



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Analysis of Breast Cancer Disease Prediction Using the Effects of Life Quality Attributes

Mrs. Ambika L G¹

Assistant professor,
Information Science and Engineering,
SJC Institute of Technology,
Chickballapur, Karnataka.

Dr. T N Anitha²

Professor,
Computer Science and Engineering,
Sir M. Visvesvaraya Institute of Technology,
Bengaluru, Karnataka.

Dr. Jayasudha K³

Associate Professor,
Information Science and Engineering,
Atria Institute of Technology,
Bengaluru, Karnataka.

Abstract— One of the most common cancers globally is breast cancer in women. According to data from reputable health organizations, breast cancer, therefore, causes a serious human setback that raises the death rate. Any health issue that is detected early may have a speedy recovery. In this research study, many factors that characterize the state of human existence now were used to analyze breast cancer. This uses machine learning approaches to anticipate the occurrence of breast cancer by identifying and analyzing human life quality variables. The information used in this evaluation inquiry was acquired from a sample of Bangalore area people as well as information that was factually stored online. Classification methods the measurements analysed to present the results utilising the various error measures RMSE, MAE, and Kappa measuring esteem include accuracy, recall, precision, and F-measures. K-means++ clustering computation and the C4.5 decision tree techniques are used in this investigation. The accuracy of the suggested machine learning model's C4.5 computation was 98.97%, and it's K-means++ grouping accuracy was 97.98% over the real-time data.

Index Terms: Prediction, breast cancer illness, K-means++ Clustering, Life quality attributes, C4.5 Algorithm.

I. Introduction

The most prevalent cancer among women is breast cancer. When normal restrictions on cell death and cell division are absent in the breast tissue, malignant tumors (tumours) form. Consequently, breast cancer is the term used to describe malignancy in breast tissue. In women, it is the cancer kind that is most common worldwide. Around 10% of all women, throughout various phases of life, find this to be distressing. Even though most breast cancers have unknown exact origins, scientists have identified a few risk factors that increase a woman's likelihood of acquiring breast cancer. Age, family history, and genetic risk are some of these characteristics. Despite being the second-leading cause of cancer mortality in women, breast cancer has a remarkable survival rate. 97% of women continue to survive after 5 years or more because of early diagnosis [2]. Chemotherapy and hormone therapy are two frequent forms of treatment that are administered to cancer patients as part of organized regimens. Although the majority of women have many recognized risk factors for breast cancer, they have never been diagnosed with the disease. The two biggest risk factors for breast cancer are being female and becoming older [3]. Breast cancer may actually have more than one cause in reality.

An efficient and effective way to extract knowledge from unstructured data is through data mining (DM). Different challenges that raw data encounters prevent the use of classic or conventional methodologies for knowledge extraction [1, 2]. Different data types should be supported by DM in all configuration forms. Data mining in medicine or healthcare has several facets and is dedicated to both data mining and medicine [3, 4]. The estimation of health expenses, diagnosis, and visualisation; the extraction of hidden values from biological information; and the discovery of a link between illnesses and

medications are among the significant medical qualities that are mentioned. A significant quantity of raw data that takes views has been produced as a result of the advancement of information storage technologies [5, 6]. These opinions on advanced algorithms and current storage technology are growing. Normal raw data may be used to learn useful and beneficial things. The minor extraction of accurate, already obscure, and practical viewpoint information from the data is known as knowledge discovery [7]. Reasonably time-multifaceted character, understandability, accuracy, and beneficial outcome are appealing qualities for extracted knowledge. The knowledge that is extracted provides fresh data for new applications and discoveries. DM was the first.

Breast cancer must emerge as the leading factor in women's deaths in industrialized nations. The most efficient strategy to lower the death rates brought on by breast cancer will be through early identification [5]. The distinction between benign and dangerous breast cancers may be made by doctors without a surgical sample, but early identification calls for a precise and steadfast diagnosis technique. One of the most interesting and difficult tasks where to develop data mining applications is predicting the outcomes of disease. Research organizations have access to vast amounts of medical data thanks to computers and automated technologies [7]. The process of information discovery that uses intellectual approaches to find patterns in data is known as data mining. In this work, C4.5 decision trees and the k-means++ clustering method have been used to diagnose and predict breast cancer.

