Research Paper

Comparison and Prediction of Breast Cancer Using Discriminant Analysis Algorithm in Active and Inactive Women

Leila Fasihi[1], Hamid Agha-Alinejad[1*], Reza Gharakhanlou[1], Francisco J. Amaro Gahete[2]

1. Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.
2. Department of Physiology, Faculty of Medical Sciences, University of Granada, Granada, Spain.

Use your device to scan and read the article online

Citation Fasihi L, Agha-Alinejad H, Gharakhanlou R, Amaro Gahete FJ. Comparison and Prediction of Breast Cancer Using Discriminant Analysis Algorithm in Active and Inactive Women. Physical Treatments. 2025; 15(3):243-252. http://dx.doi.org/10.32598/ptj.15.3.680.1

http://dx.doi.org/10.32598/ptj.15.3.680.1

Article info:
Received: 16 Dec 2024
Accepted: 22 Jan 2025
Available Online: 01 Jul 2025

Keywords:
Breast cancer, Discriminant Analysis algorithm, Active women, Prediction

ABSTRACT

Purpose: One of the most common types of malignant cancer in women is breast cancer, which has been increasing in recent years. The presence of various symptoms and characteristics of this disease makes diagnosis difficult for doctors. Data mining provides the possibility of analyzing clinical data of patients for medical decision-making. The aim of this article was to compare and predict breast cancer using the differential analysis algorithm in active and inactive women.

Methods: The medical records of 1782 women suspected of having breast cancer constituted the statistical population of this study. After reviewing the files, 642 medical records (329 active women and 313 inactive women) containing laboratory, anthropometric, and demographic information were selected as samples. The differential analysis algorithm and 15 effective features of breast cancer were used to predict the disease. Statistical analyses were performed using MATLAB software, version 2024.

Results: The results showed that using 15 risk factors, the differential analysis algorithm has an accuracy of 79.7% and a precision of 77.5% in predicting breast cancer in active women, while it achieved an accuracy of 71.6% and a precision of 69.3% in inactive women. The results also showed that the differential analysis algorithm performs better in predicting breast cancer in active women.

Conclusion: Given the high accuracy and precision of the differential analysis algorithm in predicting breast cancer, doctors and specialists in the treatment department can use this algorithm in medical and therapeutic centers to predict this type of cancer.

*** Corresponding Author:**
Hamid Agha-Alinejad, Professor.
Address: *Department of Physical Education & Sport Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran.*
E-mail: *halinejad@modares.ac.ir*

Highlights

- The most common malignancy in women worldwide is breast cancer.

- In medical decision-making, data mining enables rapid analysis of patients' clinical data.

- The results revealed that the discriminant analysis algorithm using 15 effective features of breast cancer was used to predict the disease in active and inactive women.

- Exercise has been considered a complementary therapy for the improvement of cancer patients.

Plain Language Summary

Cancer is named after the part of the body tissue that is involved in the disease, and the percentage of cancer types varies in different regions. Using data mining-based methods, extracting knowledge from the massive amount of data related to patient medical records and medical records can lead to the recognition of the laws governing the course of the disease. The results showed that using 15 risk factors, the differential analysis algorithm has using 15 risk factors, the differential analysis algorithm achieved an accuracy of 79.7% and a precision of 77.5% in predicting breast cancer in active women, while it achieved an accuracy of 71.6% and a precision of 69.3% in inactive women. Additionally, physical activity and exercise are among the intervention behaviors that can reduce the incidence of cancer.

Introduction

One of the causes of death and a major challenge in different human societies is cancer, which affects different parts of the body [1]. In this disease, abnormal cells in the body begin to multiply and create more abnormal cells instead of repairing or destroying themselves [2]. Therefore, cancer is a type of disease, in which abnormal cells are produced and begin to multiply, and the accumulation of a number of these cells in parts of the body tissue produces masses that may be benign or malignant [3]. Cancer is named after the part of the body tissue that is involved in the disease, and the percentage of cancer types varies in different regions [4]. Globally, breast cancer is the most prevalent cancer among women [5]. Among all cancers affecting women, breast cancer is the leading cause of cancer-related mortality worldwide [6]. Every year, around 1.1 million new cases of breast cancer are reported in women worldwide [7]. Breast cancer is the second most frequent cancer after lung cancer, accounting for almost one-third of all cancers in women [8]. The number of people affected by it worldwide is about 1.3 million and the annual mortality rate is estimated at 450,000 [9]. The total number of patients with breast cancer in Iran is estimated to be around 40,000, and the annual incidence of breast cancer in Iran is 20 women per 100,000, which is equivalent to 6,000 new cases per year [10].

