Applied Medical Informatics
Review Vol. 41, No. 4 /2019, pp: 129-139

Accuracy Comparison of Data Mining Algorithms Used in the Diagnosis of Breast Cancer: A Scoping Review Study

Rouhallah KHARA[1,*], Zakieh PIRI[1], Mohamad DEHGHAN[3], Morteza ARAB-ZOZANI[4]

[1] Department of Health Information Technology, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran.
[2] Student Research Committee, School of Management and Medical Informatics, Tabriz University of Medical Science, Tabriz, Iran.
[3] Social Determinants of Health Research Center, Birjand University of Medical Sciences, Birjand, Iran.
E-mails: r.khara@yahoo.com; piriz@yahoo.com; dehghanm@gmail.com; arab.hta@gmail.com

* Author to whom correspondence should be addressed; Tel.: +989169894136; Fax: +984133251378

Received: June 3, 2019/Accepted: December 18, 2019 / Published online: December 30, 2019

Abstract
Introduction: Breast cancer recognized as one of the widespread types of invasive cancer. Early diagnosis of breast cancer is crucial in treating it. The concept of data mining refers to the process of discovering and identifying information in large datasets. Data mining uses a set of techniques and algorithms that can be used for the early detection of breast cancer. The present study gives comparisons between the performances of various algorithms used to diagnose breast cancer. *Method*: This scoping review was led by the framework of the JBI methodology rules. The search was conducted in some relevant electronic databases, and PICO based extracted data were analyzed with Excel software. *Results*: The most commonly used algorithms were SVM (8 cases), j48 and Naive Bayes (7 cases), and MLP (6 cases), and 34 cases were only used in one study. The accuracy rate obtained with FSRAIRS2 (100%) is the highest among the other reported algorithms by other researchers. Moreover, Canopy was the less accurate algorithm (accuracy= 65%). *Conclusion*: Any use of a data mining and knowledge discovery method on a data set requires some discussion on the accuracy of the extracted model on some test data. In this study, we have investigated 48 common algorithms on one of the most crucial areas in medicine. Using algorithms that have high accuracy, automated, and semi-automated tools can be designed and used by professionals for the timely detection of breast cancer.

Keywords: Breast Cancer; Data Mining; Accuracy; Algorithms

Introduction

Breast cancer recognized as one of the extremely common types of invasive cancer concerning females global and is the second most frequent cancer in the world [1]. It is rising, especially in developing countries where most of the problems are a diagnosis of the disease in advanced stages. Altogether, breast cancer is the fifth reason for mortality from cancer also the primary cause of death from women cancer in those countries [2]. In the USA, it acts like to be the second leading reason for death between women and the most usual cause of death in the duration 40 to 55 years in them [3]. Mainly, one of every eight women takes breast cancer during her lifetime. This type of carcinoma is often observed in men too. Then Breast cancer is one of the most common threats to women's health and the most common cancer among them [3].

The agreeable survival rate in breast cancer is because of two factors: the first is the diagnosis of the disease at an early stage, and the second agent is due to advancements in adjuvant systemic therapy. Then the correct diagnosis is significant progress to breast cancer treatment [4]. The effectiveness of early diagnosis has been confirmed to decrease much mortality among patient [5]. In general, there are three diagnostic methods for breast cancer: physical examination, mammography, Fine needle aspiration biopsy (FNAB or FNAC) [6]. The first two methods are sometimes inaccurate and have some errors, as in the research literature, researchers have suggested that radiologists interpret mammography's differently [7]. Moreover, its accuracy varies from 68 to 79 percent [8].

The biopsy is required for the diagnosis of benign or malignant. Biopsy accuracy is about 100%, but it is invasive and costly and time-consuming and painful. FNAC, which is one of the common diagnostic methods for breast cancer, is also interpreted in a variety of ways, with accuracy about 95%. It depends on the amount of physician experience [9]. Many factors are involved in predict, diagnosis and treat this disease, and the high similarity of laboratory findings increases the possibility of misdiagnosis [8].

As noted, the initial diagnosis of breast cancer may increase the chance of treatment. In other words, early diagnosis of breast cancer is crucial in treating it [10]. To achieve this critical goal, doctors need a high-accuracy diagnostic method to understand the difference between benign and malignant tumors [11].

The correct diagnosis is one of the essential steps in treating breast cancer. Data mining techniques can support specialists to give the correct diagnosis of this disease [11]. Various data mining techniques can have a supporting role for doctors in diagnosing diseases. Also, these techniques can dramatically reduce the cost and timing of the diagnosis [12].