Breast cancer prognosis indicates when breast cancer is likely to recur in a patient's health based on their current state of health [8]. Breast cancer diagnostic separates benign from malignant breast lumps. This study's goal is to examine how human life scenario characteristics affect breast cancer treatment, post-breast cancer survival rates, and breast cancer prevention by keeping some characteristics of the disease at a certain level.

II. LITERATURE REVIEW

Nishtha Hooda suggested the disease's mortality aspect and its correlation with numerous risk factors, such as demographics, the percentage of pesticide residue in the water and soil, the way of life of women in the afflicted area, water consumption, and the patient's level of pesticide exposure [6].

T. Panduranga Vital suggested K-means, hierarchical and multidimensional scaling, as well as other statistical and clustering techniques (MDS). According to statistics data, more women than males in zone 1 of AP are affected by cancer. Breast cancer is the most common type of cancer in women, whereas lung cancer is the most common type in males. The research also yields intriguing findings about cancer-related lifestyle choices and practices. For predicting cancer, unsupervised machine learning techniques also produce strong results[10].

The probabilistic approach, which surpasses the hierarchical clustering technique, was proposed by S. M. Halawani et al. when the majority of the cluster points are grouped into one. However, this might be a disadvantage as a result of the wrong choice of distance measurement.[13]

Sahar A. Mokhtar has examined numerous classification and categorizing processes. The approaches employed for its analysis include the support vector machine, decision tree algorithm, and artificial neural network methodology [37]. The Chi squared approach and the pruning technique are used to create the decision tree algorithm. With the use of a polynomial kernel, it was used to discover the best pattern for an artificial neural network and create a support vector machine.

Milon Islam introduces the Support Vector Machine and K Nearest Neighbors, which are the supervised machine learning approaches for breast cancer diagnosis by training its characteristics, and presents a unique method for the prediction of breast cancer. The suggested technique employs 10-fold cross validation to provide reliable results. The UCI machine learning repository was used to obtain the Wisconsin breast cancer diagnostic data set. Accuracy, sensitivity, specificity, false discovery rate, false omission rate, and Matthews correlation coefficient are taken into account while evaluating the performance of the suggested system [38].

III. Implementation and System Design

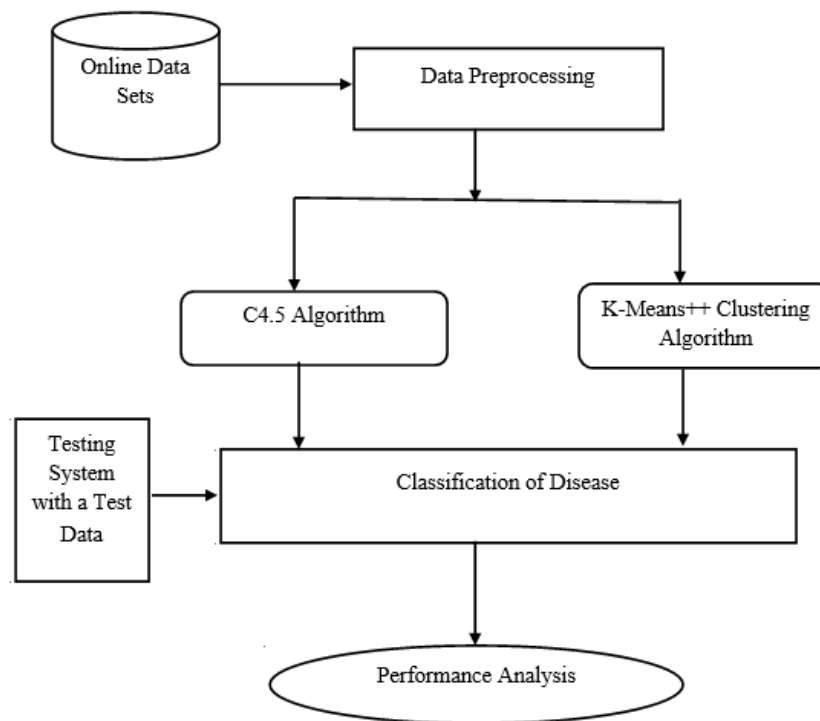


Fig. 1 Functional components of a system for predicting breast cancer disease

For the study and analysis, the data from the online repository is taken into account. This information includes a wide range of characteristics, including a family history of breast disease, Habits like smoking and consumption of tainted food, alcohol, and a hot beverage, among others. To prepare the data for usage with data mining techniques, preprocessing is done on the data. The study investigation employed the c4.5 method and K-means clustering approaches [26] [27]. The system is trained and developed using these characteristics with the goal of detecting and preventing breast disease. Depending on how the features of breast illness are impacted, these attributes are crucial in the diagnosis of breast disease.

