

# Early diagnosis and Prediction of Breast Cancer

Gagana B R, Shraddha C, Chethana H T



**Abstract:** Nowadays Cancer is one of the frequent diseases for all the humans in the world. Particularly, in women's Breast Cancer is one of the most frequent diseases. Therefore, early detection and prevention of cancer is very important to get healthy life so we need to develop new techniques for diagnosis and prediction of cancer. Many ML techniques are used for early diagnosis and prediction of cancer. In this paper, we are proposing new techniques for the classification of Breast cancer and the time prediction when that breast cancer has occurred using some ML techniques and our proposed work is medical sector application. We compared our proposed methodology result with each other techniques to get the highest accurate value.

**Keywords:** Analysis, Breast Cancer, Classification, Early detection and prevention, Machine learning techniques, women.

## I. INTRODUCTION

Breast Cancer is one of the frequent disease in women and most of the death has caused in women Manual analysis of this disease will take more hours and systems availability becomes lesser, so there is a requirement to develop a new technique for automatic identification system for early detection of cancer. A lot of ML techniques are used for the development of such system. ML classification techniques are used for classifying the breast cancer tumor using past patients records. Breast cancer is the second most dangerous type of cancer in women. According to World Cancer Research Fund in 2018, Breast cancer estimation was 25.4% and were recorded as Positive [1]. ML is nothing but it is a performance task to the machine without writing any code or instructions. So here, we are predicting the classification of Breast cancer to check whether the Person has reached initial stage or final stage of cancer using the training dataset as shown in Figure1 [2].

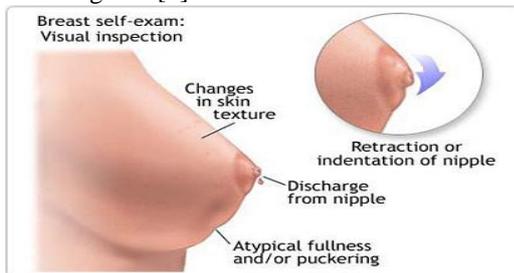


Figure 1: Breast Cancer symptoms [14]

Revised Manuscript Received on July 15, 2020.

\* Correspondence Author

**Gagana B R\***, Student, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore, Karnataka, India.

**Shraddha C**, Assistant Professor, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore, Karnataka, India.

**Chethana H T**, Assistant Professor, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore, Karnataka, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Nowadays prediction of breast cancer and classification of BC is done by using different data science algorithms like SVM, Naive bayes etc. By using these kinds of algorithms the accuracy level of detecting the cancer is very less. Our idea is to predict the classification of the BC using KNN algorithm. KNN algorithm comes under the supervised learning technique. The main of this KNN algorithm is to predict the stage of the cancer whether it's benign or malign [3]. In this paper, we are focusing on predicting the type of breast cancer by using some classification techniques in data science. Classification is the most commonly used technique in data science and it uses training datasets to classify the patient records. The main aim of the classification technique is to provide more accurate result than others. The main idea of this paper is to examine the BC dataset using a classification technique in data science. First it will classify data set and then it uses algorithms for finding and prediction of BC disease and also predicts the time when the breast cancer has occurred using same classification techniques. First it will predict the type of cancer then the time at which the breast cancer has occurred [4].

## II. LITERATURE SURVEY

In this section existing literature survey is carried out on Breast Cancer analysis.

**B. Krishnakumar et.al (2020) [5]** proposed a new features to extract the images using Convolution layer and, to check whether the tumor is in beginning stage or final stage of BC by passing the extracted image to fully linked layer. The proposed frameworks are done by using standard benchmark datasets and CNN, VGGNet Models were used to calculate the performance. Finally comparison between these two models were made. From the comparative study, CNN provides more accuracy than VGGNet.

**Sara AlGhunaim et.al (2019) [6]** Proposed a framework for breast cancer prediction using three different classification technique of GE. Our new approach is to analyze the dataset of BC using 2 kinds of big data. i.e., DM and Combined dataset of GE. Then compare the performance and efficiency of different predictive models in order to find the best classification accuracy. Final results shows that GE data is better for breast cancer classification

**ParagSinghal et.al (2018) [7]** Proposed a framework using back-propagation algorithm for Breast cancer prediction. In this paper, Breast cancer analysis is perfectly done by using back propagation technique. The outputs are analyzed by taking existing and new dataset of patients in secret layer. From the experimental result, we obtained an accuracy of 7 secret layers and it was compared with newly created models. For BC prediction feed-forward back-propagation algorithm gives best result with good accuracy.



# Early diagnosis and Prediction of Breast Cancer

*Dr. S. N. Singh et.al (2018) [8]* The purpose of this research is to compare different classifiers like CART, J48 Bayesian Logistic to find which is the best classifier and which is having more accurate and time complexity in breast cancer prediction. All this comparison experiment is done in Weka tool. The result obtained by all these classifier are then compared to find out which one is best classifier for BC prediction. From this comparison we can say that CART having best accuracy level compare to all other classifiers but it takes more time.