Many studies show that computer-aided diagnosis can reduce false positives and detect breast cancer at an early stage [13-15]. Studies have also shown that diagnosis and prediction of various diseases are possible using data mining techniques [16-18]. Data mining means the extraction of hidden information or patterns and relationships in a large volume of data in one or more large databases that simple statistical analyzes cannot do. Data mining in medicine refers to the process of extracting valid, unknown, and reliable information from medical databases that are commonly used to predict and diagnose a disease [19].

This tool can be a way to improve patient outcomes and reduce costs and improve the health status [19]. Besides, the use of automated or semi-automatic detection tools for deadly diseases such as cancer is essential. Data mining is one of the new tools for early detection of cancer. Therefore, it seems necessary to recognize and improve the various techniques of this method. Data mining can be used to classify tumors in benign and malignant categories too [20].

Also, today, with the increased storage of medical data in medical databases, appropriate tools for data evaluation, analysis, and discovery of knowledge are needed for various purposes. Data mining is a method that can accomplish this significant tasks [21]. In this study, we extracted different data mining techniques and algorithms from different studies and compared their accuracy. Accuracy is the most generally used objective indicator to estimate the performance of diagnosis results. The other used indicators are sensitivity, specificity, positive predictive value, and negative predictive value. Moreover, the authors find the beast considered technique in past studies. The compare of techniques and algorithms was evaluated to determine which technique and classifiers had better performance and better accuracy.

Material and Method

Search Strategy

This scoping review was led by the framework of the JBI (Joanna Briggs Institute Reviewers' Manual) methodology rules [22]. The search was conducted in PubMed, Web of Science, Scopus, Science Direct, Embase, Clinical Key, and Cochrane Library until February 26, 2017. The medical subject heading (MeSH) and non-MeSH terms used. Keywords included "breast cancer," "data mining," "algorithms," and "technique," combined with Boolean operators. Furthermore, cross-references listed in included studies were searched too. The search was done in English languages only. Extracted data analyzed with Excel software V.2013.

Selection Criteria

Articles were selected based on PICO inclusion and exclusion criteria. Early screening of the articles given titles and abstract was performed by three researchers the full content of the screened articles was analyzed if: (1) the study included breast cancer disease using (2) one or more the techniques or data mining algorithms to diagnose, (3) the accuracy of the diagnosis was compared with other algorithms or expert diagnosis, and (4) diagnosis accuracy has been reported.

Results

Out of 2935 articles, after eliminating duplicates, evaluating abstracts, and full text, finally, 32 articles remained in the field of the breast cancer diagnosis.

