EBPS: Effective Method for Early Breast Cancer Prediction using Wisconsin Breast Cancer Dataset

P R Anisha, B Vijaya Babu

Abstract - Machine considering is a branch of computerized reasoning that contain a dissemination of factual, probabilistic and enhancement systems that enable PCs to "examine" from past illustrations and to run over hard to recognized examples from vast , loud or muddled data units. These abilities are exceptionally pleasantly alluring to logical bundles, principally those that rely on confounded proteomic and genomic estimations. In this paper, we dissected the bosom Cancer actualities to be had from the Wisconsin dataset from UCI gadget learning with the reason for creating exact expectation rendition for bosom growth and proposed Effective Breast Cancer Prediction System. The proposed variant is in examination with introducing approaches in expressions of exactness, specificity and missteps cost.

Keyword: UCI gadget learning

I. INTRODUCTION

Bosom growth is the most well-known most diseases in ladies around the world. It is in like manner the statute thought process of passing on from growth among women all around. The best method to decrease bosom disease passings is recognize it prior. Early forecast requires an exact and reliable visualization strategy that lets in doctors to recognize considerate bosom tumors from threatening ones without going for careful biopsy. Bosom most tumors is the greatest frequently perceived disease and is the fundamental driver of most malignancies passing among ladies worldwide. Like clockwork, some place round the field an instance of bosom disease is recognized among young ladies. At regular intervals, some place inside the worldwide, somebody kicks the bucket from bosom growth [1-3].

We can recognize three sorts of bosom tumors: amiable bosom tumors, in situ malignancies, and obtrusive growths. The dominant part of bosom tumors identified by means of mammography is considerate. They are non-malignant developments and can't unfurl outside of the bosom to different organs. In a couple of occurrences it is hard to recognize certain amiable burdens from threatening injuries with mammography. In the event that the threatening cells have not long experienced the basal film anyway is completely contained in the lobule or the channels, the malignancy is alluded to as in situ or noninvasive. In the event that the malignancy has harmed by means of the basal layer and spread into the incorporating tissue, its miles called obtrusive. In this manner, early identification of bosom most growths is essential.

Today framework considering methodologies is being utilized in an awesome scope of restorative bundles together with recognizing and arranging tumors. It has been utilized first and primary as a helpful asset to most malignancies guess and recognition. Malignancy scientists have at present attempt and relate device acing toward most tumors forecast and investigation. Machine considering by the by pulls in from measurements and possibility, anyway it is profoundly more noteworthy great as it grants derivations or decisions to be made that couldn't in some other case be made the use of customary factual systems.

Foreseeing the aftereffect of confusion is one of the most extreme fascinating and critical commitments in which to expand measurements mining bundles. As the utilization of PC frameworks controlled with mechanized apparatus, enormous volumes of restorative measurements are being assembled and made to be had to the clinical examinations offices. As a final product, Knowledge Discovery in Databases (KDD) [20-24], which incorporates realities mining methodologies, has wind up a well known examinations gadget for logical scientists to select and make the most styles and connections among expansive scope of factors and made them ready to expect the last consequences of a disease the utilization of the old cases put away inside datasets. The objective of this paper is to be expecting the bosom disease the utilization of Wisconsin dataset accumulated shape UCI Machine Learning Repository [19]. A powerful method for early Breast Cancer Prediction (EBPS) is proposed.

The unwinding of the paper is sorted out as takes after: Scetion 2 scrutinizes the past work on bosom malignancy expectation, segment three features the proposed set of guidelines, area four portrays the trial discourse and impacts, and segment 5 examines the ends drawn seen followed by references.

Revised Manuscript Received on December 28, 2018.

P.R. Anisha, Research scholar, Dept of CSE, KLEF University, Guntur, India.

Dr. B. Vijaya babu , Professor , Dept of **CSE**, KLEF University**,** Guntur, India.