For the purpose of developing and testing a framework for forecasting breast disease, the information taken into account is split 80:20. The C4.5 decision tree model is then developed for more research and evaluation, with the weighting of a large example and the order tactics around it. The bunching model is created using the K-implies grouping technique, and the model's performance is examined using several accuracy measures, such as precision, recall, and specificity. He table I in this research study and analysis lists the aspects of life quality that were taken into consideration.

TABLE 1. Breast Cancer creating life quality attributes and its ranges

Sl n.	Life Quality Attribute	Ranges	Risk Level
1.	Age	Less than 30	2
		31 to 40	3
		41 to 60	4
		Above 60	5
2.	Gender	Female	5
		Male	1
3.	Living Area	Urban	4
		Rural	2
4.	Habits	Smoking	3
		Alcohol	4
5.		Chewing	5
		Hot beverage	2
6.	Education	Uneducated	5
		School / College	3
7.	Occupational Hazards	Radiation Exposure	5
		Chemical Exposure	4
		Sunlight Exposure	2
		Thermal Exposure	2
8.	Family History of Cancer	Yes	5
		No	2
9.	Past History of Disease	Nil	1
		Still Continues	4
10.	Physical Activity	Less	5
		Medium	2
		Heavy	1
11.	Blood Group	A	4
		o	3
		B	2
		AB	1
12.	Left/Right side occurrence of tumors	Left	3
		Right	1
13.	Stress	No emotion control	2
		Rare control	4
		Heavy control over emotions	5
14.	Present Problem	Benign	1
		Malignant	4

Data sets that cause breast disease and data sets that don't cause illness are divided into two categories. Making use of the k-means++ clustering technique. The collection's underlying focal point is determined by using the important example weightage, which is separately established and whose mean value is used. In this approach, the estimation of "k" stands for groups, whereas the estimation of "D" stands for informative collections with "n" objects. Choose the arrangement of mean extents from the well-known example weightage as the first stage in this methodology, and then select the commencement or beginning point. The second step is to map the best appropriate cluster for each item using the mean value of significant pattern weighting.

Pseudo Code of K-Means++ clustering Algorithm

Step 1: Randomly select the first centroid (c_j) from the data points.

Step 2: For each data point compute its distance from the nearest, previously chosen centroid.

Step 3: Compute distance of all points in the dataset from the selected centroid. The distance of x_i point from the farthest centroid can be computed by

$$d_i = \max(j:1 \rightarrow m) \| X_i - C_j \|^2$$

Where d_i : Distance of X_i point from the farthest centroid.

m : number of centroids already picked.

Step 4: Make the point x_i as the new centroid that is having maximum probability proportional to d_i .

Step 5: Stop if no data points were changed; otherwise, go back to step 3 and repeat again.

B. C4.5 Algorithm

The data gain percentage, which is determined by entropy, is what the algorithm C4.5 depends on. At each hub in the tree, the test highlights are selected using the data gain proportion metric. Component (property) determination measure is the phrase used to describe such a measure. The test feature for the current hub is chosen to be the character with the most stunning data increase proportion. Assume that D is a collection that contains $(D_1 \dots D_j)$ data samples [33]. Assume that the class mark characteristic, C_i (for $i = 1, \dots, m$), has m distinctive characteristics that define m -specific classes. Give D_i a chance to take more tests than D in class C_i . The typical information needed to order a particular sample is provided by

$$\text{SplitinfoA}(D) = - \sum (|D_j|/|D|) * \log_2(|D_j|/|D|)$$

Where

$$\text{Gain} = \text{Info}(D) - \text{InfoA}(D)$$

$$\text{Info}(D) = - \sum p_i \log_2(p_i)$$

and

$$\text{Info A}(D) = - \sum (|D_j|/|D|) * \text{Info}(D_j)$$

Where

P_i = probability, D = data Set, S is a subattribute of the attribute, and $(|D_j|/|D|)$ serves as the j th partition's weight. $\text{Gain}(A)$ is, in other words, the anticipated decrease in entropy brought on by knowing the value of feature A .

