# A SURVEY ON DIFFERENT ALGORITHM FOR BREAST CANCER PREDICTION

S.Lincy Jemina[1], Dr. I. Muthulakshmi[2]
[1]PG Scholar, [2]Professor
[1,2] Department of CSE, V V College of Engineering, Tisaiyanvilai – 628 656, India.

## ABSTRACT

**A breast cancer is one of the leading kill disease in the human in today world. Mostly women can affect the breast cancer compared to men. To solve this problem, we are proposed a Different machine-learning techniques can be used for the prediction of breast cancer, and to achieve the accuracy for prevent the breast cancer in today life. In this paper, we have discussed the breast cancer and finally the study of various literatures available on prevention of breast cancer using the various machine learning algorithms.**

**Keywords- Breast cancer, accuracy, Machine learning algorithms**

## 1. INTRODUCTION

Breast cancer occurs when a malevolent tumor originates in the breast. As breast cancer tumors established, they may spread to other parts of the body. The principal route of metastasis is the lymphatic system which, ironically enough, is also the human body's for producing and transporting white blood cells and other cancer-fighting immune system cells throughout the body. Metastasized cancer cells that aren't cracked by the lymphatic system's white blood cells move through the lymphatic vessels and settle in isolated body locations, forming new tumors and perpetuate the disease process.

Breast cancer is practically common. Because of its well exposed nature, and potential for lethality, breast cancer is arguably the scary type of cancer diagnosis someone can receive.

However, it is important to keep in mind that, if identified and properly treated while still in its early stages, breast cancer can be cured.

Breast cancer is not now a woman's disease. It is fairly possible for men to get breast cancer, even though it occurs less habitually in men than in women. Our discussion will focus primarily on breast cancer as it relates to women but it should be noted that much of the information is also germane for men.

Symptoms of breast cancer is Lump in your breast or underarm, Swelling in the armpit, collarbone, or breast, Pain or tenderness in your breast, and flat or indented area on your breast etc..

## 2. TYPES OF BREAST CANCER

Breast cancer can be classified into : invasive and noninvasive.

- Invasive (infiltrating) breast cancer cells smash through normal breast tissue barriers and extend to other parts of the body through the bloodstream and lymph nodes.
- Noninvasive (in situ) breast cancer cells remain in a particular location of the breast, without dispersion to surrounding tissue, lobules or ducts.

**AGGRESSIVE BREAST CANCERS**

Ductal carcinoma in situ (DCIS) is a non-invasive cancer where abnormal cells have been found in the lining of the breast milk duct. The a typical cells have not broaden outside of the ducts into the surrounding breast tissue. Ductal carcinoma in situ is very early cancer that is highly treatable, but if it’s left untreated or unnoticed, it can spread into the surrounding breast tissue.

The invasive ductal carcinoma abnormal cancer cells that began forming in the milk ducts have spread beyond the ducts into other parts of the breast tissue. Invasive cancer cells can also spread to other parts of the body. It is also sometimes called infiltrative ductal carcinoma.

- IDC is the most common type of breast cancer, making up nearly 70- 80% of all breast cancer diagnoses.
- IDC is also the type of breast cancer that can most commonly affects men.

**DCIS VS IDC**

DCIS means the cancer is still contained in the milk duct and has not invaded any other area. IDC is cancer that began growing in the duct and is invading the surrounding tissue. Cancer staging done by a physician, along with a physical exam and medical history can help identify the best treatment options.

Triple Negative Breast Cancer is a diagnosis of triple negative breast cancer means that the three most common types of receptors known to fuel most breast cancer growth–estrogen, progesterone, and the HER-2/neu gene– are not present in the cancer tumor. This means that the breast cancer cells have tested negative for hormone epidermal growth factor receptor (HER-2), estrogen receptors (ER), and progesterone receptors (PR). Since the tumor cells lack the necessary receptors, common treatments like hormone therapy and drugs that target estrogen, progesterone, and HER-2 are ineffective. Using chemotherapy to treat triple negative breast cancer is still an effective option. In fact, triple negative breast cancer may respond even better to chemotherapy in the earlier stages than many other forms of cancer. [2] discussed that Biomedical and anatomical data are made simple to acquire because of progress accomplished in computerizing picture division. More research and work on it has improved more viability to the extent the subject is concerned. A few tech- niques are utilized for therapeutic picture division, for example, Clustering strategies, Thresholding technique, Classifier, Region Growing, Deformable Model, Markov Random Model and so forth. This work has for the most part centered consideration around Clustering techniques, particularly k-implies what's more, fluffy c-implies grouping calculations. These calculations were joined together to concoct another technique called fluffy k-c-implies bunching calculation, which has a superior outco- me as far as time usage. The calculations have been actualized and tried with Magnetic Resonance Image (MRI) pictures of Human cerebrum. The proposed strategy has expanded effectiveness and lessened emphasis when contrasted with different techniques. The nature of picture is assessed by figu- ring the proficiency as far as number of rounds and the time which the picture takes to make one emphasis. Results have been dissected and recorded. Some different strategies were surveyed and favorable circumstances and hindrances have been expressed

as special to each. Terms which need to do with picture division have been characterized nearby with other grouping strategies.