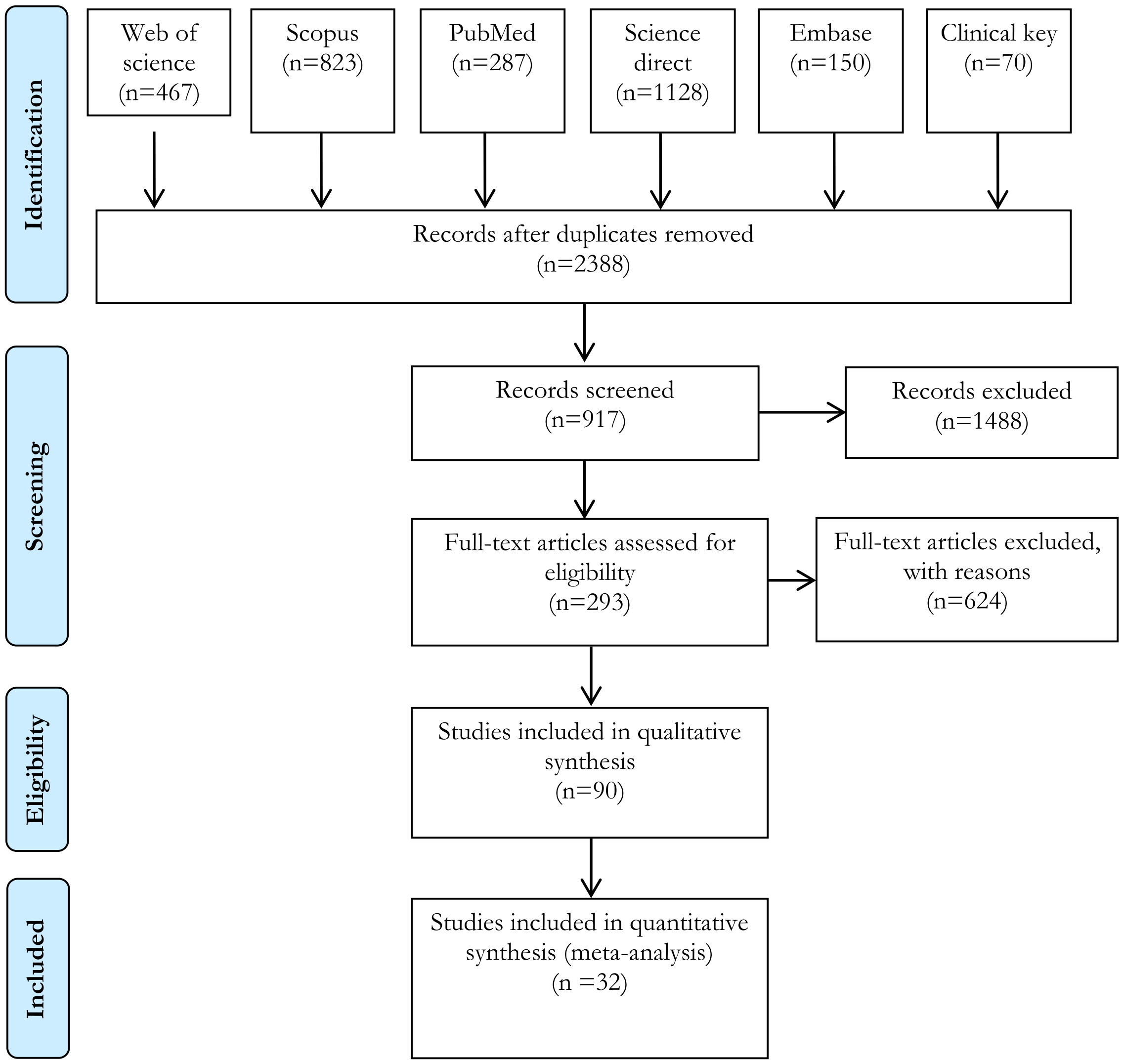

Figure 1. Articles selection flowchart

As shown in the selection flowchart of articles, 32 articles related to the diagnosis of breast cancer using data mining were selected.

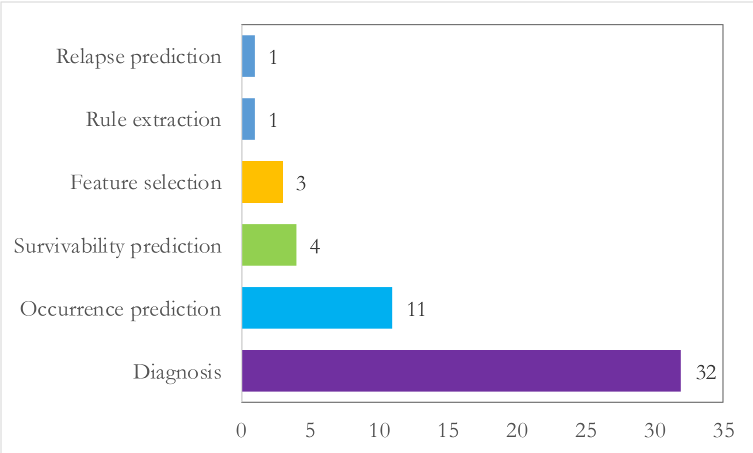

Figure 2. Data mining applications in breast cancer

Table 1. Reviewed articles

Ref	Article propose	Techniques and algorithms	Data base	Software
Diz et al. (2016)	using Data Mining Techniques to Improve Breast Cancer Diagnosis	Libsvm,Random forest, SVM_SMO Naive bayes, K-nearest neighbor	(BCDR)[1]	WEKA
Alickovic et al. (2017)	using GA feature selection and Rotation Forest to Breast cancer diagnosis	Logistic Regression, Decision Trees (C4.5) Random Forest Bayes Net, ANN (MLP), RBFN, SVM Rotation Forest	NR[2]	WEKA
Asri et al. (2016)	Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis	C4.5, SVM, NB, K-NN	Wisconsin	WEKA
Azar Shereen et al. (2013)	Using Decision tree classifiers for automated medical diagnosis	Single decision tree (SDT), boosted decision tree (BDT), decision tree forest (DTF)	Wisconsin	DTREG
Daoudi et al. (2013)	Using An Immune-Inspired Approach for Breast Cancer Classification	CLONALG, CSA, AIRS1, IMMUNOS1, CLONAX, Median Filter AIS	Wisconsin	NR
Dong et al. (2015)	Using An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms	novel automated segmentation and classification SVM, GA-SVM, PSO-SVM, decision tree	DDSM database and MIAS	NR
Fiuzy et al. (2012)	Introduction of a New Diagnostic Method for Breast Cancer	New algorithm	Wisconsin	NR
Ghosh et al. (2014)	Introduction of A novel Neuro-fuzzy classification technique for data mining	NFS, RBFNN, ANFIS	benchmark UCI data sets	MATLA B