Cancer treatment is divided into two types: Local and systemic, with surgery and radiotherapy being examples of local treatments and chemotherapy and hormone therapy being examples of systemic treatments, which are usually used together for better results [11]. For years, researchers have been looking for a better solution for the treatment of breast cancer, among which physical activity has been accepted as a safe intervention to improve the quality of life of people with this disease, but its therapeutic aspect is still controversial [12]. Galva et al. reported that 12 weeks of combined resistance and endurance training resulted in increased cardiopulmonary endurance and muscle strength in cancer patients [13]. Furthermore, in the last decade, knowledge regarding the effect of exercise and physical activity on cellular and molecular processes involved in the regulation of tumor metastasis has increased significantly [14]. In this regard, Hejazian et al. reported that a period of exercise training resulted in a decrease in tumor antigenic genes in patients, thus playing an important role in reducing cancer progression [15]. Sheri et al showed that combined training induced the expression of tumor suppressor genes in cancer [16]. Jones et al. showed that aerobic exercise in cancer mice reduced vascular endothelial growth factor (VEGF) levels in muscle and also stabilized tumor mass and tumor tissue while increasing muscle VEGF compared to the control group [17].

Discovering and extracting knowledge from the vast amount of data related to patient medical records using data mining algorithms can lead to the identification and recognition of the laws governing the development and progression of the disease and provide valuable information to health professionals and specialists to identify the causes of diseases and predict and treat the disease according to the prevailing environmental factors [18]. The data relating to the symptoms of patients with various diseases and auxiliary methods for identifying these diseases is so extensive that it seems difficult for a single person to analyze and consider all the effective factors [7]. Using data mining-based methods, extracting knowledge from this substantial amount of data can lead to the recognition of the patterns governing the course of the disease [19]. One of the important applications of data mining is its use in the prediction and early diagnosis of the disease [20]. Early diagnosis of cancer and the use of new therapeutic options can help prevent mortality [19]. Nilashi et al. used a decision tree to classify breast cancer data and the output of their model was fuzzy, which was converted to a definite state using expert opinion [21]. Kang et al. examined the impact of exercise on inflammatory markers in breast cancer survivors. They demonstrated that exercise interventions dramatically decreased fasting insulin levels in breast cancer survivors, highlighting that insulin plays a crucial role in carcinogenesis in various human tissues, including breast tissue, after reviewing 18 studies involving 681 breast cancer survivors [22].

The use of risk factors has been shown to be very effective in cancer prediction using data mining algorithms. Nine modifiable risk factors cause more than 1/3 of cancers in the world, including physical inactivity, smoking, alcohol, obesity, low fruit and vegetable consumption, air pollution, indoor fuel smoke, and contaminated injections [23]. In addition to the above, environmental factors (physical activity, nutrition, etc.) can also affect cancer [24]. Physical activity and exercise are among the intervention behaviors in reducing the incidence of cancer [25]. Epidemiological studies have shown that increasing physical activity not only reduces the incidence of cancer but is also an effective intervention that has attracted the attention of many researchers in recent years [26]. Currently, the investigation of the effects of exercise training and physical activity as therapeutic supplements for improving the conditions of cancer patients has drawn the interest of researchers [27]. To our knowledge, no study has used physical activity as an effective factor in predicting cancer incidence by data mining. Therefore, the aim of this study was to compare and predict breast cancer using a discriminant analysis algorithm in active and inactive women. The discriminant analysis algorithm is a dimensionality reduction technique mainly used in supervised classification problems. This modeling facilitates the distinction between groups and effectively separates two or more classes.

Materials and Methods

The statistical population was all female patients with breast cancer in the age range of 25-75 years who were referred to Ayatollah Kashani and Imam Khomeini hospitals in Tehran from 2011 to 2024. These patients had medical records (containing laboratory, personal, and lifestyle information) stored in the computer archive files of those hospitals. The initial number of files reviewed was 1,782. After examining the files, recording laboratory characteristics and values, and completing the physical activity questionnaire (distributed through social media platforms such as WhatsApp, Telegram, etc.), a total of 642 available patients were ultimately selected as a sample for participation in this study. They were divided into two groups: Active (329 individuals) and inactive (313 individuals) according to the study criteria. In this study, active individuals were defined as those who had exercised regularly three times a week for the past six months [28].

By reviewing medical records, 30 variables were initially selected, and then, using the opinions of two physicians specializing in breast cancer and reviewing the results of scientific articles, 15 anthropometric and physiological variables were finally selected as input features for the algorithm. These variables included exposure to cigarette smoke, age, height, albumin, weight, beta lipoprotein, body mass index, systolic blood pressure, serum selenium levels, cholesterol, high-density lipoprotein cholesterol, use of oral contraceptives, breastfeeding, use of oral hormone replacement therapy (HRT), and family history.

Inclusion criteria included female gender, age between 25 and 75 years, having a medical history and clinical tests in the hospital, and being available by phone or internet. Exclusion criteria included unusual fatigue, anemia, physical dysfunction due to disease, kidney, liver, parathyroid, thyroid diseases, and diabetes mellitus. At the time of the study, the patients had not started any cancer treatment.

