World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Earlier diagnosis of breast cancer using e- health file with hybrid machine learning algorithms

P Priya * and T Meyyappan

Department of Computer Science, Alagappa University, Tamil Nadu, India.

World Journal of Advanced Engineering Technology and Sciences, 2022, 07(02), 212–219

Publication history: Received on 08 November 2022; revised on 21 December 2022; accepted on 24 December 2022

Article DOI: https://doi.org/10.30574/wjaets.2022.7.2.0163

Abstract

Breast Cancer denotes the major disease which makes a high range of deaths worldwide. Earlier forecasting of breast cancer can help the breast cancer patients by improving their survival. Data mining methodologies and machine learning models are widely used in the health sector fields to offer effective diagnosis in various diseases. The main theme of this research work is to implement a hybrid approach of breast cancer predictions by using the SVM (Support vector Machine) and Naïve Bayes classifiers to predict the breast cancer using the E-Health file data of a patient. The proposed approach is aimed to offer the maximum accuracy than the existing approaches.

Keywords: Diagnosis; Breast Cancer; E-Health; Machine Learning; SVM; Naïve Bayes; Hybrid

1. Introduction

Nowadays Breast cancer is treated as one of the severe disease in India due to the increase of death rates. The count of breast cancer patients are increasing rapidly in every day. Earlier prediction of any disease can reduce the treatment complexity. Also the earlier detection of breast cancer helps to make the precautions to control the spread of the disease and death. It is mandatory is provide the guidance about the disease, effect of the disease and the earlier forecasting methodologies and its importance's into the peoples to reduce the count of patients and death ratio.

The earlier detection is the best way to know the possibilities and impact of certain disease before the deep infect. Hence, the earlier detection of breast cancer also helps the patients to avoid or cure the disease at earlier stage.

Machine Learning is the feature which performs automated analysis and reporting features without manual interactions. Machine learning techniques are applied in various fields and it is s mandatory feature in medical arena. The machine learning is vital to detect and diagnosis the diseases in accurate and effective manner. The machine learning approaches can be implemented in earlier breast cancer detection and diagnosis to get accurate results.

The researches which are made in health sectors reports that the big data and machine learning approaches should be integrated to help the people by predicting diseases, health related risk factors, disease management and health planning.

The term E-HR represents the Electronic Health Record is a computerized edition of every patient's health information's. The E-Health File systems are implemented to maintain the concurrent, centralized health records management and accessibility through a secured way. It can contain the entire profile about the patients, diagnosis profiles, medical, ongoing treatments and everything about a patient. The Health record analysis with the machine learning approaches can help to predict the major diseases like breast cancer and others.

* Corresponding author: P Priya

The Support Vector Machine is categorized under the supervised machine learning algorithm. It can try to predict a result by using the various features. The target is called dependent variable and the given features are called as independent variables. The SVM performs the predictions by using the mapping processes. The SVM is chosen because the SVM can work in structured, semi structured and unstructured data. The SVM can work efficiently in text and images and provide better results than other machine leaning algorithms.

The Naïve Bayes is also categorized under the supervised machine learning algorithm. It can perform with the probability approaches and predict the future by using the previous result values. The Naïve bayes algorithm provides faster and accurate prediction results. The Naïve bayes algorithm can be implemented for any type of data in easier way.

The SVM and Naïve bayes machine learning algorithms are implemented in this framework to provide faster and accurate prediction report of breast cancer with the health data and mammography images from the E-HR source.

2. Related Work

There are huge number of researches are performed in the area of breast cancer early diagnosis and prediction. The Prediction of breast cancer using supervised machine learning techniques [1] compared the following machine learning algorithms Logistic regression, SVM and K Nearest Neighbor and concluded that SVM is the effective algorithm for breast cancer prediction than the others.

The authors of the research "Breast cancer Biopsy predictions based on mammographic diagnosis" used there are five machine learning algorithms called Logistic regression, Linear Discriminate Analysis, Quadratic Discriminate Analysis, Random forest and Support vector machine. The authors proved that the SVM model can provide high accuracy of 99.9% than the others [2].

The authors of the research used Gaussian naïve bayes classifier to the breast cancer classification. The authors concluded that the naïve bayes can perform faster and accurate predictions with less attributes. The authors tested with health values and the naïve bayes provide 96.6% of accuracy [3].

3. Proposed Approach

The earlier diagnosis of breast cancer is necessary to reduce the breast cancer death ratio. The HMLA model is chosen because of the result of SVM and Naïve bayes in the previous researches. The Hybrid machine learning algorithms model implements the Support vector machine and Naïve Bayes to provide accurate results in the prediction of breast cancer using the E-HR health records.