*YounessKhourdifi et.al (2018) [4]* Provided explanations of different ML approaches by using different algorithms like SVM, Random-forest, Naive Bayes to find the performance and accuracy for breast cancer diagnosis and prognosis. SVM has proven that it give more efficient and more accurate result compare to other existing techniques.

### III. PROPOSED SYSTEM

Our proposed framework focuses on classification of Breast cancer using Machine Learning technique. Figure 2 shows the architecture of proposed framework.

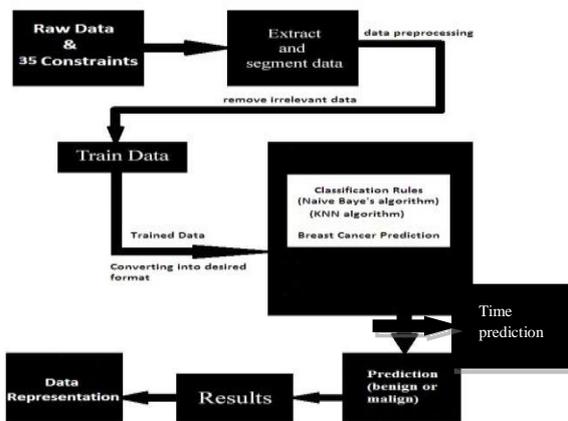


Figure2: Architecture of proposed system

First it will collect the BC Data from UCI dataset then remove the irrelevant data by extracting and segmenting the collect data. This extracted data is considered as training dataset. Using this training dataset we are going to classify the breast cancer by applying ML classification rules. Before performing this classification rule, we need to convert the training data into some desired format. Same classification rule is applied for time prediction of BC then that results are presented like data representation of BC

#### A. Dataset

Dataset which we are used in this proposed system are collected from UCI Machine Learning Repository. Collected from 1000 datasets for BC analysis which are classified as benign and malignant. Also, dataset has 9 attributes values [9].

The 9 attributes values of dataset are;

- Clump Thickness
- Uniformity of Cell Size

- Uniformity of Cell Shape
- Marginal Adhesion
- Single Epithelial Cell Size
- Bare Nuclei
- Bland Chromatin
- Normal Nuclei
- Mitoses
- Result (2: Benign, 4: Malignant)

#### B. Methodology

Methodology is a collection of techniques, which provides step by step information about our proposed framework.

- *K – Nearest Neighbor (KNN)*

KNN algorithm uses training dataset to makes predictions directly. K most similar instances of training dataset are used for the Predictions of new instance and summarizing the output variable for those K instances. In classification technique this value might be class value [10]. To determine the distance between the new input and k instances of training dataset Euclidean distance is used. Euclidean distance is the most popular formula to measure the distance. Euclidean distance is calculated for the entire input attribute by square root of sum of squared difference among fresh instances and already obtained instance. Euclidean Distance  $(A, A_i) = \sqrt{\sum ((A_j - A_{ij})^2)}$ . By collecting K instance values after performing Euclidean Distance we need to sort this value in order to find the highest value and then we will get result for prediction [11].

- *Naives Bayes*

It is not just a single algorithm but it is a collection of classification algorithms based on Bayes' Theorem but it share a common principle, i.e. it's independent of one another for each pairs. Bayes' Theorem is stated as:  $P(h|d) = (P(d|h) * P(h)) / P(d)$  Naive Bayes is a classification algorithm for two or more class of classification problems [12]. When this classification technique is described in binary input value it is very easy to understand. Assume that we have 2 classes of data within this dataset. We have probability equation for class 1:  $q_1(h, d)$  and class 2:  $q_2(h, d)$ . In order to find measurement for new instance, we use some following rules i.e., if  $q_1(h, d) > q_2(h, d)$  then the result is class 2 [11].

### IV. EXPERIMENTAL RESULT

This section presents the experimental result of the proposed method based on dataset collected by the user.

Our Proposed system focuses on classification of the breast cancer and time prediction of BC using ML techniques. We have used KNN and Naive Bayes for this classification. Table 2 shows the comparative analysis of both the algorithms along with figure 1 shows graphical representation of both algorithm. Table 3 shows the time prediction analysis of BC. Finally result shows that KNN algorithm will provide more accuracy and it takes less time than naive bayes's algorithm.

▪ Test cases

Software Testing is one of the major roles to find bugs and defects in software before and after developing application [13].Table 3 shows test cases of Breast cancer Prediction after developing software.