In total, this study was conducted in two stages: The first stage involved collecting patient-related data by reviewing hospital records, while the second stage focused on training data mining algorithms using the collected data. A discriminant analysis algorithm was used to predict the disease. This algorithm was created by utilizing the input variables and determining the target variable. To optimally use the data, they had to be adjusted to suit data mining algorithms [29]. For questions that had yes and no answers, the numbers zero and one were assigned, with one indicating a "yes" response and zero indicating a "no" response.

In the next stage, the data were divided into two groups: Training (70%) and testing (30%). The data in the training section were used to build the model, and the data in the testing section evaluated the created model. The data set was transferred to MATLAB software, version 2024 in Excel format for analysis.

Discriminant analysis algorithms

In machine learning and pattern recognition, discriminant analysis is a statistical technique that determines the linear combination of features that best distinguishes between two or more object classes. In these statistical techniques, the dependent variable is represented as a linear combination of other variables.

The algorithm used in discriminant analysis is more akin to logistic regression. Statistical techniques for combining variables in a way that best explains the data include both linear and quadratic discriminant analysis algorithms. Reducing the dimensionality of the data is a significant application of each of these techniques. However, there is one key distinction between these approaches: principal component analysis ignores class differences, whereas linear discriminant analysis models them. In both linear and quadratic discriminant analysis models, this network seeks a combination of variables that best describes the data and also attempts to differentiate between various classes of data [30].

Evaluation criteria

The evaluation criterion was the accuracy criterion, which calculates the classification accuracy and determines the extent to which the algorithm used in this study has performed classification and diagnosis correctly. The recall criterion calculates the rate of correctly predicted positive outcomes by the system; that is, it measures what percentage of the total cancer samples in the tested database were correctly identified as cancer by the system.

Table 1 shows the accuracy and precision criteria based on the data evaluation method, where accuracy is defined as "how many of the selected samples are correct" and precision is defined as "how many of the available correct samples are correctly selected" [31].

$$1.\ \text{"Accuracy"}=\frac{TP+TN}{TP+FP+TN+FN}$$

$$2.\ \text{Precision}=\frac{TP}{TP+FP}$$

TP: The number of cancer subjects that the system correctly diagnosed as cancer. FP: The number of healthy subjects that the system correctly diagnosed as cancer. FN: The number of cancer subjects that the system correctly diagnosed as healthy. TN: The number of healthy subjects that the system correctly diagnosed as healthy.

In this study, the accuracy and precision of the algorithm's performance were evaluated. The accuracy of the algorithm indicates its value in prediction, which is obtained from the number of correct predictions divided by the total number of predictions (Equation 1). The precision of the algorithm indicates its power to distinguish between sick and healthy individuals and is achieved by dividing the number of correct predictions by the number of predictions in each row (Equation 2).

Results

Table 2 shows the anthropometric characteristics and Table 3 presents the descriptive statistics related to the quantitative and qualitative variables of the subjects' files.

Exposure to cigarette smoke, the use of oral contraceptive pills, breastfeeding, the use of oral HRT, and family history were qualitative variables, and age, height, weight, body mass index, systolic blood pressure, serum levels of selenium, cholesterol, beta lipoprotein, albumin, high-density lipoprotein cholesterol were quantitative variables. In this research, 30% of the data were considered for testing and 70% for training the algorithm. A discriminant analysis algorithm was used to predict breast cancer. The results of the clutter matrix of this algorithm are shown in Figures 1 and 2.

The results showed that the discriminant analysis algorithm could predict active women with breast cancer with an accuracy of 79.7% and a precision of 77.5%. The results showed that the discriminant analysis algorithm could predict inactive women with breast cancer with an accuracy of 71.6% and a precision of 69.3%.