II. Relevant Work

Machine picking up learning of, a branch of engineered insight, is an efficient subject stressed with the outline and change of calculations that enable PCs to adjust practices construct absolutely in light of experimental measurements, comprising of from sensor data or databases. Machine Learning (ML) gives methods, methodologies, and devices that can help comprehending indicative and prognostic issues in a determination of logical areas. ML is getting utilized for the examination of the significance of clinical parameters and their combos for forecast, e.g. Forecast of illness movement, extraction of clinical learning for definite outcomes examines, cure arranging and help, and for the general influenced individual administration. Foundation records on propensities in these fields is outfitted all together that logical overseers can hold to supplant their way to deal with the assessment of bosom most tumors danger. The reason for a writing assessment is to genuinely examine and complete an in-power assessment of going before ponders. Immense investigations is been done in the train of Cancer utilizing Data mining and framework picking up learning of methodologies.

Afzan Adam[4] et al. Have built up a programmed bosom most tumors forecast with the guide of joining hereditary arrangement of standards and Back engendering neural network which progress toward becoming developed as speedier classifier model to diminish the analyze time notwithstanding expanding the precision in grouping mass in bosom to either amiable or dangerous. In these two particular purifying methodologies progress toward becoming accomplished on the dataset. This exploration gave a precision of roughly 83.36%.

Kim W et al. [5] has propelled a forecast form the utilization of assistance vector framework. The system made utilization of insights on 679 patients, who experienced bosom most tumors surgery somewhere in the range of 1994 and 2002, have been gathered reflectively from a Korean tertiary instructing center. The accompanying factors have been chosen as fair factors for the prognostic model, by utilizing the set up therapeutic mastery and univariate investigation: histological review, tumor length, assortment of metastatic lymph hub, estrogen receptor, lymph vascular attack, adjacent intrusion of tumor, and scope of tumors. Three expectation calculations, with each utilizing SVM, counterfeit neural network and Cox-relative shot relapse display, were built and in correlation with one another.The examine gave a precision of roughly 89%.

Stop C et al. [6] proposed a special semi-regulated learning set of standards construct absolutely with respect to a chart regularization technique. Advance the examinations depictions changed over the quality articulation realities directly into a diagram shape for semi-administered picking up learning of and included protein interaction data with the quality articulation records to choose practically related quality sets. At that point, the artworks anticipated the repeat of most diseases by utilizing making utilization of a regularization system to the developed chart containing each marked and unlabeled hubs.

Tseng C-J et al. [7] connected 3 framework learning techniques together with help vector gadget, C5.Zero and serious acing gadget were thought about to find fundamental risk components to expect the repeat inclination for cervical tumor. Test outcomes outline that C5.0 demonstrate is the most helpful technique to the disclosure of repeat inclination factors. It gave an exactness of roughly 68%.

A.Punitha [8] et al. 2007 have examined the hereditary calculation and versatile reverberation idea neural network for bosom malignancy examination the utilization of Wisconsin Breast Cancer Data (WBCD). They instructed 699 examples which end up taken from Fine Needle Aspirates (FNA) with 16 missing data, and 683 examples with bosom tumors are utilized in this works of art of which sixty five% end up turned out to be kindhearted and 35% threatening. The creator has likewise analyzed the final product of Adaptive Resonance Theory (ART) with Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Multi Layer Perceptron (MLP), in which the general execution of these blended strategy has now not handiest advanced the precision anyway moreover decreased the time taken to teach the network.

Dr.S.Santhosh baboo and S.Sasikala [9] have accomplished an overview on data digging strategies for quality decision class. This article treated greatest utilized actualities digging procedures for quality decision and most diseases grouping, basically they have concentrated on 4 essential rising fields. They are neural network fundamentally based calculations, contraption picking up information of calculations, hereditary arrangement of guidelines and bunch based absolutely calculations and they have assigned future change in this subject.

Val'erie Bourd`es [10] et al., have presented the thing by means of contrasting counterfeit neural network and strategic relapse. The maker has in examination multilayer perceptron Neural Networks (NNs) with Standard Logistic Regression (SLR) to recognize key covariates affecting on mortality from most malignancies causes, Disease-Free Survival (DFS), and Disease Recurrence the utilization of Area Under Receiver-Operating Characteristics (AUROC) in bosom tumor patients.

Anant Madabhushi and Dimitris N. Metaxas[11] blessing a novel technique to mechanically discover sore edges in ultrasound pix, through consolidating power and surface with exact zone exact know-how along the edge of directional inclination and a deformable frame based rendition. The photos are first separated to discard spot commotion after which differentiate more attractive to stress the tumor regions. For the essential time, a scientific segments of the exact arrangements utilized by radiologists in identifying ultrasonic

bosom sores, prevalently called the "Stavros Criteria" is provided in this paper.