Pseudo Code of C4.5 decision tree Algorithm :

//INPUT: A dataset with attribute values D

Step1: tree = { }

Step2: If D meets other stopping requirements or is "Pure," then

Step3: Stop

Step4: end if

Step5: \forall Attribute $s \in D$ do

Step6: If we split based on 's,' compute information-theoretic criteria.

Step7: end for

Step8: s_{best} = Best characteristic based on the computations made above

Step9: tree = Make a decision node that tests the root s_{best} option.

Step10: D_v = sub-datasets that D induced based on s_{best}

Step11: $\forall D_v$ do

Step12: tree $v=C4.5(D_v)$

Step13: Connect tree $_v$ to the appropriate tree branch.

Step14: end for

Step15: Return Tree

IV. Results and Discussions

Equation (1) & (4) provides the mathematical representations for the three statistical techniques of precision, recall, accuracy, and F-measure used to examine the performance of the breast disease prediction system. With test data, the system's performance is evaluated, and accurate and incorrect categorizations are noted to calculate the system's accuracy. This system uses occurrences that have been classified favorably and negatively to determine their accuracy.

$$\text{Recall} = \frac{\text{True Positive}(TP)}{\text{True Positive}(TP) + \text{False Negative}(FN)}$$

(1)

$$\text{Precision} = \frac{\text{True Positive}(TP)}{\text{True Positive}(TP) + \text{False Positive}(FP)}$$

(2)

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

(3)

$$F - \text{Measure} = \frac{2 * \text{Re call} * \text{Pr ecision}}{\text{Pr ecision} + \text{Re call}}$$

(4)

A. Mean Absolute Error

The equation provides a mathematical description of mean absolute error (MAE) (5). The absolute difference between

$$MAE = \frac{1}{N} \sum_{j=1}^n |y_i - y'_i|$$

expected and actual outcomes is measured by the mean test.

(5)

B. RMSE (Root Mean Squared Error)

Equation(6) is used to determine the size of RMSE. It is the average squared difference between the forecast and the actual observation.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_i - y'_i)^2}$$

(6)

C. Kappa Statistics

The representation of the equation for estimating kappa statistics is (7). Kappa statistics gives information to identify the extent to which doctors influence one another beyond what you may expect to observe based only on chance.

$$Kappa = \frac{y_i - y'_i}{1 - y'_i}$$

(7)

Table II. K-Means++ Clustering Statistical Values

Cluster	TP	FP	Precision	Recall	F - Measure
Cluster 1	1	0.001	0.998	0.9991	0.999
Cluster 2	1	0	0.998	1	1
Cluster 3	1	0	1	0.993	0.994
Cluster 4	0.002	0	0.994	1	0.996
Cluster 5	0.932	0.001	1	0.996	1

Table II presents the statistical results of many metrics produced from the k-means++ clustering method. The true positive values are relatively near to one, according to the statistical information in the table. The near-zero or zero number of false positives shows that this method provides an improved early diagnosis of breast cancer sickness.

Table III. Statistical Values of C4.5

Cluster	TP	FP	Precision	Recall	F - Measure
Breast cancer Malignant	0.997	0.048	0.946	0.963	0.956
Breast cancer Benign	0.989	0.059	0.998	0.982	0.949
Weighted average	0.979	0.049	0.974	0.978	0.956

The statistical values of various accuracy metrics utilising the C4.5 decision tree technique are shown in Table III. The real positive values and false positive values in this method are both extremely close to one. These findings demonstrate that a life quality-based early diagnostic strategy performs better in predicting breast cancer illness.

Table IV. Error Statistics

Cluster	K-Means++	C4.5
Root mean squared error (RMS)	0.0256	0.2241
Mean absolute error (MAE)	0.0082	0.1730
Kappa statistic (KS)	0.9758	0.9108

The error measurements produced by the categorization processes are shown in Table IV. It is employed to verify the veracity of the analysis's findings. This research uses three calibrated error measures, and the error statistics clearly demonstrate how well this early diagnostic strategy predicts breast cancer.