The proposed HMLA approach is implemented with the following stages:

- Train Dataset
- Data Selection
- Data Preprocessing
- Data Transformation
- Pattern Matching
- Prediction

3.1. Train Dataset

The very first stage of this proposed model is the dataset training stage. The E-health records are collected from the authorized sources at this stage. The breast cancer patients' health information's and the mammograms are collected and organized into the following categories Normal, benign and malignant.

The Health data contains the following data like age, breast depth, breast size and position. All the above data are organized under a patient ID and stored in a data table. Totally 150 instances are collected to train the dataset which has high quality mammograms in Normal benign and malignant.

3.2. Data Selection

The HMLA Framework can permit simple mammogram image up loader to predict the breast cancer possibilities. Once the mammogram image is uploaded the framework detects the exact location of mammogram sample and calculates

the pixels count and density, height and weight of the mammogram image. The selection unit also detects the pixels(X, Y) range of the breast position.

3.3. Data Preprocessing

The data preprocessing stage inputs the mammogram and the patient age to be predict. The given age is converted into a age range with higher and lower bounds. This stage verifies the training dataset and detects the Health records which have the age under the higher and lower bounds. The records are maintained in a temporary dataset with its mammogram images.

3.4. Data Transformation

The Data transformation stage loads the processed records and the mammogram image to be predicted. The unit verifies the health records with the range of age. The data transformation unit loads the mammogram of every patient from the temporary dataset who have the same age range. The unit involves loading the mammogram image with auto cropping feature. This feature helps to set same height and width to perform exact breast cancer detection.

3.5. Pattern Matching

The Pattern matching stage does a deep matching process between the new mammogram image and the benign, malignant and normal breast mammogram images in the train dataset. The pattern matching process involves in pixel by pixel matching approach for breast cancer possibility detection. Once the pattern matching is completed between a pair of mammogram images it load the next set of images and does the comparison again and again. The comparison is done with all the mammogram images in the benign, malignant and normal breast which are in the given age range.

3.6. Prediction

The SVM and Naïve Bayes machine learning algorithms starts is learning process at every comparison done in pattern matching stage. The total available pixels, matching pixels and mismatch pixels are calculated by the algorithms and the overall normal breast possibilities, benign possibilities and malignant possibilities are calculated with the both SVM and Naive Bayes algorithms. The matching results are shown and the overall result is predicted and shown to the user.

4. Proposed Approach HMLA Architecture

The HMLA Approach is implemented with the following architecture:

The Proposed Architecture shows the sequential flow of the entire stages involved in the HMLA Approach to perform effective predictions.