Hamid Karim Khani Zand[12] have been given their review on class of Breast Cancer records might be helpful to anticipate the last consequences of a couple of maladies or discover the hereditary conduct of tumors. To display a near study on data mining methods inside the examination and expectation of bosom most diseases and furthermore an assessment of the forecast of

survivability charge of bosom most tumors patients.

Vikas Chaurasia et.Al.[13] have given their test an anticipation machine for distinguishing bosom growth construct absolutely in light of RepTree, RBF Network and Simple Logistic. In test degree, 10-overlap move approval approach changed into actualized to the University Medical Center, Institute of Oncology, Ljubljana, Yugoslavia database to evaluate the proposed framework exhibitions. The right class charge of proposed gadget is 74.Five%. To set up that the Simple Logistic can be utilized for bringing down the measurement of highlight zone and proposed Rep Tree and RBF Network model might be utilized to achieve quick programmed demonstrative frameworks for various maladies.

Abdelghani Bellaachia, Erhan Guven [14] in their exploration have examined three data mining methodologies: the Naïve Bayes, the lower back-proliferated neural system, and the C4.Five choice tree calculations. The completed forecast exhibitions are like current strategies. In any case, we found out that C4.Five calculation has a greatly improved in general execution than alternate systems with an exactness of 81. 3%.

Xu et al., [15] offered A half breed technique for probabilistic neural system (PNN) and discrete paired adaptation of Particle swarm enhancement. PSO is to a great extent utilized for best selection of qualities and dimensionality decrease. Feed forward neural network is utilized to put into impact neural network arrangements. This approach is practically investigated enormous B-cell lymphoma dataset with 80% class exactness. Conventional methods for prediction of cancer in medical practice can be often imperfect or confusing while molecular level diagnostics with gene expression microarray technique is very precise, objective and efficient for cancer classification.

Won et al., [16] presented a technique using ensemble of neural network classifiers studied from negatively correlated characteristics to accurately categorize cancer and it estimates the functioning of the proposed technique with using three benchmark datasets. Experimental performance proves the ensemble classifier with negatively correlated characteristics provides best recognition rate on these benchmark datasets.

David B.fogel [17] et al. has discussed the evolving neural networks for detecting breast cancer and the related works used for breast cancer diagnosis using back propagation method with multilayer perceptron. In contrast to back propagation David B.fogel et al. found that evolution computational method and algorithms were used often, outperform more classic optimization techniques.

Wiselin Jiji, J. R. Marsilin [18] proposed scheme involves low level feature extraction from images like shape and texture features. Shape features used in this scheme are Zernike moments and Radial Chebyshev moments. Texture features of contrast, energy and run length matrix features are also used with the shape features. These extracted features are then classified using SVM. The output of the SVM is considered as patterns. The similarity between two patterns is estimated as a function of the similarity of both their structures and the measure components

III. EBPS Algorithm

Read the dataset.
Generate a sorted attribute list
Evaluate the interval range using formula 1

$$Interval\ Range = \frac{a\max - a\min}{Group\ size} \quad (1)$$

Where amax is the maximum value for the particular attribute, amin is the minimum value for the particular attribute and Group size is to be fixed by user. Upon the experimentation 3 is giving as maximum accuracy levels when compared with other levels.

Based on the interval range, evaluate the split points using formula 2

Split Point = (Attribute value of Changed Class1+Attribute Value of Canged Class2)/4 (2)

Calculate the Index_info value for the class label by using formula 3

$$Index_info = 1 - \sum_{i=1}^{M} P_i^4 \quad (3)$$

Calculate the Index_infoDvalue for every attribute by using formula 4.

$$Index_info_D = \sum_{j=1}^{N} P_j[1 - \sum_{i=1}^{M} P_i^4] \quad (4)$$

The Index is obtained by finding the difference between Index_info and $Index_info_D$ values by using formula 5

$$Index = Index_info - Index_info_D \quad (5)$$

The greatest Index esteem is considered as the best split point and is the root hub, as appeared in equation 6.

$$Best\ Split\ point = Maximum(Index) \quad (6)$$

Rehash this methodology until the point that each hub closes with a unique class name.