[1] - Breast Cancer Digital Repository
[2] - Not reported

Table 1 (continued).

Ref	Article propose	Techniques and algorithms	Data base	Software
Jeyasingh et al. (2017)	Modified Bat Algorithm for Feature Selection	Bat Algorithm	WDBC	NR
Ghosh et al. (2014)	A Comparative Study of Breast Cancer Detection Classifier	MLP BPN, SVM	Wisconsin	MATLAB
Holsbach et al. (2014)	using A data mining method for breast cancer identification	K-nearest neighbor algorithm	Wisconsin	
Yeh et al. (2009)	Introduction A new hybrid approach for mining breast cancer pattern	DPSO	NR	NR
Sasikala et al. (2015)	Introduction A Novel Feature Selection Technique for Improved Survivability Diagnosis of Breast Cancer	SVEGA K-NN, NB, SVM, J48	Kent ridge biomedical repository	NR
Anh Pham et al. (2009)	An application of a new meta-heuristic for optimizing	SVM-HBA, DT-HBA, ANN-HBA	(BC)machine learning data repository	NR
Chen et al. (2016)	Introduction An Approach Based on Biclustering and Neural Network for Classification of Lesions in Breast	Biclustering, BPNN, Fuzzy, cerebellar, mode NN, Fuzzy svm, SVM	Cancer Center of Sun Yat-sen University	NR
Sutha et al. (2015)	Introduction An Efficient Method for Detection of Breast Cancer	MCFI-DS	Wisconsin	NR
Kumar et al. (2013)	Introduce An data mining technique for classification and detection of breast cancer	rule mining classifier	DDSM[3]	NR
Radovic et al. (2013)	Application of Data Mining Algorithms for Mammogram Classification	Naive Bayes, Logistic regression, SVM, KNN, C4.5, Random forest, MLP	MINIMIAS database	NR
Kaushik et al. (2016)	Using Data Mining for Prediction of Breast Tissue Biopsy Results	Multilayer Perceptron , Random Forest , Random Tree ,Ensemble Classifier, DT,SVM, DGP, CBR, ANN	Mammographic Masses Dataset [4]	WEKA
Doreswamy et al. (2015)	Introduce A Bio Inspired Model for Prediction of Breast Cancer	BAT-ELM	Wisconsin	NR
Zheng et al. (2014)	Using data mining for Breast cancer diagnosis	K-SVM, hybrid, k means and SVM	WDBC	NR
Gao et al. (2012)	Using Support Vector Machine to Breast Cancer Diagnosis	Support Vector Machine	WDBC	NR
Kuol et al. (2001)	Using Data mining for diagnosis of breast tumor	decision tree	Digital US images	NR
Jin et al. (2012)	Construction of an system to breast cancer diagnosis and prognosis	Naıve Bayes, FT, IB1, Random Tree, J48, LMT, 1-R	Wisconsin	NR

3 - digital database for screening mammography (DDSM) university of South Florida
4 - obtained from the UCI

Table 1 (continued).

Ref	Article propose	Techniques and algorithms	Data base	Software
Hassan Al-Hagery (2016)	Assessment Classifiers' Accuracy Based on Breast Cancer Medical Data	J48, Bayes Net, Naïve Bayes, MLP, RBF, LMT	Wisconsin	WEKA
Saybani et al. (2015)	Diagnosing breast cancer with a recognition system	FSRAIRS2	Wisconsin	NR
Mahsal et al. (2014)	Breast Cancer Classification using Neural Networks	CGPWNN, CGPWNN-NL	DDSM[5]	NR
Salama et al. (2016)	Experimental Comparison of Classifiers for Breast Cancer Diagnosis	NB, MLP, J48, SMO, IBK	WDBC	WEKA
Doreswamy et al. (2015)	Classification of Breast Cancer using Neural Networks	FM−ANN, Feed forward, MLP, RBF, MNN	WBCD KDD	NR
Huang et al. (2009)	Breast Cancer Diagnosis using Neural Network Classifier	BPNN with PSO	WBCD	NR
Markkongkeaw et al. (2013)	Breast Cancer Cell Classifying Based on GLCM	GLCM	NR	NR
Devi et al. (2015)	Comparison of Various Clustering Techniques for Diagnosis of Breast Cancer	DBSCAN, Farthest First, Canopy, LVQ	Wisconsin	WEKA