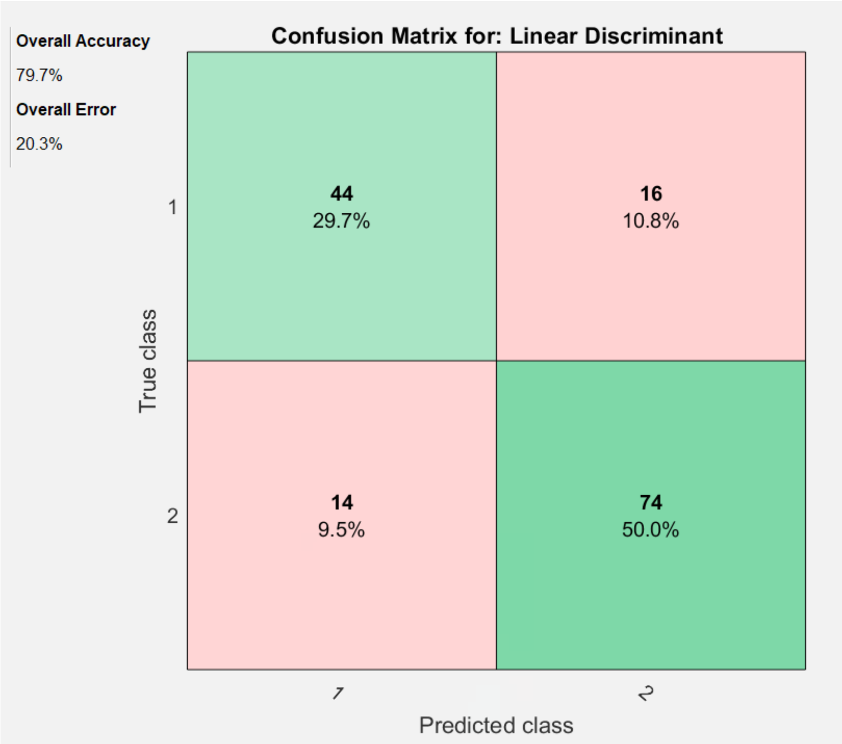

Figure 1. Discriminant analysis algorithm confusion matrix for active women

PHYSICAL TREATMENTS

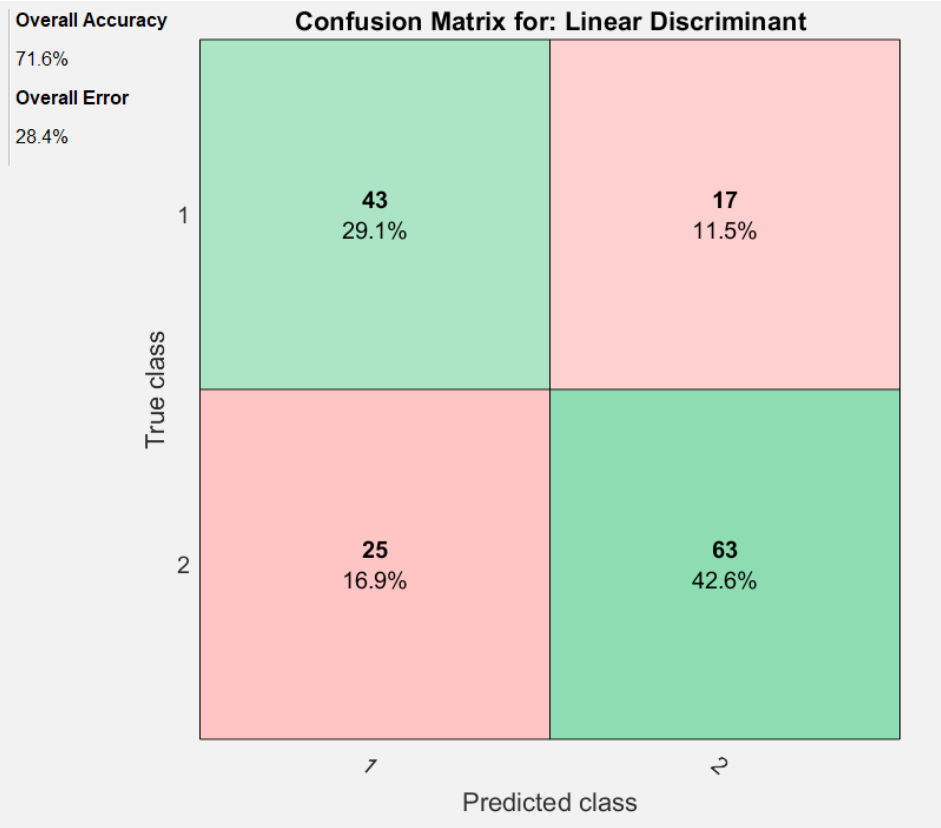

Figure 2. Discriminant analysis algorithm confusion matrix for inactive women

PHYSICAL TREATMENTS

Table 1. Evaluation method

Variable		Correct Values	
		Positive	Negative
Predicted values	Positive	TP	FP
	Negative	FN	TN

Abbreviations: TP: True positive; FN: False negative; FP: False positive; TN: True negative.

Discussion

The emergence of large data sets and the development of databases over the past few decades have created new needs, such as automatic data summarization, extraction of stored information, and discovery of patterns from raw data, of which data mining is an example. Some data mining algorithms, such as machine learning methods, can predict different situations in the future by gradually learning these patterns and existing conditions, in addition to analyzing data and extracting hidden patterns from them [32]. The possible relationship between risk factors and breast cancer in women has been discussed for years. In addition, cancer treatment workers are seeking ways to predict the likelihood of developing cancer. A topic that has received less attention is the prediction of diseases related to breast cancer using data mining. The aim of the present study was to compare and predict breast cancer using a discriminant analysis algorithm in active and inactive women. Seventy percent of the data was allocated for training and 30% for testing the algorithm.