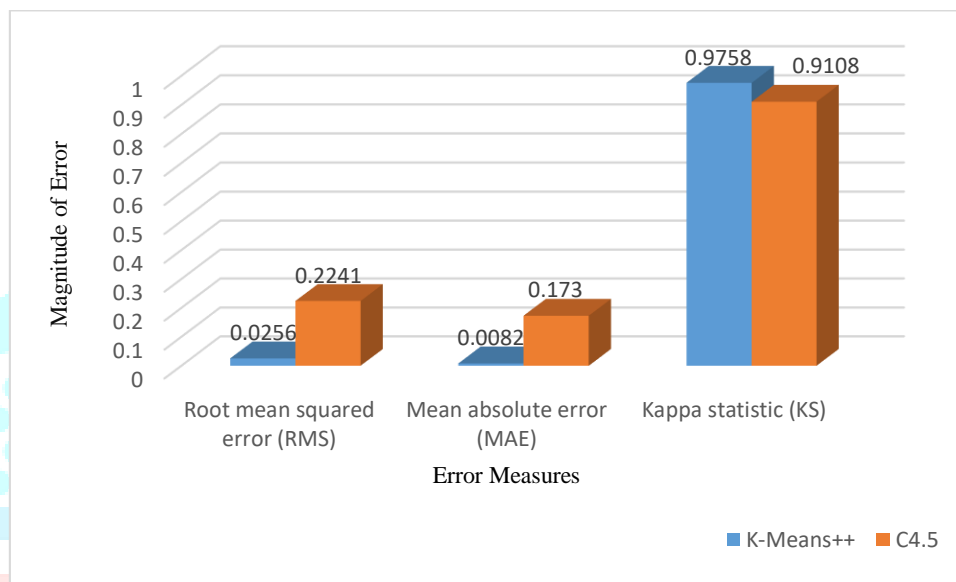


Fig. 2 Representations of error measurements

It is important to recalculate the separation between each data point and the newly discovered cluster centres. The erroneous readings are visually represented in Figure 2. This illustration makes it clear that the k-means++ technique results in lower RMSE when compared to the C4.5 strategy. The following error measure kappa statistic favours the k-means++ method over the C4.5 decision tree strategy.

V. Accuracy Evaluations

Cluster	K-Means++	C4.5
Accuracy	97.98%	98.87%

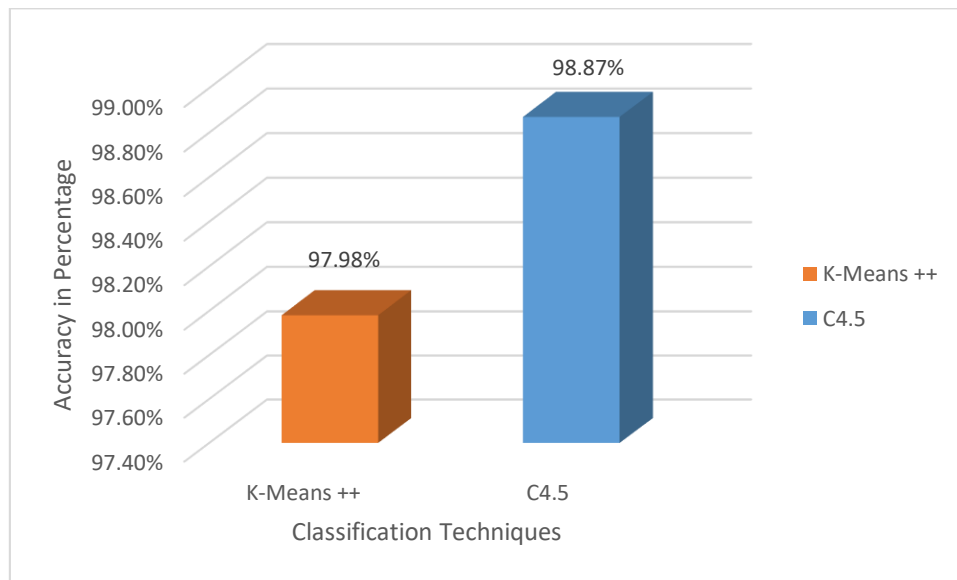


Fig. 3 Accuracy representations

Figure 3 illustrates classification process accuracy in a visual manner. These illustrations demonstrate that the c4.5 technique outperforms K-means++ in terms of accuracy and precision. The data employed in the decision tree technique and the variety and type of data contained in the various clusters are to blame for this.