The period of analyzed studies was between 2009 and 2017, most of these studies were conducted on the WBCD database (about 56%), and the rest were based on other databases. Most of these studies did not report on what software they used, but among the reported articles, WEKA software was used more than others.

Studies show that 48 algorithms were used in papers. Some algorithms have been used in several studies, and others have only been used in one study. Reports of the period, number of studies, the accuracy mean and standard deviation and type of the algorithms (used for clustering or classification) is in Table 2.

Table 2. Algorithms accuracy

No.	Algorithms or technique	Time period	Studies no.	Accuracy mean (%)	Standard deviation	Classification or clustering	Refs.
1	Bayes Net	2016-2017	2	96.52	0.72	Classification	[23, 24]
2	Logistic Model Tree (LMT)	2012-2016	2	95.59	1.59	Classification	[24, 25]
3	Radial basis function (RBF)	2014-2017	4	94.55	0.88	Classification	[6,7,23, 24]
4	Decision tree	2001-2015	3	93.95	2.37	Classification	[9, 16, 26]
5	J48 (c4.5)	2012-2017	7	92.01	5.25	Classification	[23-25, 27-30]
6	SVM	2012-2017	8	90.32	9.67	Classification	[23, 26, 27, 29-33]
7	Naive Bayes	2012-2016	7	89.48	7.98	Classification	[24, 25, 27-30, 34]

5 - Digital Database for Screening Mammography

Table 2 (continued).

No.	Algorithms or technique	Time period	Studies no.	Accuracy mean (%)	Standard deviation	Classification or clustering	Refs.
8	MLP	2013-2017	7	89.68	9.36	clustering	[7, 23, 24, 27, 28, 33, 35]
9	Random tree	2012-2016	2	87.98	4.88	Classification	[25, 35]
10	K-nearest neighbor	2013-2016	5	86.49	8.69	Classification	[13, 27, 29, 30, 34]
11	Logistic regression	2013-2017	2	85.59	11.59	Classification	[23, 27]
12	Sequential minimal optimization (SMO)	2016	2	79.39	17.59	Classification	[28, 34]
13	Random forest	2013-2017	4	78.14	13.1	classification	[23, 27, 34, 35]
14	FSRAIRS2	2015	1	100	-	Classification	[36]
15	FM−ANN	2015	1	99.8	-	clustering	[7]
16	Multiresolution Nearest Neighbor (MNN)	2015	1	99.22	-	Classification	[7]
17	Decision tree forest (DTF)	2013	1	99.08	-	Classification	[10]
18	BPNN with PSO	2009	1	98.83	-	Classification	[37]
19	Discrete Particle Swarm Optimization (DPSO) - statistical method	2009	1	98.71	-	Classification	[38]
20	SVM-HBA	2009	1	98.6	-	Classification	[39]
21	ANN-HBA	2009	1	98.6	-	Classification	[39]
22	Number Field Sieve (NFS)	2014	1	98.4	-	Classification	[6]
23	Boosted decision tree (BDT)	2013	1	97.7	-	Classification	[9]
24	Rotation forest	2017	1	97.41	-	Classification	[23]
25	K-SVM, hybrid k-means and SVM	2014	1	97.38	-	Classification	[40]
26	Functional Trees FT	2012	1	97.33	-	clustering	[25]
27	DT-HBA	2009	1	97.2	-	Classification	[39]
28	Adaptive neuro-fuzzy inference system	2014	1	95.3	-	Classification	[6]
29	Gray level co-occurrence matrix (GLCM)	2013	1	95.21	-	Classification	[41]
30	Instance-based learning IB1	2012	1	94.92	-	Classification	[25]
31	IBK	2016	1	94.56	-	Classification	[28]
32	Fuzzy SVM	2016	1	94.25	-	Classification	[31]
33	Biclustering + BPNN	2016	1	94	-	clustering	[31]
34	Novel automated segmentation and classification	2015	1	93.24	-	Classification	[26]
35	Fuzzy cerebellar mode NN	2016	1	92.31	-	Classification	[31]
36	"One Rule" 1R	2012	1	90.43	-	Classification	[25]

Table 2 (continued).