IV. Experimental Discussion and Results

To clarify the proposed calculation in component an example records is taken from the one of a kind dataset with 20 information for preparing cause and 20 records for experimenting with thought process incorporates 9 traits saw through superbness name, appeared in Table 1 and Table 2 individually. In also tables, properties A, B, C, D, E, F, G, H and I speaks to Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses.

Table 1: Sample train dataset for illustration

A	B	C	D	E	F	G	H	I	Class
5	4	4	5	7	10	3	2	1	benign
3	1	1	1	2	2	3	1	1	benign
6	8	8	1	3	4	3	7	1	benign
4	1	1	3	2	1	3	1	1	benign
8	10	10	8	7	10	9	7	1	malignant
1	1	1	1	2	10	3	1	1	benign
2	1	2	1	2	1	3	1	1	benign
2	1	1	1	2	1	1	1	5	benign
4	2	1	1	2	1	2	1	1	benign
1	1	1	1	1	1	3	1	1	benign
2	1	1	1	2	1	2	1	1	benign
5	3	3	3	2	3	4	4	1	malignant
1	1	1	1	2	3	3	1	1	benign
8	7	5	10	7	9	5	5	4	malignant
7	4	6	4	6	1	4	3	1	malignant
4	1	1	1	2	1	2	1	1	benign
4	1	1	1	2	1	3	1	1	benign
10	7	7	6	4	10	4	1	2	malignant
6	1	1	1	2	1	3	1	1	benign
7	3	2	10	5	10	5	4	4	malignant

Table 2: Sample test dataset for illustration

A	B	C	D	E	F	G	H	I	Class
4	1	1	1	2	1	3	1	1	benign
6	1	3	2	2	1	1	1	1	benign
4	1	1	1	1	1	2	1	1	benign
7	4	4	3	4	10	6	9	1	malignant
4	2	2	1	2	1	2	1	1	benign
1	1	1	1	1	1	3	1	1	benign
3	1	1	1	2	1	2	1	1	benign
2	1	1	1	2	1	2	1	1	benign
1	1	3	2	2	1	3	1	1	benign
5	1	1	1	2	1	3	1	1	benign
5	1	2	1	2	1	3	1	1	benign
4	1	1	1	2	1	2	1	1	benign
6	1	1	1	2	1	2	1	1	benign
5	1	1	1	2	2	2	1	1	benign
3	1	1	1	2	1	1	1	1	benign
5	3	1	1	2	1	1	1	1	benign
4	1	1	1	2	1	2	1	1	benign
2	1	3	2	2	1	2	1	1	benign
5	1	1	1	2	1	2	1	1	benign
6	10	10	10	4	10	7	10	1	malignant

The split points computed using formula 2 for the sample dataset is shown in table 3. The decision tree obtained for the sample dataset is shown in Fig. 1, followed by rules obtained.

Table 3: No. of split points computed for the sample dataset

Attribute	No. of Split Points
Clump Thickness	2
Uniformity of Cell Size	3
Uniformity of Cell Shape	3
Marginal Adhesion	2
Single Epithelial Cell Size	2
Bare Nuclei	4
Bland Chromatin	2
Normal Nucleoli	3
Mitoses	3

```
b (1)
  |--- < 3.0
  |    benign
  |--- >= 3.0
  |    i (8)
  |       |--- < 2.25
  |       |    f (5)
  |       |       |--- < 3.25
  |       |       |    malignant
  |       |       |--- >= 3.25
  |       |       |    f (5)
  |       |       |       |--- < 5.0
  |       |       |       |    benign
  |       |       |       |--- >= 5.0
  |       |       |       |    h (7)
  |       |       |       |       |--- < 2.75
  |       |       |       |       |    benign
  |       |       |       |       |--- >= 2.75
  |       |       |       |       |    malignant
  |       |--- >= 2.25
  |       |    malignant
```

Fig. 1: Decision tree obtained for the sample dataset

Rules:

1: If [(b < 3.0)] Then (prediction = benign)

2: If [(b >= 3.0) and (i < 2.25) and (f < 3.25)] Then (prediction = malignant)