Inflammatory breast cancer is an an aggressive and fast growing breast cancer in which cancer cells infiltrate the skin and lymph vessels of the breast. It often produces no distinct tumor or lump that can be felt and isolated within the breast. But when the lymph vessels become blocked by the breast cancer cells, symptoms begin to appear.

Metastatic breast cancer is also classified as Stage 4 breast cancer. The cancer has spread to other parts of the body. This usually includes the lungs, liver, bones or brain.

## 3. LITERATURE REVIEW

### 3.1 PREDICTING BREAST CANCER RECURRENCE USING DATA MINING TECHNIQUES

Predicting breast cancer recurrence using data mining techniques is presented by Qi Fan et al: proposed to investigate five data mining algorithms: Artificial Neural Network with BP algorithm, and four different decision tree algorithms (C 5.0, CHAID, C&RT, and QUEST). In this paper, we applied these algorithms to predict whether or not breast cancer will recur for the breast cancer patient based on SEER data set. And we compared the results to find the most suitable one as the predictive model of breast cancer recurrence. The SEER Public-Use Data 2005 is used, and the c5 algorithm has the best performance of accuracy.

### 3.2 APPLYING BEST MACHINE LEARNING ALGORITHMS FOR BREAST CANCER PREDICTION AND CLASSIFICATION

Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification paper authors have proposed to analyze data from a breast cancer dataset using a classification technique in the field of medical bioinformatics to accurately predict the class in each case, using the weka data-mining tool and its use for classification. The main contributions of this work are: Select the best classifier for breast cancer prediction, Comparison of different data mining algorithms on the breast cancer dataset, Identification of the best performance-based algorithm for disease prediction.

The authors using several machine-learning algorithms that are Random Forest, Naïve Bayes, Support Vector Machines , and K-Nearest Neighbors are used effectively. The SVM provide better accuracy.

### 3.3 PREDICTION OF BREAST CANCER USING VOTING CLASSIFIER TECHNIQUE

Prediction of Breast Cancer using Voting Classifier Technique is presented by U. Karthik Kumar, M.B. Sai Nikhil and K. Sumangali to proposed to reduce the growth of the cells in breast tissue and control the cancer. The main objective of this paper is to compare the results of supervised learning classification algorithms and combination of these algorithms using voting classifier technique. Voting is the ensemble approach ,can combine multiple models for the better classification. The dataset is taken from Wisconsin University database.

### 3.4 USING RANDOM FOREST ALGORITHM FOR BREAST CANCER DIAGNOSIS

Using Random Forest Algorithm for Breast Cancer Diagnosis is presented by Bin Dai, Rung-Ching Chen et al: proposed to analyze the medical case diagnosis of breast cancer and combine the characteristics of multiple eigen values. Based on the ensemble learning method of random forests, the results of multiple weak classifiers are combined and to produce better accuracy. We used auxiliary medical diagnosis data sets.

**TABLE 1 COMPARISON OF DIFFERENT ALGORITHMS**

| Algorithm | Description | Performance | Data set Name |
|---|---|---|---|
| C 5.0, CHAID, C&RT and QUEST are used. | To predict whether or not breast cancer will recur for the breast cancer patient based on SEER data set. | The c5 algorithm has the best performance of accuracy. | SEER Public-Use Data 2005 |

| | | | |
|---|---|---|---|
| Random Forest, Naïve Bayes, Support Vector Machines and K-Nearest Neighbors | To analyze data from a breast cancer dataset using a classification technique in the field of medical bioinformatics to accurately predict the breast cancer, using the weka data-mining tool and its use for classification. | SVM gives the highest accuracy 97.9%. | Wisconsin Hospitals Madison Breast Cancer Database |
| Voting classifier algorithm | The objective of the paper is reduce the growth of the cells in breast tissue and control the cancer. | Provide better performance | Wisconsin University database. |
| A random forest algorithm | To analyze the medical case diagnosis of breast cancer and combine the characteristics of multiple eigen values. | High prediction accuracy | Auxiliary medical diagnosis. |
| Forward back propagation Algorithm | we develop a system that can classify “Breast Cancer Disease” tumor and to classify the tumor from a symptom that causes the breast cancer disease. | Backpropagation algorithm provide 96.63 accuracy. | University of Wisconsin (UCI) Machine Learning Repository |
| Non-symmetrical C-SVM algorithm | To develop the breast cancer diagnosis SVM based application. | The application provide better performance. | University of Wisconsin Hospitals dataset used. |

