bağçivan [3], Ezgi Bilmiç [4], Hilal Benzer [5], Ferda Akyüz Özdemir [1], Halise Taşkın Duman [1], Özlem Ovayolu [6]

1- Muğla Sıtkı Koçman University Fethiye Faculty of Health Sciences, 2- Ege University Department of Biostatistics, 3- Koç University Faculty of Nursing, 4- Dr Sadi Konuk Training and Research Hospital, 5- Hasan Kalyoncu University Vocational School, 6- Gaziantep University Faculty of Health Sciences

ABSTRACT

Introduction:

The number of breast cancer survivors has increased recently with advances in early detection and treatment modalities. Self-efficacy is crucial for symptom management and improving the quality of life in breast cancer survivors. This study aims to determine predictors of self-efficacy in breast cancer survivors and identify vulnerable groups.

Methods:

This descriptive study was conducted between November 2023 and April 2024 at three hospitals in Türkiye and involved 430 breast cancer survivors. Data were collected through face-to-face surveys using a patient identification form and the Breast Cancer Survivor Self-Efficacy Scale. This study identified patient characteristics that indicate a tendency towards higher self-efficacy using four machine learning models; Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), XGBoost (XGB).

Results :

The mean age of participants was 50.7±11.5 years. Majority of the participants (n=425) were female. There were statistically significant differences between the groups regarding work status, educational level, income level, and treatment modalities ($p<0.05$). Education level ranked first in the LR (0.3874), RF (0.3290), and SVM (0.1250) models, and was the second most important variable in the XGB (0.2327) model. Conversely, the cancer stage stood out in the LR (0.2466) and RF (0.1935) models, ranking third and fourth, respectively, while it ranked third in SVM (0.0683) and fourth in XGB (0.1872). Additionally, comorbidity ranked third in importance in the LR (0.2213) and RF (0.1681) models, but second in SVM (0.0705) and seventh in XGB (0.1393).

Conclusion:

In conclusion, the study found that the self-efficacy of breast cancer survivors was associated with the sociodemographic and medical characteristics of the patients. These characteristics should be considered in patient care and in further studies to improve the self-efficacy of breast cancer survivors.

INTRODUCTION

The number of breast cancer survivors is growing, and the death rate is declining as a result of recent advancements in treatment and healthcare services. Therefore, better understanding and managing the sequels of cancer and its treatment, including both emotional and physical needs among survivors, are essential for the progression of cancer care in breast cancer (1). Studies have shown that higher levels of self-efficacy are associated with better health outcomes. "Machine learning (ML)" is a broad term encompassing various models and strategies centered on algorithmic modeling. (2). Using machine learning to predict self-efficacy and associated factors in breast cancer involves applying advanced algorithms to analyze complex datasets, which include clinical, demographic, and psychological variables. Among cancer survivors, high levels of self-efficacy were associated with high levels of self-care and self-management (3). For this reason, it is important to identify the predictors of self-efficacy in breast cancer survivors. This study aimed to identify predictors of self-efficacy in breast cancer survivors.

METHODS AND MATERIALS

This study was conducted in Türkiye with breast cancer survivors. A total of 430 patients participated in this study. There is no reported method for calculating the sample in machine learning. Therefore, the aim was to reach the maximum number of patients in the sample of this study. The data were collected face-to-face. The data collection process for each patient lasted approximately 10 minutes. Basic statistics were performed using IBM SPSS V25 (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.). The comparison of means/medians between two independent groups for continuous data, the independent sample t-test/Mann-Whitney U test was employed. The chi-square test was utilized to evaluate categorical data. Statistical significance was determined at the $p<0.05$ level. Also, machine learning analyses were performed using Ddsv4-series Azure Virtual Machines with a vCPU count of 32 and a memory capacity of 128 GiB.

RESULTS

Table 1. Comparison of descriptive statistics of participants in groups

Variables	Category	Group		Total	Test statistics; p-value
		Below the Mean	Above the Mean		
Gender	Female	208 (48.9)	217 (51.1)	425 (100)	X^2=0.158 p=0.999
	Male	2 (40.0)	3 (60.0)	5 (100)	
Age	Mean±SD	50.9±12.0	50.6±11.0	50.7±11.5	t=0.277 p=0.782
	Med [Min-Max]	51.0 [20.0-80.0]	50.0 [24.0-80.0]	50.0 [20.0-80.0]	
Marital status	Married	163 (47.7)	179 (52.3)	342 (100)	X^2=0.926 p=0.336
	Single	47 (53.4)	41 (46.6)	88 (100)	
Work status	Housewife	146 (50.7)	142 (49.3)	288 (100)	X^2=10.702 p=0.013
	Working	9 (23.7)	29 (76.3)	38 (100)	
	Not Working	38 (52.8)	34 (47.2)	72 (100)	
	Retired	17 (53.1)	15 (46.9)	32 (100)	
Educational Level	Illiterate	71 (60.7)	46 (39.3)	117 (100)	X^2=14.242 p=0.003
	Primary Education	95 (49.0)	99 (51.0)	194 (100)	
	High School	24 (33.3)	48 (66.7)	72 (100)	
	University	20 (42.6)	27 (57.4)	47 (100)	
Income Level	Income Less Than Expenses	134 (55.8)	106 (44.2)	240 (100)	X2=11.232 p=0.004
	Equal to Income	57 (38.5)	91 (61.5)	148 (100)	
	Income Exceeds Expenses	19 (45.2)	23 (54.8)	42 (100)	
Disease duration (months)	Mean±SD	32,0±46.1	37,0±45.5	30.6±45.8	U=2098.0 p=0.102
	Med [Min-Max]	12.0 [1.0-264.0]	24.0 [1.0-300.0]	18.0 [1.0-300.0]	
Cancer stage	Cured	1 (25.0)	3 (75.0)	4 (100)	X^2=4.402 p=0.367
	Stage 1	37 (44.0)	47 (56.0)	84 (100)	
	Stage 2	70 (48.3)	75 (51.7)	145 (100)	
	Stage 3	67 (55.8)	53 (44.2)	120 (100)	
	Stage 4	35 (45.5)	42 (54.5)	77 (100)	
Comorbidity	Absence	88 (54.3)	74 (45.7)	162 (100)	X^2=3.128 p=0.077
	Presence	122 (45.5)	146 (54.5)	268 (100)	
Treatment	Chemotherapy	87 (51.5)	82 (48.5)	169 (100)	X^2=12.199 p=0.007
	Immunotherapy+ Surgery	9 (34.6)	17 (65.4)	26 (100)	
	Chemotherapy+Surgery	95 (45.5)	114 (54.5)	209 (100)	
	Other	17 (81.0)	4 (19.0)	21 (100)	
Mastectomy	Absence	104 (53.3)	91 (56.7)	195 (100)	X^2=2.887 p=0.089
	Presence	106 (45.1)	129 (54.9)	235 (100)	
Lymphedema	Absence	150 (48.1)	162 (51.9)	312 (100)	X^2=0.263 p=0.608
	Presence	60 (50.8)	58 (49.2)	118 (100)	