3: If [(b >= 3.0) and (i < 2.25) and (f >= 3.25) and (f < 5.0)] Then (prediction = benign)

4: If [(b >= 3.0) and (i < 2.25) and (f >= 3.25) and (f >= 5.0) and (h < 2.75)] Then (prediction = benign)

5: If [(b >= 3.0) and (i < 2.25) and (f >= 3.25) and (f >= 5.0) and (h >= 2.75)] Then (prediction = malignant)

6: If [(b >= 3.0) and (i >= 2.25)] Then (prediction =

malignant)

BPS	130	36	2	3
EBPS	130	38	0	3

Results

The general execution of the EBPS calculation changed into as contrasted and the elective decision tree and non-choice tree calculations the utilization of Wisconsin Breast Cancer measurements set accumulated from UCI Machine Learning Repository. All results that we at last record depend on ten times cross approval. Table 4 gives an inside and out case on the dataset. Amid the experimentation, the dataset is part in to tutoring set with 75% of actualities and experimenting with set with 25% of data. The arrangement of principles is connected the utilization of JAVA programming dialect. Computed split points for the dataset; attribute wise is listed in Table 5. Comparison of split points between BPS and EBPS is listed in Table 5. It is clearly visible that in EBPS number of split points is less compared with BPS.

Table 4: Breast Cancer dataset description

Dataset	Records	No.of attributes	Train	Test
Wisconsin	682	10	511	171

Table 5: Split points comparision of BPS and EBPS

Attribute	BPS	EBPS
Clump Thickness	13	14
Uniformity of Cell Size	13	10
Uniformity of Cell Shape	15	10
Marginal Adhesion	13	10
Single Epithelial Cell Size	17	10
Bare Nuclei	15	12
Bland Chromatin	11	10
Normal Nucleoli	15	14
Mitoses	9	4

A perplexity grid is a work area that is frequently used to clarify the general execution of a class form on an arrangement of investigate measurements for which the best possible qualities are known.True Positives (TP) are the effectively foreseen fine qualities which implies that the expense of genuine polish is certain and the charge of anticipated style is in like manner yes. Genuine Negatives (TN) are the viably anticipated awful qualities which implies that the cost of real style isn't any and cost of anticipated class is additionally no. False Positives (FP) are while genuine style isn't any and expected polish is indeed, False Negatives (FN) are while genuine class is certain anyway anticipated heavenliness in no, appeared in Table 6. The comparison between correct and in correct prediction are shown in Table 7. In Table 7, CP represents correct predictions and ICP represents in correct prediction.

Table 6: Comparison of confusion matrix with BPS

Algorithm	TP	TN	FP	FN

Table 7: Comparison of correct and In-correct predictions with BPS

Algorithm	CP	ICP
BPS	166	05
EBPS	168	03

Accuracy is the most natural execution degree and its miles extremely a proportion of effectively anticipated that comment would the aggregate perceptions, appeared in Table 8 and graphically spoke to in Fig. 2. The top notch accuracy is 100.0, while the most noticeably bad is 0.0. Accuracy may be assessed the utilization of segments 7, demonstrated underneath.

Accuracy = (TP+TN) / (TP+FP+FN+TN) (7)

Table 8: Performance comparision (Accuracy) with existing approaches

Algorithm	Accuracy (%)
HPBCR	85
Naïve Bayes	84.5
ANN	86.5
C4.5	86.7
Bayes Net	47.67
jRip	74.42
REP Tree	74.42
KNN	74.42
BPS	97.07
EBPS	98.24

Specificity (SP) is figured as the amount of exact unpleasant expectations separated by utilizing the aggregate assortment of negatives. It is otherwise called genuine negative charge (TNR), demonstrated in Table 9 and graphically appeared in Fig. 3. The agreeable specificity is a 100.0, while the most exceedingly bad is 0.0. Specificty might be assessed the utilization of framework 8, appeared underneath:

Specificty = TN / (TN+FP) (8)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