The results of this study showed that the discriminant analysis algorithm could predict the likelihood of developing breast cancer with 79.7% accuracy and 77.5% precision in active individuals, and with 71.6% accuracy and 69.3% precision in inactive individuals, using 15 quantitative and qualitative indicators. Physical activity, through its effects on the characteristics and abilities of individuals' bodies, acts as a barrier against diseases and improves physical conditions. In line with the results of this study, Rabiei et al. predicted the probability of breast cancer using the performance of four algorithms: Random forest (RF), multilayer perceptron (MLP), gradient boosting trees (GBT), and genetic algorithm (GA), along with 24 demographic, laboratory, and mammographic

Table 2. Quantitative variables of the subjects

Parameters	Mean±SD			
	Healthy		Patients	
	Active	Inactive	Active	Inactive
Age (y)	48.86±12.18	48.37±11.81	49.76±11.17	49.47±12.71
Height (cm)	157.38±17.14	156.47±16.33	158.29±21.24	156.74±19.41
Weight (kg)	65.42±12.57	66.31±26.61	62.31±12.52	63.82±12.57
Body mass index (kg/m^2)	23.58±4.75	24.84±2.81	27.59±2.12	28.78±3.91
Serum selenium (µg/dL)	23.85±9.80	22.56±8.28	16.24±8.21	15.55±7.83
BLP (g/L)	10.31±2.27	12.42±2.54	14.14±2.46	14.82±2.18
Albumin (g/L)	40.6±2.19	39.96±2.91	39.25±2.56	38.42±2.89
Systolic blood pressure (mm Hg)	126.9±27.9	135.8±26.7	146.9±28.6	156.12±34.75
Cholesterol (mg/dL)	255.12±41.48	256.22±42.67	277.54±41.43	278.66±92.78
HDL-C (mmol/L)	1.88±0.16	1.81±0.15	1.11±0.09	1.25±0.69

BLP: Beta-lipoprotein.

Table 3. Qualitative variables of the subjects

Characteristic	Type (Discontinuous)	%			
		Healthy		Patients	
		Active	Inactive	Active	Inactive
Smoking	Yes (1) No (0)	64 36	67 33	62 38	65 35
Using oral contraceptive pills	Yes (1) No (0)	30 70	33 67	25 75	34 66
Using oral HRT	Yes (1) No (0)	55 45	56 44	59 41	60 40
Breastfeeding	Yes (1) No (0)	33 67	30 70	28 72	80 20
Family history	Yes (1) No (0)	66 34	69 31	67 33	72 28

PHYSICAL TREATMENTS

HRT: Hormone replacement therapy.

characteristics. They obtained the highest performance with the RF algorithm, achieving an area under the curve (AUC) of 0.56, sensitivity of 95%, specificity of 80%, and accuracy of 80% [33].

Naji et al. used machine learning algorithms in their study to diagnose and predict breast cancer. They compared the results of five algorithms: logistic regression, support vector machine (SVM), k-nearest neighbor (KNN), RF, and decision tree (C4.5) using a breast cancer diagnostic dataset. Ultimately, they reported the SVM algorithm as having the highest accuracy at 97.2% [34].

Land et al. in their study in their study titled "multiclass primitive SVMs for breast cancer classification," used SVM to predict breast cancer on a breast cancer dataset and ultimately reported an accuracy of 96.7% [35]. Lavanya et al. achieved an accuracy of 94.84% using data from the WBCD database and a two-stage classification decision tree [36]. Kiyan et al. reported accuracies of 96.18% and 95.74% using the RBF and MLP methods for breast cancer prediction, respectively [37]. Chaurasia et al. achieved an accuracy of 96.84% in their study using the SVM method [38]. Sarvestani et al. compared the mean square error in multilayer, competitive, and radial basis neural networks to predict the grade of breast cancer malignancy, finding that the radial basis neural network had the best accuracy [39].

Mosayebi et al. investigated the prediction of breast cancer recurrence using three data mining techniques. They reported accuracies of 0.936, 0.947, and 0.957 for the results of three data mining algorithms, namely decision tree, ANN, and SVM, respectively [40].

Compared to the current study, some machine learning studies have claimed higher sensitivity and accuracy for breast cancer prediction. This is probably because various databases, such as Wisconsin and SEER, as well as different indices and methods were used [41-43]. Behravan et al. used a database with 695 entries, including genetic information and demographic risk variables to predict breast cancer. Their results demonstrated that the multi-factor boost model outperformed a model using a single set of factors [44]. In a study by Feld et al. modeling was done on 738 records that included genetic, demographic, and mammographic anomalies for breast cancer prediction [45]. In addition, studies show that the performance of the algorithm improves by considering multiple indices in the modeling. For example, Ayvaci et al. found that applying logistic regression to analyze demographic, mammographic, and biopsy data increased accuracy [46]. Additionally, Rajendran et al. used the Naïve Bayes, RF, and C4.5 algorithms to predict breast cancer by analyzing 2.4 million mammography screening records and demographic risk factors linked to breast cancer. The results revealed that Naïve Bayes achieved the greatest AUC (0.993) [47].