### 3.5 BREAST CANCER PREDICTION BASED ON BACKPROPAGATION ALGORITHM

Breast Cancer Prediction Based On Back propagation algorithm is presented by Muhammad Sufyian Bin Mohd Azmi, and Zaihisma Che Cob to develop a system that can classify “Breast Cancer Disease” tumor using neural network with Feed forward Back propagation Algorithm to classify the tumor from a symptom that causes the breast cancer disease. Breast cancer tumor database used University of Wisconsin (UCI) Machine Learning Repository. [5] discussed about the

combination of Graph cut liver segmentation and Fuzzy with MPSO tumor segmentation algorithms. The system determines the elapsed time for the segmentation process. The accuracy of the proposed system is higher than the existing system. The algorithm has been successfully tested in multiple images where it has performed very well, resulting in good segmentation. It has taken high computation time for the graph cut processing algorithm. In future work, we can reduce the computation time and improves segmentation accuracy.

### 3.6 BREAST CANCER DIAGNOSIS VIA SUPPORT VECTOR MACHINES

Breast Cancer Diagnosis via Support Vector Machines is presented by Wang Yi et al: proposed to develop the application of SVM to breast cancer prediction and give the better accuracy. The authors used the algorithm as C-SVM, it used a soft margin optimization, which can be handle with the general case of non-linearly separable classes with linear and non-linear surface. The University of Wisconsin Hospitals dataset are used to predict the breast cancer diagnosis.

## 4. CONCLUSION

In this study, Breast cancer diagnosis has been a typical machine learning benchmark problem for many years. It has been dealt with using various machine learning algorithms. The problem an diagnosis system for predicting the breast cancer based on different algorithm is presented. And explain the different ML approaches and their applications in breast cancer diagnosis and prognosis used to analyze the dataset in different biomedical database.

## REFERENCES

[1] Uma Ojha , Dr. Savita Goel, [IEEE 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence (Confluence) - Noida, India (2017.1.12-2017.1.13)] 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence - A study on prediction of breast cancer recurrence using data mining techniquesDOI:10.1109/CONFLUENCE.2017.7943207

[2] Christo Ananth, S.Aaron James, Anand Nayyar, S.Benjamin Arul, M.Jenish Dev, “Enhancing Segmentation Approaches from GC-OAAM and MTANN to FUZZY K-C-MEANS”, Investigacion Clinica, Volume 59, No. 1, 2018,(129-138).

[3] Youness Khourdifi , Mohamed Bahaj, [IEEE 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) - Kenitra, Morocco (2018.12.5-2018.12.6)] 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) - Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification DOI: 10.1109/ICECOCS.2018.8610632

[4] U. Karthik Kumar, M.B. Sai Nikhil and K. Sumangali, [IEEE 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM) - Chennai, India (2017.8.2-2017.8.4)] – "Prediction of breast cancer using voting classifier technique", DOI: 10.1109/ICSTM.2017.8089135

[5] Christo Ananth, D.R.Denslin Brabin, "ENHANCING SEGMENTATION APPROACHES FROM FUZZY K-C-MEANS TO FUZZY-MPSO BASED LIVER TUMOR SEGMENTATION", Agrociencia, Volume 54, No. 2, 2020,(72-84).

[6] Parag Singhal, Saurav Pareek, [IEEE 2018 2nd International Conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) - Palladam, India (2018.8.30-2018.8.31)] 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on – "Artificial Neural Network for Prediction of Breast Cancer", DOI: 10.1109/I-SMAC.2018.8653700

[7] S. Muthuselvan, Dr. K. Soma Sundaram, Dr. Prabasheela, [IEEE 2016 International Conference on Information Communication and Embedded Systems (ICICES) –Chennai, Tamilnadu, India (2016.2.25-2016.2.26)] 2016 International Conference on Information Communication and Embedded Systems (ICICES) – "Prediction of breast cancer using classification rule mining techniques in blood test datasets", DOI:10.1109/ICICES.2016.7518932.

AUTHOR(S) BIOGRAPHY

**S. Lincy Jemina** received the B.E degree in Computer Science and Engineering from Kings Engineering College, Chennai in 2018. She is pursuing M,E degree in Computer Science and Engineering from VV college of Engineering, Tisaiyanvilai. Her area of interest include Data mining ,Cloud computing, and Bigdata analysis.

**I.Muthulakshmi** received the B.E degree in Electronics and Communication Engineering from Government College of Engineering, Tirunelveli in 1989, the M.E degree in Computer Science and Engineering from Manonmanium Sundaranar University in 2004 and the Ph.D degree in Computer Science and Engineering from Manonmanium Sundaranar University in 2013. She is currently working as a Professor of Computer Science and Engineering in V V College of Engineering, Tisaiyanvilai. Her research interests include Digital Image Processing, Artificial Neural Networks, Networking and Data Mining.