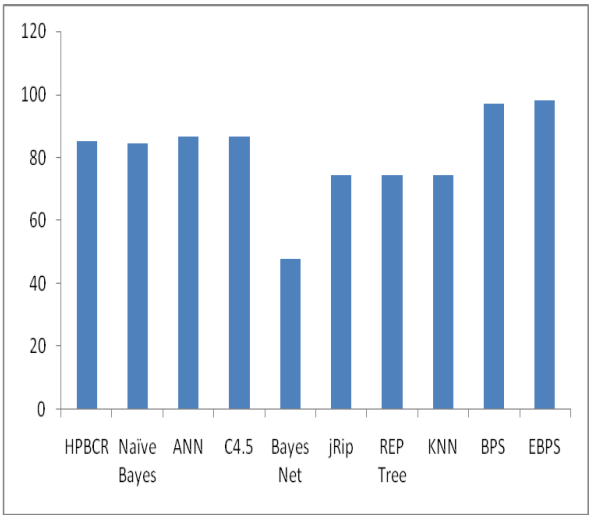

Fig. 2: Performance comparision (Accuracy) with existing approaches

Table 9: Perfromance comparision (Specificity) with existing approaches

Dataset	HPBCR	C4.5	ANN	BPS	EBPS
Wisconsin	93	90.7	92.8	94.73	100

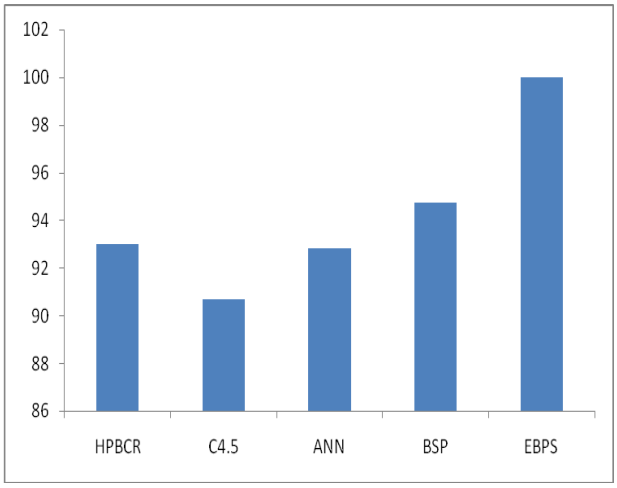

Fig. 3. Perfromance comparision (Specificity) with existing approaches

Error rate is calculated and is shown in Table 10 and graphically represented in fig. 4. The best error rate is 0.0, whereas the worst is 100.0. Error rate will be evaluated using formula 9, shown below

Error Rate = (FP+FN) / (TP+FP+FN+TN) (9)

Table 10: Perfromance comparision (Error rate) with existing approaches

Algorithm	Error rate (%)
HPBCR	15
Naïve Bayes	15.5
ANN	13.5
C4.5	13.3
Bayes Net	52.33
jRip	25.58
REP Tree	25.58
KNN	25.58
BPS	2.93
EBPS	1.76

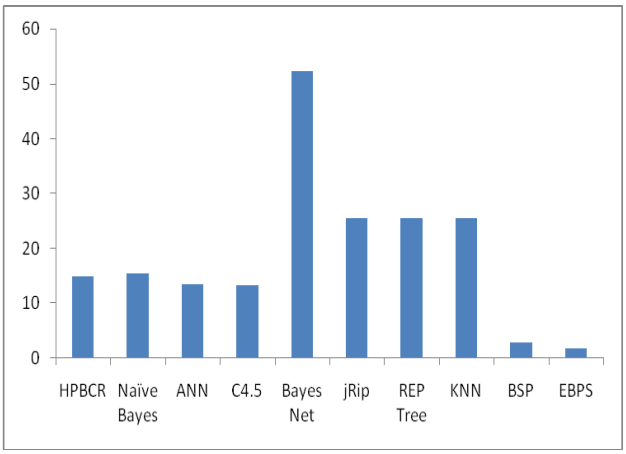

Fig. 4: Error rate comparison

V. Conclusions

An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for Medical applications. In this paper, Effective Breast Cancer Prediction System is proposed. The algorithm uses Wisconsin breast cancer dataset collected from UCI Machine Learning Repository. Performance measures such as accuracy, specificity and error rate are comapred with existing approaches. The performance of EBPS shows the high level compare with other classifiers.