V. CONCLUSION

Major essential parameters taken into account in this research are lifestyle factors and their ranges with various levels of risk. This study uses the categorization techniques c4.5 decision tree and the k-means++ mechanism to more precisely forecast breast cancer sickness. When compared to the K-means++ clustering approach, the life quality characteristics and the c4.5 algorithm offer superior accuracy. The features targeted for prediction in this study have a significant impact on breast cancer diagnosis, prognosis, and prevention. This study aids both the patient and the doctor in projecting breast cancer illness early and affordably. The improvement of life quality characteristics and the development of a framework for early illness identification are the next steps. The ongoing development of this study include improving life quality features and developing a framework for multiple illness early diagnosis.

REFERENCES

- [1] Hussein AttyaLafta, Noor KdhimAyoob, Asraa Abdullah Hussein. "Breast Cancer Diagnosis Using Genetic Algorithm for Training Feed Forward Back Propagation", International Conference on New Trends in Information & Communications Technology Applications- (NTICT'2017).
- [2] Sureshkumar "Diagnosis Prognosis and Prevention of Breast Cancer Based on Present Scenario of Human Life" , 2018 International Conference on Communication, Information & Computing Technology (ICCICT), Feb. 2-3, Mumbai, India.
- [3] Nishtha Hooda "Prediction of Malignant Breast Cancer Cases using Ensemble Machine Learning: A Case Study of Pesticides Prone Area" IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS 2020.
- [4] Williams K, Idowu PA, Balogun JA, Oluwaranti A. Breast cancer risk prediction using data mining classification techniques. Transactions on Networks and Communications. 2015; 3(2):1–11.
- [5] Nishtha Hooda, Ruchika Gupta, "Prediction of Malignant Breast Cancer Cases using Ensemble Machine Learning: A Case Study of Pesticides Prone Area" IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS 2020.
- [6] Saravana Kumar K, Arthanaraj AM Evaluate the multiple breast cancer factors and calculates the risk by software tool breast cancer risk evaluator. Indian Journal of Science and Technology. 2015 Apr;8(S7):686–91.
- [7] Ennour-Idrissi K, Ayotte P, Diorio C. Persistent Organic Pollutants and Breast Cancer: A Systematic Review and Critical Appraisal of the Literature. Cancers. 2019 Aug;11(8):1063.