No.	Algorithms or technique	Time period	Studies no.	Accuracy mean (%)	Standard deviation	Classification or clustering	Refs.
37	CGPWNN	2014	1	89.57	-	Classification	[42]
38	Libsvm	2016	1	89.3	-	classification	[34]
39	CGPANN	2014	1	89.11	-	Classification	[42]
40	CGPWNN-NL	2014	1	89.11	-	Classification	[42]
41	PSO-SVM	2015	1	84.96	-	Classification	[26]
42	GA-SVM	2015	1	83.8	-	Classification	[26]
43	Ensemble Classifier	2016	1	83.5	-	clustering	[35]
44	Farthest First	2015	1	74	-	clustering	[43]
45	Learning vector quantization	2015	1	67	-	clustering	[43]
46	DBSCAN	2015	1	66	-	clustering	[43]
47	Hierarchy	2015	1	66	-	clustering	[43]
48	Canopy	2015	1	65	-	clustering	[43]

As shown in Table 2, thirteen algorithms have been used in more than one study (Number 1 to 13 in Table 2). Moreover, thirty-four of them were used just in one study. The most commonly used algorithms were SVM (8 study), j48 and Naive Bayes (7 study) and MLP (6 study).

Discussion

In this study, authors have investigated 48 common algorithms on one of the most critical areas in medicine; most of these Studies used the Wisconsin breast cancer dataset that contains essential data and risk factors about breast cancer patients. It is found that among various classifications and clustering algorithms, FSRAIRS2 was better than all other techniques with the highest accuracy rate (100%). Furthermore, Canopy was the least accurate algorithm (accuracy = 65%). Of the 48 algorithms used in the articles, there were 38 in the field of classification and 10 in the field of clustering.

As mentioned above, many algorithms have been used in only one study. Furthermore, the result of one study is not enough to determine the accuracy of an algorithm. Therefore, deciding on the accuracy of an algorithm requires more than one study. In this study, in the algorithms that used in more than one case, the Bayes Net algorithm with the accuracy of 96.52% had the highest accuracy, of course, this algorithm was only used in two studies, but it is reliable because of the low standard deviation. Moreover, the Random Forrest algorithm (78.14%) has the least degree of accuracy among them.

Various studies have been conducted to compare data mining algorithms in the field of breast cancer, but most have compared only two or three algorithms, although, in this study, a large number of data mining algorithms and techniques have been compared.

As shown in Figure 2, data mining in the field of diagnosis of breast cancer has been used more than other areas; this can be due to the importance of timely diagnosis in treatment.

Most of these algorithms were used on Wisconsin database, which seems it is due to data precise, easy to use and free access to this database. Sensitivity and specificity were only expressed in some studies, so we do not mention them in this study. Authors tried to compare performance, efficiency and effectiveness of algorithms in terms of accuracy. Because some of the algorithms have been used in only one study and their results are not very reliable, we are only discussing algorithms that were used in more than one study. In Delen et al. study, the results indicated that the C5 algorithm is the best predictor with 93.6% accuracy and the logistic regression models had the lowest accuracy (89.2%) [44] in another study results show that Farthest First clustering has higher prediction accuracy i.e., 72% than another employed clustering method [43]. In general, in the literature, the accuracy of classification algorithms was greater than clustering. The results were similar in this study.

Another notable point is that several heuristic algorithms have also been used in the studies, reflecting the interest of specialists in the use of data mining in the diagnosis of breast cancer.

In this study, only English language studies were selected, and there is the possibility of losing data from other studies in other languages.

Conclusions

Research literature shows that researchers are interested in comparing algorithms and identifying the best algorithms for use in data mining of breast cancer data. This paper reviewed several studies which are having been varied out for diagnosis, treatment, or prognosis breast cancers. Based on the results of this study, any use of a data mining and knowledge discovery method on a data set requires some discussion on the accuracy of the extracted model on some test data. In terms of time and cost, the accuracy of an algorithm indicates its superiority to other algorithms. Such studies suggest that higher accuracy rates may be achieved by using more and more data points as training data. It shows that even billions of observations may still not be sufficient to capture the behavior of a system under situation accurately.