REFERENCES

1. Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M and Razavi AR, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence", Journal of Health & Medical Informatics, 2013.
2. Mohammad R. Mohebian, Hamid R. Marateb, Marjan Mansourian, Miguel Angel Mañanas, and Fariborz Mokarian, "A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning, Journal of Computational and Structural Bio-Technology Elsevier", 2016.
3. Siddhant Kulkarni and Mangesh Bhagwat, "Predicting Breast Cancer Recurrence using Data Mining Techniques", International Journal of Computer Applications, 2015.
4. Afzan Adam, Khairuddin Omar, "Computerized Breast Cancer Diagnosis with Genetic Algorithms and Neural Network" "-fitt.mmu.edu.my/caiic/papers/afzaniCAIET.pdf.
5. Kim W, Kim KS, Lee JE, Noh D-Y, Kim S-W, Jung YS, et al. Development of novel breast cancer recurrence prediction model using support vector machine. J Breast Cancer 2012;15:230–8
6. Park C, Ahn J, Kim H, Park S (2014) Integrative Gene Network Construction to Analyze Cancer Recurrence Using Semi-Supervised Learning. PLoS ONE 9(1): e86309. https://doi.org/10.1371/journal.pone.0086309.
7. Tseng C-J, Lu C-J, Chang C-C, Chen G-D. Application of machine learning to predict the recurrence-proneness for cervical cancer. Neural Computer & Application 2014;24: 1311–6.
8. Dr.C.P.Sumathi, Dr.T.Santhanam, A.Punitha "Combination of genetic algorithm and ART neural network for breast cancer diagnosis", 2007 -- Asian Journal of information technology, Medwell journals.

9. Dr.Santhosh baboo, S.Sasikala "A Survey on data mining techniques in gene selection and cancer classification"-April 2010 International journal of Computer science and information technology.
10. Val´erie Bourd`es, St´ephane Bonnevay, Paolo Lisboa, R´emy Defrance, David P´erol, Sylvie Chabaud, Thomas Bachelot, Th´er`ese Gargi and Sylvie Negrier "Comparison of Artificial Neural Network with Logistic regression as Classification Models for Variable Selection for Prediction of Breast Cancer Patient outcomes".
11. A. MADABHUSHI, D. METAXAS., (2003). "Combining low-, high-level and Empirical Domain Knowledge for Automated Segmentation of Ultrasonic Breast Lesions". IEEE Transactions Medical Imaging, Vol. 22, No. 2, 2003, pp: 155–169.
12. Hamid Karim Khani Zand "a comparative survey on data mining techniques for breast cancer diagnosis and prediction" Indian Journal of Fundamental and Applied Life Sciences, Vol.5 (S1), pp. 4330-4339, 2015.
13. Vikas Chaurasia , Saurabh Pal "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability" International Journal of Computer Science and Mobile Computing, Vol.3 Issue.1, January-2014, pg. 10-22.
14. Abdelghani Bellaachia, Erhan Guven, "Predicting Breast Cancer Survivability Using Data Mining Techniques", http://ieeexplore.ieee.org/document/5608818/.
15. Rui Xu, Xindi Cai, Donald C. , Wunsch II. Gene Expression Data for DLBCL Cancer Survival Prediction with A Combination of Machine Learning Technologies; In Proceedings of the IEEE International Conference on Medicine and Biology, 2005, p. 894-897.
16. Hong-Hee Won, Sung-Bae Cho. Paired neural network with negatively correlated features for cancer classification in DNA gene expression profiles; In Proceedings of the International Joint Conference on Neural Networks; 3; 2003; p. 1708 – 1713.
17. David B.Fogel, Eugene C, Wasson, Edward M.Boughton "Evolving neural networks for detecting breast cancer". 1995 Elsevier Science Ireland Ltd.
18. Wiselin Jiji,J. R. Marsilin, "Diagnose the Stages of Breast Cancer using SVM", International Journal of Computer Applications (0975 – 8887) Volume 38– No.11, January 2012.
19. https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original).
20. J. Han: Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2011.
21. Arun k pujari: Data Mining Techniques", Universities Press 2004.
22. Hastie TJ, Tibshirani, RJ, Friedman JH, "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Second Edition, Springer; 2009.
23. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy, "Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies", The MIT Press, 2015.