Given the importance of breast cancer, the use of data mining algorithms for the timely prediction and diagnosis of this disease is essential. In this regard, more data and different algorithms can be used in future studies and the results can be compared. It is also recommended that the field of data mining science be prepared to predict disease recurrence after surgery in hospitals so that doctors and specialists can use an appropriate environment to examine and treat these patients and, as a result, prevent irreparable harm in women with breast cancer.

This study had some limitations that can be addressed in future research, including the limited number of features, the geographical limitation of the data collection location, and the presence of null and missing values in the dataset. The larger the number of subjects and data sets, the more accurate and complete the results of breast cancer prediction will be and can be used to identify people at risk of cancer, improve quality of life, and prevent its consequences. Therefore, future studies are recommended to use a larger sample size and a wider dataset in different regions.

Conclusion

Predicting and correctly diagnosing breast cancer using artificial intelligence and machine learning increases the chances of correct diagnosis and successful treatment because timely and suitable therapeutic actions can help decrease the disease's progression and lower mortality if this disease is detected early. This study optimized data mining results for breast cancer prediction and diagnosis using a discriminant analysis technique. The algorithm's performance can be enhanced by utilizing various machine learning techniques, gaining access to bigger datasets, and taking into account important characteristics from numerous relevant data sources.

Ethical Considerations

Compliance with ethical guidelines

This study was approved by the Ethics Committee of Tarbiat Modares University, Tehran, Iran (Code: IR.MODARES.REC.1402.185). Participants entered the study after providing written informed permission and had the opportunity to withdraw at any time.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

All authors contributed equally to the conception and design of the study, data collection and analysis, interpretation of the results, and drafting of the manuscript. Each author approved the final version of the manuscript for submission.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgments

The authors thank all participants in this study.

References

[1] Ma RJ, Ma C, Hu K, Zhao MM, Zhang N, Sun ZG. Molecular mechanism, regulation, and therapeutic targeting of the STAT3 signaling pathway in esophageal cancer (Review). International Journal of Oncology. 2022; 61(3):105. [DOI:10.3892/ijo.2022.5424] [PMID]

[2] Saba T. Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons And Challenges. Journal of Infection and Public Health. 2020; 13(9):1274-89. [DOI:10.1016/j.jiph.2020.06.033] [PMID]

[3] Alshammary RA, Al-Attar MM. Overview of pathophysiology of cancer, types, causes, treatment. Journal of Prospective Researches. 2024; 24(4):17–24. [DOI:10.61704/jpr.v24i4.pp17-24]

[4] Kashyap D, Pal D, Sharma R, Garg VK, Goel N, Koundal D, et al. Global increase in breast cancer incidence: Risk factors and preventive measures. BioMed Research International. 2022; 2022:9605439. [DOI:10.1155/2022/9605439] [PMID]

[5] Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. 2022; 66:15-23. [DOI:10.1016/j.breast.2022.08.010] [PMID]

[6] Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast cancer-epidemiology, risk factors, classification, prognostic markers, and current treatment strategies-An updated review. Cancers. 2021; 13(17):4287. [DOI:10.3390/cancers13174287] [PMID]

[7] Kaur G, Gupta R, Hooda N, Gupta NR. Machine learning techniques and breast cancer prediction: A review. Wireless Personal Communications, 2022; 125(3):2537-564. [DOI:10.1007/s11277-022-09673-3]

[8] Alsayadi HA, Abdelhamid AA, El-Kenawy ES, Ibrahim A, Eid MM. Ensemble of machine learning fusion models for breast cancer detection based on the regression model. Fusion. 2022; 9(2):19-26. [DOI:10.54216/FPA.090202]

[9] Silva-Aravena F, Núñez Delafuente H, Gutiérrez-Bahamondes JH, Morales J. A hybrid algorithm of ml and xai to prevent breast cancer: A strategy to support decision making. Cancers. 2023; 15(9):2443. [DOI:10.3390/cancers15092443] [PMID]

[10] Biglu MH. Breast cancer in Iran: The trend of Iranian researchers' studies in MEDLINE database. Basic & Clinical Cancer Research. 2014; 6(1):22-32. [Link]

[11] Abbas Z, Rehman S. An overview of cancer treatment modalities. Neoplasm. 2018; 1:139-57. [DOI:10.5772/intechopen.76558]