- [8] Gray JM, Rasanayagam S, Engel C, Rizzo J. State of the evidence 2017: an update on the connection between breast cancer and the environment. *Environmental Health*. 2017 Dec;16(1):94
- [9] L'otsch, J'orn, et al. "Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy." *Breast cancer research and treatment*. 2018 171.2:399-411.
- [10] T. Panduranga Vital, K. Dileep Kumar "Analysis of Cancer Data Set with Statistical and Unsupervised Machine Learning Methods" Springer Nature Singapore Pte Ltd. 2019.
- [11] Eiko Kai et al., "Empowering the healthcare worker using the Portable Health Clinic", *IEEE Transactions*, DOI 10.1109/AINA.2014.108, 2014.
- [12] Chandresh Arya, Ritu Tiwari. "Expert System for Breast Cancer Diagnosis: A Survey", 2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Jan. 07 – 09, 2016, Coimbatore, INDIA.
- [13] Hussein AttyaLafta, Noor KdhimAyoob, Asraa Abdullah Hussein. "Breast Cancer Diagnosis Using Genetic Algorithm for Training Feed Forward Back Propagation", *International Conference on New Trends in Information & Communications Technology Applications- (NTICT'2017)*.
- [14] Dana Bazazeh, Raed Shubair. "Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis" *IEEE Conference* (2016).
- [15] S M Halawani "A study of digital mammograms by using clustering algorithms" *Journal of Scientific & Industrial Research* Vol. 71, September 2012, pp. 594-600.
- [16] Rajashree Dash "A hybridized K-means clustering approach for high dimensional dataset" *International Journal of Engineering, Science and Technology* Vol. 2, No. 2, 2010, pp. 59-66.
- [17] Sapiah binti sakri, Nuraini binti abdul rashid, "Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction" *IEEE* June 20, 2018. 2843443VOLUME 6, 2018,
- [18] Daniel Duong, Morgan George, Brendan Abraham, "Using Patient-Generated Health Data to Facilitate Preoperative Decision Making for Breast Cancer Patients", 978-1-5386-1848-6/17/\$31.00 ©2017 IEEE, 2017.
- [19] Varalatchoumy.M 1, Ravishankar.M 2, "Comparative Study of Four Novel Approaches Developed for Early Detection of Breast Cancer and its Stages", *Proceedings of the ICICI 2017, IEEE Xplore Compliant - Part Number: CFP17L34-ART*, ISBN: 978-1-5386-4031-9, 2017.
- [20] Abdelouhab Zeroual, Sara Haddou Bouazza, Khalid Auhmani, "Selecting significant marker genes from microarray data by filter approach for cancer diagnosis", *Published by Elsevier B.V. Selection and peer-review under responsibility of International Neural Network Society Morocco Regional Chapter*. 10.1016/j.procs.2018.01.126, 2018.
- [21] Naresh Khuriwall, Nidhi Mishra, "Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm" 978-1-5386-1138-8/18/\$31.00 ©2018 IEEE, 2018.
- [22] Preetha R1, S Vinila Jinny 2, "Margin based feature selection: An algorithmic approach for a set of attributes extrication," *Proceedings of 2017 IEEE International Conference on Circuits and Systems (ICCS2017)*, 2017.
- [23] Koya Kawashima, Wenjun Bai, Changqin Quan, "Text Mining and Pattern Clustering for Relation Extraction of Breast Cancer and Related Genes" 978-1-5090-5504-3/17/\$31.00 ©2017 IEEE SNPD 2017, June 26-28, 2017, Kanazawa, Japan, 2017
- [24] Soolmaz Sohrabi, Alireza Atashi, Ali Dadashi, "A Comparative Study of Multilayer Neural Network and C4.5 Decision Tree Models for Predicting the Risk of Breast Cancer" Atashi, et al. *Arch Breast Cancer* 2018; Vol. 5, No. 1: 11-14, 2018.
- [25] S. Di Meo 1, P.F. Espin-Lopez 1, "Experimental Validation of the Dielectric Permittivity of Breast Cancer Tissues up to 50 GHz", *IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes (IMWS-AMP 2017)*, 20-22 September 2017, Pavia, Italy, 2017.
- [26] Usman Naseem; Junaid Rashid; Liaqat Ali; Jungeun Kim; Qazi Emad Ul Haq; Mazhar Javed Awan; Muhammad Imran "An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers", *IEEE Access* 2022, Volume 10.
- [27] Noreen Fatima; Li Liu; Sha Hong; Haroon Ahmed "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis", *IEEE Access* 2020, Volume 8.
- [28] Ravi K. Samala; Heang-Ping Chan; Lubomir Hadjiiski; Mark A. Helvie; Caleb D. Richter; Kenny H. Cha "Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning Using Deep Neural Nets", *IEEE Transactions on Medical Imaging* 2019, Volume 38, Issue 3.

- [29] Zhiqiong Wang;Mo Li;Huaxia Wang;Hanyu Jiang;Yudong Yao;Hao Zhang;Junchang Xin “Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features”, IEEE Access 2019, Volume 7.
- [30] Michiel Kallenberg;Kersten Petersen “Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring”, IEEE Transactions on Medical Imaging 2016, Volume 35, Issue 5.
- [31] Bo Fu;Pei Liu;Jie Lin;Ling Deng;Kejia Hu;Hong Zheng “Predicting Invasive Disease-Free Survival for Early Stage Breast Cancer Patients Using Follow-Up Clinical Data”, IEEE Transactions on Biomedical Engineering 2019, Volume 66, Issue 7.
- [32] P. Esther Jebarani;N. Umadevi;Hien Dang;Marc Pomplun “A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection”, IEEE Access 2021, Volume 9.
- [33] Amin Ul Haq;Jian Ping Li “Detection of Breast Cancer Through Clinical Data Using Supervised and Unsupervised Feature Selection Techniques”, IEEE Access 2021, Volume 9.
- [34] Abdelali Elmoufidi “Deep Multiple Instance Learning for Automatic Breast Cancer Assessment Using Digital Mammography”, IEEE Transactions on Instrumentation and Measurement 2022, Volume 71.
- [35] K Nagaiah“Automated Assistance for Breast Cancer Identification on Mammograms using Computer Vision Algorithms” 2021 International Conference on Intelligent Technologies (CONIT) ISBN:978-1-7281-8583-5 2022.
- [36] Zexian Huang;Daqi Chen “A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm”, IEEE Access, 2022, Volume 10.
- [37] Sahar A. Mokhtar, “Predicting the Servery of Breast Masses with Data Mining Methods” International Journal of Computer Science Issues, Vol. 10, Issues 2, No 2, March 2013.
- [38] Milon Islam “Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors” 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.