Considering all the previous papers and our work, the accuracy of diagnosis classification hardly reaches 100%. The study suggests that some algorithms in data mining provide essential tools for researchers and physicians to improve disease prevention, diagnostic methods, and therapeutic plans.

Since there is no tool that automatically diagnoses or prognoses breast cancer, it seems better to use more efficient algorithms for designing automated tools for detecting or predicting breast cancer. However, more work is needed in evaluating the performance of the proposed method on other medical datasets such as histopathological images.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

Authors would like to thanks all the researchers who sent us their articles by email or made it available by free access.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 2018;68(6):394-424.
2. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and mortality and epidemiology of breast cancer in the world. Asian Pac J Cancer Prev. 2016;17(S3):43-6.
3. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J. Breast Cancer: Epidemiology and Etiology. Cell Biochemistry and Biophysics 2015;72(2):333-8.
4. Henriksen EL, Carlsen JF, Vejborg IM, Nielsen MB, Lauridsen CA. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. Acta Radiologica 2019;60(1):13-8.
5. Alqudah AM, Algharib HM, Algharib AM, Algharib HM. Computer aided diagnosis system for automatic two stages classification of breast mass in digital mammogram images. Biomedical Engineering: Applications, Basis and Communications. 2019;31(01):1950007.
6. Ghosh S, Biswas S, Sarkar D, Sarkar PP. A novel Neuro-fuzzy classification technique for data mining. Egyptian Informatics Journal 2014;15(3):129-47.

7. Doreswamy, Salma MU. Fast Modular Artificial Neural Network for the Classification of Breast Cancer Data. Proceedings of the Third International Symposium on Women in Computing and Informatics; 2015, pp. 66-72.
8. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Internal Medicine. 2015;175(11):1828-37.
9. Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Computing and Applications 2013;23(7-8):2387-403.
10. Alqudah A, Alqudah AM. Sliding Window Based Support Vector Machine System for Classification of Breast Cancer Using Histopathological Microscopic Images. IETE Journal of Research 2019. doi: 10.1080/03772063.2019.1583610
11. Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian Journal of Computer Science and Engineering (IJCSE) 2011;2(2):188-95.
12. Holsbach N, Fogliatto FS, Anzanello MJ. A data mining method for breast cancer identification based on a selection of variables. Ciencia & Saude Coletiva 2014;19(4):1295-304.
13. Kuo W-J, Chang R-F, Moon WK, Lee CC, Chen D-R. Computer-aided diagnosis of breast tumors with different US systems. Academic Radiology 2002;9(7):793-9.
14. Lakshmi K, Krishna MV, Kumar SP. Performance comparison of data mining techniques for prediction and diagnosis of breast cancer disease survivability. Asian Journal of Computer Science & Information Technology 2013;3(5):81-7.
15. Li P, Bi T, Huang J, Li S. Breast cancer early diagnosis based on hybrid strategy. Bio-Medical Materials and Engineering 2014;24(6):3397-404.
16. Kuo W-J, Chang R-F, Chen D-R, Lee CC. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. Breast Cancer Research and Treatment 2001;66(1):51-7.
17. Majali J, Niranjan R, Phatak V, Tadakhe O. Data Mining Techniques for Diagnosis and Prognosis of Breast Cancer. International Journal of Computer Science and Information Technologies (IJCSIT) 2014;5(5):6487-90.
18. Mina LM, Isa NAM. A Review of computer-aided detection and diagnosis of breast cancer in digital mammography. J Med Sci. 2015;15:110-21.
19. Oskouei RJ, Kor NM, Maleki SA. c. American Journal of Cancer Research 2017;7(3):610.
20. Rodríguez‐Jiménez JM, Cordero P, Enciso M, Mora A. Data mining algorithms to compute mixed concepts with negative attributes: an application to breast cancer data analysis. Mathematical Methods in the Applied Sciences 2016;39(16):4829-45.
21. Wang H, Yoon SW, editors. Breast Cancer Prediction Using Data Mining Method. IIE Annual Conference Proceedings; 2015: Institute of Industrial and Systems Engineers (IISE).
22. Peters MDJ GC, McInerney P, Baldini Soares C, Khalil H, Parker D. Scoping Reviews. In: Aromataris The Joanna Briggs Institute: The Joanna Briggs Institute; 2017. Available from: https://reviewersmanual.joannabriggs.org/
23. Aličković E, Subasi A. Breast cancer diagnosis using GA feature selection and Rotation Forest. Neural Computing and Applications 2017;28(4):753-63.
24. Al-Hagery MAH. Classifiers' Accuracy Based on Breast Cancer Medical Data and Data Mining Techniques. International Journal of Advanced Biotechnology and Research (IJBR) 2016;7(2): 760-72.
25. Jin SY, Won JK, Lee H, Choi HJ. Construction of an automated screening system to predict breast cancer diagnosis and prognosis. Basic and Applied Pathology 2012;5(1):15-8.
26. Dong M, Lu X, Ma Y, Guo Y, Ma Y, Wang K. An efficient approach for automated mass segmentation and classification in mammograms. Journal of Digital Imaging 2015;28(5):613-25.
27. Radovic M, Djokovic M, Peulic A, Filipovic N. Application of data mining algorithms for mammogram classification. 13th IEEE International Conference on BioInformatics and BioEngineering, Chania, 2013, pp. 1-4. doi: 10.1109/BIBE.2013.6701551.
28. Salama GI, Abdelhalim M, Zeid MA. Experimental comparison of classifiers for breast cancer diagnosis. 2012 Seventh International Conference on Computer Engineering & Systems (ICCES), Cairo, 2012, pp. 180-5. doi: 10.1109/ICCES.2012.6408508