[12] Franzoi MA, Agostinetto E, Perachino M, Del Mastro L, de Azambuja E, Vaz-Luis I, et al. Evidence-based approaches for the management of side-effects of adjuvant endocrine therapy in patients with breast cancer. The Lancet. Oncology. 2021; 22(7):e303-13. [DOI:10.1016/S1470-2045(20)30666-5] [PMID]

[13] Galvão DA, Nosaka K, Taaffe DR, Spry N, Kristjanson LJ, McGuigan MR, et al. Resistance training and reduction of treatment side effects in prostate cancer patients. Medicine and Science in Sports and Exercise. 2006; 38(12):2045-52. [DOI:10.1249/01.mss.0000233803.48691.8b] [PMID]

[14] Perego S, Sansoni V, Ziemann E, Lombardi G. Another weapon against cancer and metastasis: Physical-activity-dependent effects on adiposity and adipokines. International Journal of Molecular Sciences. 2021; 22(4):2005. [DOI:10.3390/ijms22042005] [PMID]

[15] Hejazian MB, Barari A, Abbasi-Daloii A, Hasrak K. [The effect of a period of physical exercise on the plasma and gene expression levels of hypoxia-inducible factor-1 (HIF-1) and serum prostate specific antigen levels in men with prostate cancer (Persian)]. Journal of Isfahan Medical School. 2019; 36(505):1434-43. [DOI:10.22122/jims.v36i505.10266]

[16] Sheri A, Martin LA, Johnston S. Targeting endocrine resistance: Is there a role for mTOR inhibition? Clinical Breast Cancer. 2010; 10(Suppl 3):S79-85. [DOI:10.3816/CBC.2010.s.016] [PMID]

[17] Jones LW, Antonelli J, Masko EM, Broadwater G, Lascola CD, Fels D, et al. Exercise modulation of the host-tumor interaction in an orthotopic model of murine prostate cancer. Journal of Applied Physiology. 2012; 113(2):263-72. [DOI:10.1152/japplphysiol.01575.2011] [PMID]

[18] Zadeh AH, Alsabi Q, Ramirez-Vick JE, Nosoudi N. Characterizing basal-like triple negative breast cancer using gene expression analysis: A data mining approach. Expert systems With Applications. 2020; 148:113253. [DOI:10.1016/j.eswa.2020.113253]

[19] Alsabry A, Algabri M, Ahsan AM. Breast cancer-risk factors and prediction using machine-learning algorithms and data source: A review of literature. Sana'a University Journal of Applied Sciences and Technology. 2023; 1(2):145-66. [DOI:10.59628/jast.v1i2.361]

[20] Conte L, Rizzo E, Civino E, Tarantino P, De Nunzio G, De Matteis E. Enhancing breast cancer risk prediction with machine learning: Integrating BMI, smoking habits, hormonal dynamics, and BRCA gene mutations—A game-changer compared to traditional statistical models? Applied Sciences. 2024; 14(18):8474. [DOI:10.3390/app14188474]

[21] Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L. A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics. 2017; 34(4):133-44. [DOI:10.1016/j.tele.2017.01.007]

[22] Kang DW, Lee J, Suh SH, Ligibel J, Courneya KS, Jeon JY. Effects of exercise on insulin, IGF axis, adipocytokines, and inflammatory markers in breast cancer survivors: A systematic review and meta-analysis. Cancer Epidemiology, Biomarkers & Prevention. 2017; 26(3):355-65. [DOI:10.1158/1055-9965.EPI-16-0602] [PMID]

[23] Krishnaiah V, Narsimha G, Chandra NS. Diagnosis of lung cancer prediction system using data mining classification techniques. International Journal of Computer Science and Information Technologies. 2013; 4(1):39-45. [Link]

[24] Hong BS, Lee KP. A systematic review of the biological mechanisms linking physical activity and breast cancer. Physical Activity and Nutrition. 2020; 24(3):25-31. [DOI:10.20463/pan.2020.0018] [PMID]

[25] McTiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical activity in cancer prevention and survival: A systematic review. Medicine and Science in Sports and Exercise. 2019; 51(6):1252-61. [DOI:10.1249/MSS.0000000000001937] [PMID]

[26] Brown JC, Winters-Stone K, Lee A, Schmitz KH. Cancer, physical activity, and exercise. Comprehensive Physiology. 2012; 2(4):2775-809. [DOI:10.1002/j.2040-4603.2012.tb00476.x] [PMID]

[27] Ferioli M, Zauli G, Martelli AM, Vitale M, McCubrey JA, Ultimo S, et al. Impact of physical exercise in cancer survivors during and after antineoplastic treatments. Oncotarget. 2018; 9(17):14005-34. [DOI:10.18632/oncotarget.24456] [PMID]

[28] Cannioto RA, Hutson A, Dighe S, McCann W, McCann SE, Zirpoli GR, et al. Physical activity before, during, and after chemotherapy for high-risk breast cancer: Relationships with survival. Journal of the National Cancer Institute. 2021; 113(1):54-63. [DOI:10.1093/jnci/djaa046] [PMID]