**Table 1: Test cases of Breast cancer prediction**

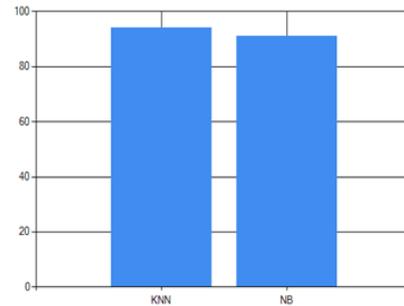
TC#	Description	Expected Result	Actual Result	Status of Execution Pass/Fail
TC01	Execute run the application	Application should run without any interrupts.	Application is executing properly	Pass
TC02	Verification of Admin Login Input User Name and Password then click on Login button.	Admin User Name & Password should be check: verify with database.	Admin User Name & Password successfully checked with database.	Pass
TC03	Verification of Input User Name & Password of Admin.	When username & Password of admin is correct then it should navigate to respective Admin home page.	When username & Password is correct or valid then successfully navigating respective home page.	Pass
TC04	Verification of Input User Name & Password of Admin. (Invalid Case)	When admin username and password is incorrect then this display message that user Name & Password is invalid	When username & Password is incorrect or invalid then message box display that whether use name or Password wrong	Pass

**Table2: Comparative Analysis of Algorithms**

Constraint	KNN	Naive Bayes
Accuracy	94%	91%
Time (Milisecs)	1010	1163
Correctly Classified	94%	91%
Incorrectly Classified	6%	9%

**Table 3: Time Prediction Analysis**

Constraint	KNN Algorithm
Accuracy	90%
Time (Milisecs)	160
Correctly Classified	90%
Incorrectly Classified	10%



**Figure 3: Graphical Representation of Algorithms**

**IV. CONCLUSION**

Breast cancer is one of the frequent issues in female and most of the humans lose their life .In past 20 years many research are going on Breast Cancer prediction and many of the technologies are also developed but efficiency and accuracy is less. In this paper, we have used different classification Techniques on ML for Time prediction of BC and classification of cancer using UCI Dataset. The effectiveness of using classification technique is shown by compare performance obtained in terms of accuracy and time. Based on comparisons of the performance tell that KNN algorithms gives More accurate result than other technique.

**REFERENCES**

- Ch. Shravya, K. Pravalika and Shaik Subhani, "Prediction of Breast Cancer Using Supervised Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-6, April 2019.
- Mamatha Sai Yarabarla, Lakshmi Kavya Ravi and Dr. A. Sivasangari, "Breast Cancer Prediction via Machine Learning", International Conference on Trends in Electronics and Informatics, ISBN: 978-1-5386-9439-8, 2019.
- Mr.P.Sathiyarayanan, Ms. Pavithra.S, Ms. Sai Saranya.M and Ms. Makeswari.M "Identification of Breast Cancer Using the Decision Tree Algorithm", International conference, 2019.
- Youness Khouridfi and Mohamed Bahaj, "Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification", DOI: 978-1-5386-4225-2/18, 2018.
- B.Krishnakumar, K.Kousalya, R.S. Mohana, K.Dinesh and S.Santhiya, "Classification of Breast Cancer using Deep Learning Architecture", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
- Sara AlGhunaim and Heyam Al-Baity, "On the Scalability of Machine-Learning Algorithms for Breast Cancer Prediction in Big Data Context", journal, DOI: 10.1109, 2019.
- Parag Singhal and Saurav Pareek, "Artificial Neural Network for Prediction of Breast Cancer", International conference, ISBN: 978-1-5386-1442-6, 2018.
- Dr. S. N. Singh and Shivani Thakral, "Using Data Mining Tools for Breast Cancer Prediction and Analysis", International Conference on Computing Communication and Automation (ICCCA), DOI: 978-1-5386-6947-1, 2018.
- Ebru Aydımdag Bayrak, Pinar Kırıcı and Tolga Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis", International Conference, DOI: 978-1-7281-1013-4, 2019.
- Bhondve Arti T, Bhame Vaishali S, Kadam Aishwarya R and Kopnar Komal D," Breast Cancer Disease Prediction: Using Machine Learning Approach", International Research Journal of Engineering and Technology (IRJET) E-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019.



## Early diagnosis and Prediction of Breast Cancer

11. Anusha Bharat, Pooja N and R Anishka Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis", International Conference on Circuits, Control, Communication and Computing, 2018.
12. Megha Rathi and Arun Kumar Singh, "Breast Cancer Prediction using Naïve Bayes Classifier", International Journal of Information Technology & Systems, Vol. 1; No. 2: ISSN: 2277-9825, July-Dec-2012.
13. Ingrid A. Buckley and Winston S. Buckley," Teaching Software Testing using Data Structures", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, 2017.
14. <https://steemit.com/air-clinic/@eunireal1/breast-self-examination>.

### AUTHORS PROFILE



**Gagana B R**, M-tech Student, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore - 570006, Karnataka, India.



**Shraddha C**, Asst. Professor, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore - 570006, Karnataka, India.



**Chethana H T**, Asst. Professor, Department of Computer Science & Engineering, Vidyavardhaka College of Engineering, Mysore - 570006, Karnataka, India.