29. Sasikala S, alias Balamurugan SA, Geetha S. A novel feature selection technique for improved survivability diagnosis of breast cancer. Procedia Computer Science 2015;50:16-23.
30. Asri H, Mousannif H, Al Moatassime H, Noel T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Procedia Computer Science 2016;83:1064-9.
31. Chen Y, Huang Q. An approach based on biclustering and neural network for classification of lesions in breast ultrasound. International Conference on Advanced Robotics and Mechatronics (ICARM) 2016; pp. 597-601. doi: 10.1109/ICARM.2016.7606988
32. Gao S, Li H. Breast cancer diagnosis based on Support vector machine. 2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering, Jalarta, 2012, pp. 240-243. doi: 10.1109/URKE.2012.6319555
33. Ghosh S, Mondal S, Ghosh B. A comparative study of breast cancer detection based on SVM and MLP BPN classifier. 2014 First International Conference on Automation, Control, Energy and Systems (ACES), Hooghy, 2014, pp. 1-4. doi: 10.1109/ACES.2014.6808002
34. Diz J, Marreiros G, Freitas A. Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. Journal of Medical Systems 2016;40(9):203.
35. Kaushik D, Kaur K. Application of Data Mining for high accuracy prediction of breast tissue biopsy results. 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), Moscow, 2016, pp. 40-45. doi: 10.1109/DIPDMWC.2016.7529361
36. Saybani MR, Wah TY, Aghabozorgi SR, Shamshirband S, Kiah MLM, Balas VE. Diagnosing breast cancer with an improved artificial immune recognition system. Soft Computing 2016;20(10):4069-84.
37. Huang ML, Hung YH, Chen WY. Neural network classifier with entropy based feature selection on breast cancer diagnosis. Journal of Medical Systems 2010;34(5):865-73.
38. Yeh W-C, Chang W-W, Chung YY. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications 2009;36(4):8204-11.
39. Pham HNA, Triantaphyllou E. An application of a new meta-heuristic for optimizing the classification accuracy when analyzing some medical datasets. Expert Systems with Applications 2009;36(5):9240-9.
40. Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications 2014;41(4):1476-82.
41. Markkongkeaw A, Phinyomark A, Boonyapiphat P, Phukpattaranont P. Preliminary results of breast cancer cell classifying based on gray-level co-occurrence matrix. The 6th 2013 Biomedical Engineering International Conference, Amphur Muang, 2013, pp. 1-4. doi: 10.1109/BMEiCon.2013.6687634
42. Khan MM, Chalup SK, Mendes A. Evolving wavelet neural networks for breast cancer classification. 12th Australasian Data Mining Conference, Brisbane, Australia, 27th-28th November; 2014. Available from: https://pdfs.semanticscholar.org/8d41/7bb42d5aceefa64e92e0495e7ddf5dffcf2c.pdf
43. Devi RDH, Deepika P. Performance comparison of various clustering techniques for diagnosis of breast cancer. 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, 2015, pp. 1-5. doi: 10.1109/ICCIC.2015.7435711
44. 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.