[29] Rafiei FM, Manzari SM, Bostanian S. Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. Expert Systems With Applications. 2011; 38(8):10210-7. [DOI:10.1016/j.eswa.2011.02.082]

[30] Kuhn M, Johnson K. Discriminant analysis and other linear classification models. In: Kuhn M, Johnson K, editors. Applied predictive modeling. New York: Springer; 2013. [DOI:10.1007/978-1-4614-6849-3_12]

[31] Abreu PH, Santos MS, Abreu MH, Andrade B, Silva DC. Predicting breast cancer recurrence using machine learning techniques: A systematic review. ACM Computing Surveys (CSUR). 2016; 49(3):1-40. [DOI:10.1145/2988544]

[32] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal. 2017; 15:104-16. [DOI:10.1016/j.csbj.2016.12.005] [PMID]

[33] Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaeili M, Atashi A. Prediction of breast cancer using machine learning approaches. Journal of Biomedical Physics & Engineering. 2022; 12(3):297-308. [DOI:10.31661/jbpe.v0i0.2109-1403] [PMID]

[34] Naji MA, El Filali S, Aarika K, Benlahmar EH, Ait Abdelouhahid R, Debauche O. Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Computer Science. 2021; 191:487-92. [DOI:10.1016/j.procs.2021.07.062]

[35] Land WH Jr, Verheggen EA. Multiclass primal support vector machines for breast density classification. International Journal of Computational Biology and Drug Design. 2009; 2(1):21-57. [DOI:10.1504/IJCBDD.2009.027583] [PMID]

[36] Lavanya D, Rani KU. Ensemble decision tree classifier for breast cancer data. International Journal of Information Technology Convergence and Services. 2012; 2(1):17-24. [DOI:10.5121/ijitcs.2012.2103]

[37] Kıyan T, Yıldırım T. Breast cancer diagnosis using statistical neural networks. IU-Journal of Electrical & Electronics Engineering. 2004; 4(2):1149-53. [Link]

[38] Chaurasia S, Chakrabarti P. An approach with support vector machine using variable features selection on breast cancer prognosis. International Journal of Advanced Research in Artificial Intelligence. 2013; 2(9):38-42. [DOI:10.14569/IJARAI.2013.020907]

[39] Sarvestani AS, Safavi AA, Parandeh NM, Salehi M. Predicting breast cancer survivability using data mining techniques. Papr presented at: 2010 2nd International Conference on Software Technology and Engineering. 2010 October 3; San Juan, USA. [DOI:10.1109/ICSTE.2010.5608818]

[40] Mosayebi A, Mojaradi B, Bonyadi Naeini A, Khodadad Hosseini SH. Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. Plos One. 2020; 15(10):e0237658. [DOI:10.1371/journal.pone.0237658] [PMID]

[41] Bayrak EA, Kırcı P, Ensari T. Comparison of machine learning methods for breast cancer diagnosis. Paper presented at: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science. 2019 April 24; Istanbul, Turkey. [DOI:10.1109/EBBT.2019.8741990]

[42] Alghunaim S, Al-Baity HH. On the scalability of machine-learning algorithms for breast cancer prediction in big data context. Ieee Access. 2019; 7:91535-46. [DOI:10.1109/ACCESS.2019.2927080]

[43] Memon MH, Li JP, Haq AU, Memon MH, Zhou W. Breast cancer detection in the IOT health environment using modified recursive feature selection. Wireless Communications and Mobile Computing. 2019; 2019(1):5176705. [DOI:10.1155/2019/5176705]

[44] Behravan H, Hartikainen JM, Tengström M, Kosma VM, Mannermaa A. Predicting breast cancer risk using interacting genetic and demographic factors and machine learning. Scientific Reports. 2020; 10(1):11044. [DOI:10.1038/s41598-020-66907-9] [PMID]

[45] Feld SI, Fan J, Yuan M, Wu Y, Woo KM, Alexandridis R, et al. Utility of genetic testing in addition to mammography for determining risk of breast cancer depends on patient age. AMIA Summits on Translational Science Proceedings. 2018; 2018:81. [PMID]

[46] Ayvaci MU, Alagoz O, Chhatwal J, Munoz del Rio A, Sickles EA, Nassif H, et al. Predicting invasive breast cancer versus DCIS in different age groups. BMC Cancer. 2014; 14:584. [DOI:10.1186/1471-2407-14-584] [PMID]

[47] Rajendran K, Jayabalan M, Thiruchelvam V. Predicting breast cancer via supervised machine learning methods on class imbalanced data. International Journal of Advanced Computer Science and Applications. 2020; 11(8):54-63. [DOI:10.14569/IJACSA.2020.0110808]