X^2: Chi square test statistics, t: Independent sample t test statistics, U: Mann Whitney U test statistics, $p<0.05$ Significance level

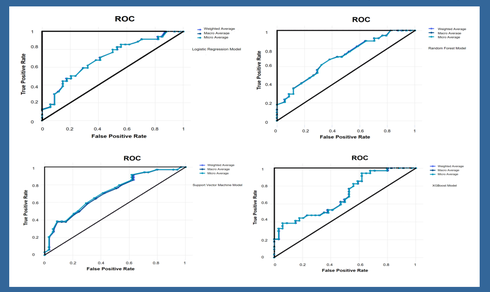

Figure 1. Performance metrics of the prediction models

DISCUSSION

The current study demonstrated a positive causal relationship between higher levels of education and higher levels of self-efficacy among breast cancer survivors. This allows individuals to take a more active role in the decision-making process regarding their care (4).

Cancer stage was also found to be a significant predictor of breast cancer survivors' self-efficacy.. On the other hand, self-efficacy can be influenced by individual, social and cognitive factors. That's why it's not possible to link levels of self-efficacy to cancer stage. We recommend that all factors that influence self-efficacy are considered together.. Comorbidity was another predictor of the self-efficacy of breast cancer survivors in the current study. We believe that coping mechanisms may have been strengthened because they may have previously experienced similar psychological problems due to chronic illness. Working status is another important variable in all the prediction models. There is a two-sided interaction between the work status and self-efficacy. Self-efficacy plays a role in the return to work for breast cancer survivors. It shows that higher self-efficacy is linked to outcomes in returning to work after battling cancer (5). On the other hand, workability, the capability to meet work demands in relation to current health status, is closely related to self-efficacy. (6). We believed that these two conditions are the reason for the differences between the groups regarding the working status. Among the treatment options, those including the combination of surgery and immunotherapy or chemotherapy has higher level of self-efficacy. Enien et al. (2018) reported that patients treated with breast conservative surgery has higher level of quality of life (7). We believed that surgical approaches would lead to higher self-efficacy in cancer survivors. Income was another predictor of self-efficacy in the current study. We recommend healthcare practitioners to be aware of this vulnerable group and implement interventions to increase self-efficacy in this group. The duration of the disease was found to be longer in those with higher levels f self-efficacy. It is recommended that healthcare practitioners should be aware of this issue.

CONCLUSIONS

Education level, cancer stage, comorbidity, and age were identified as the most significant predictors in all models. Furthermore, the logistic regression model revealed that work status, income level, and disease duration were also among the leading variables. It is recommended that healthcare professionals be aware of this vulnerable group. which includes individuals with lower education levels, those diagnosed at stage 3, those with low-income levels, and those with shorter disease durations.

REFERENCES

1. Heidary Z, Ghaemi M, Hossein Rashidi B, Kohandel Gargari O, Montazeri A. Quality of life in breast cancer patients: A systematic review of the qualitative studies. Cancer Control. 2023;30:10732748231168318.
2. Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annu Rev Public Health. 2020;41(1):21-36.
3. Foster C, Fenlon D. Recovery and self-management support following primary cancer treatment. British journal of cancer. 2011;105(1):S21-S8.
4. Liu Y-B, Liu L, Li Y-F, Chen Y-L. Relationship between health literacy, health-related behaviors and health status: a survey of elderly Chinese. International journal of environmental research and public health. 2015;12(8):9714-25.
5. Cheng AS, Lee S, Li N, Tsang S, Zeng Y. Chinese translation and cross-cultural adaptation of the return-to-work self-efficacy scale among chinese female breast cancer survivors. International Journal of Environmental Research and Public Health. 2023;20(5):4225.
6. Rogers LQ, McAuley E, Courneya KS, Verhulst SJ. Correlates of physical activity self-efficacy among breast cancer survivors. American journal of health behavior. 2008;32(6):594-603.
7. Enien MA, Ibrahim N, Makar W, Darwish D, Gaber M. Health-related quality of life: Impact of surgery and treatment modality in breast cancer. Journal of cancer research and therapeutics. 2018;14(5):957-63.