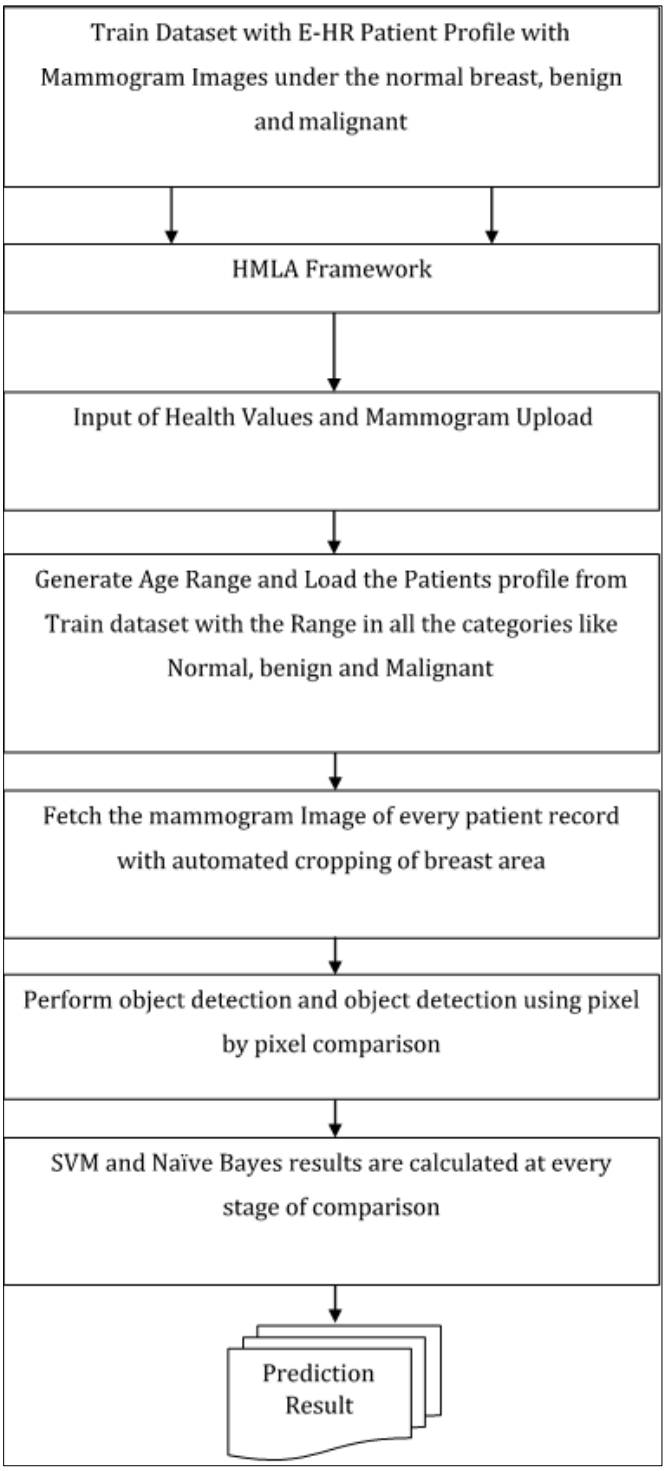

Figure 1 Proposed Architecture

5. Experimental Results

The HMLA Approach is implemented as a GUI based application using C Sharp windows application. The stage by stage results are illustrated below:

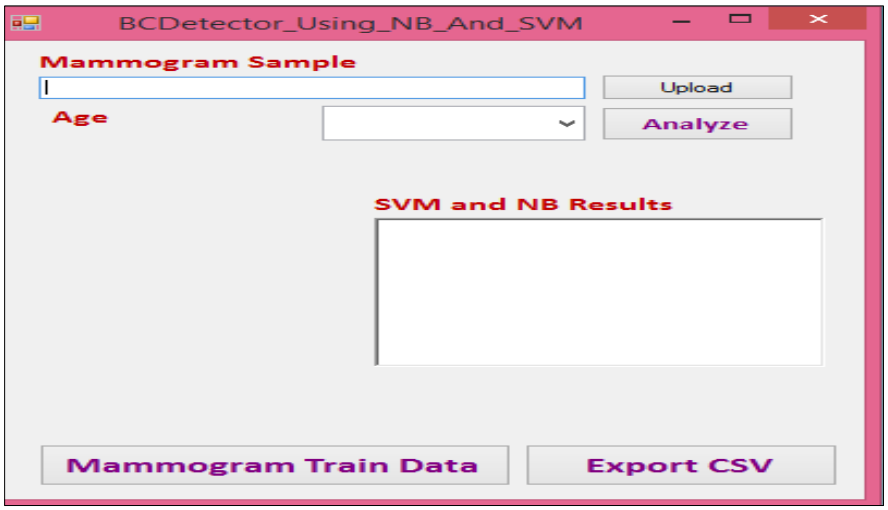

Figure 2 The Starter

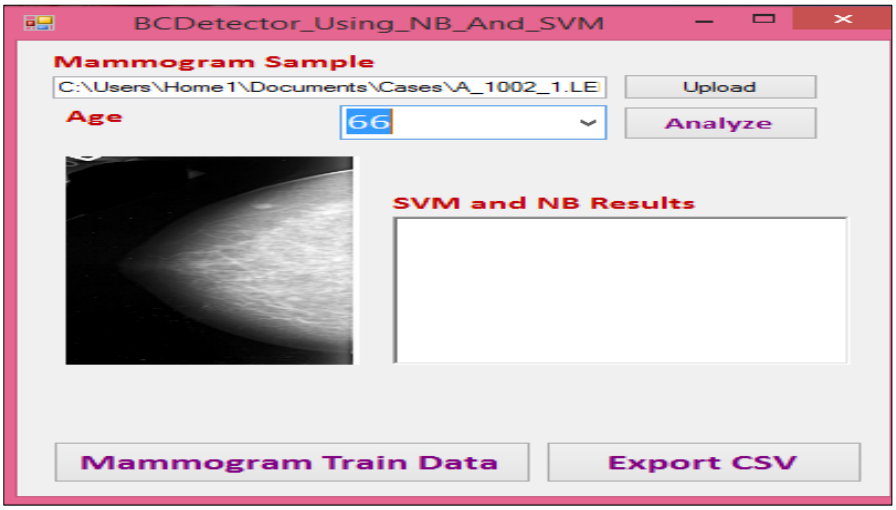

Figure 3 Mammogram Scan Uploads With Patient Age from E-Health Record

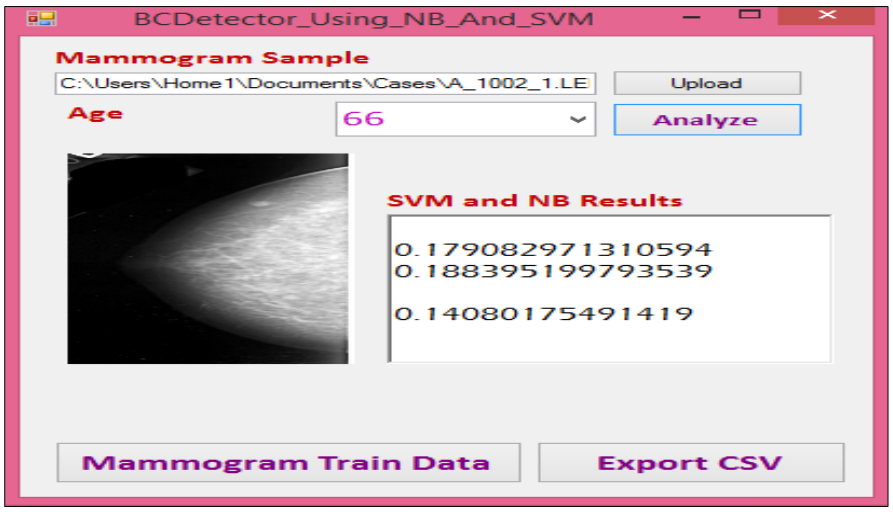

Figure 4 Result analysis with SVM and NB Machine Learning Algorithm

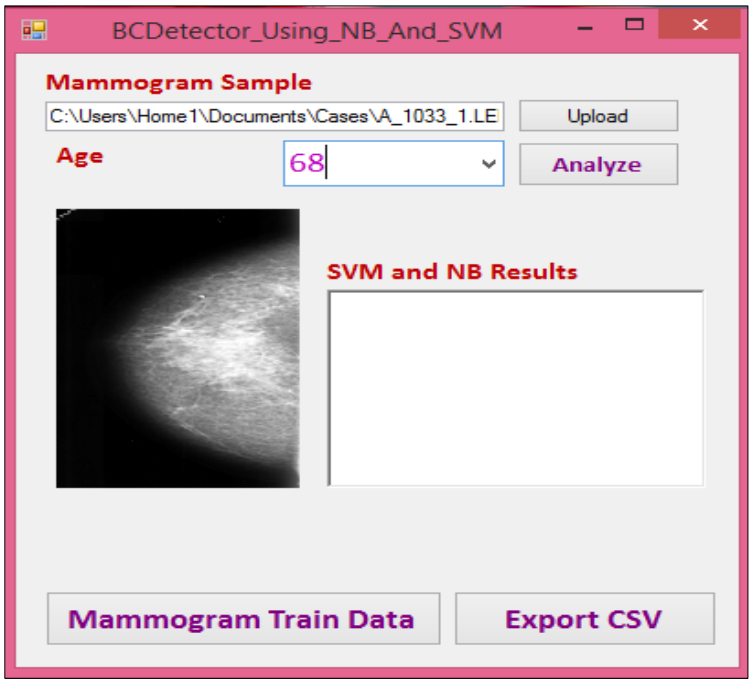

Figure 5 Mammogram Scan Uploads With Patient Age from E-Health Record

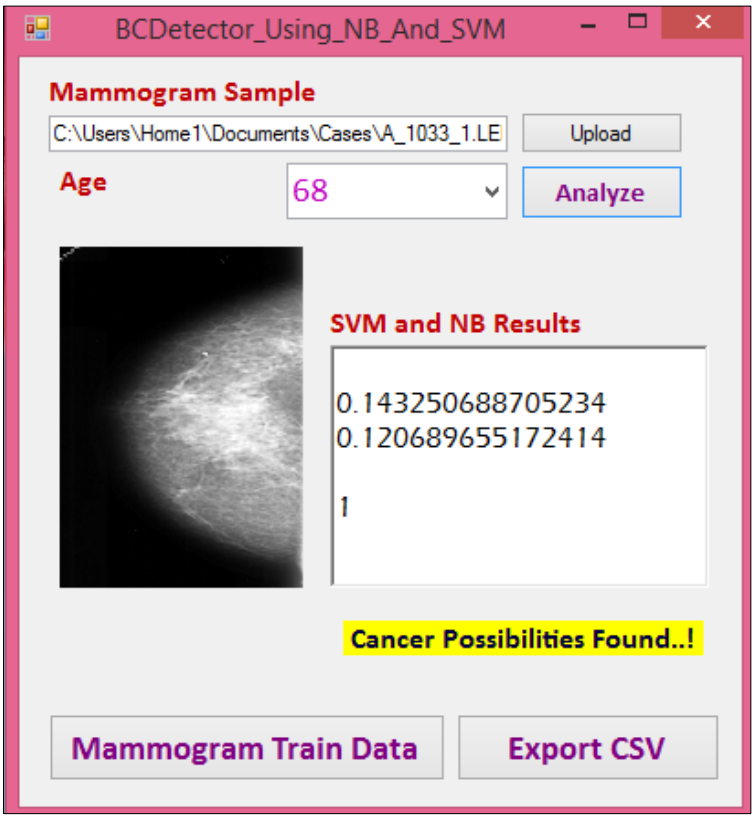

Figure 6 Result analysis with SVM and NB Machine Learning Algorithm

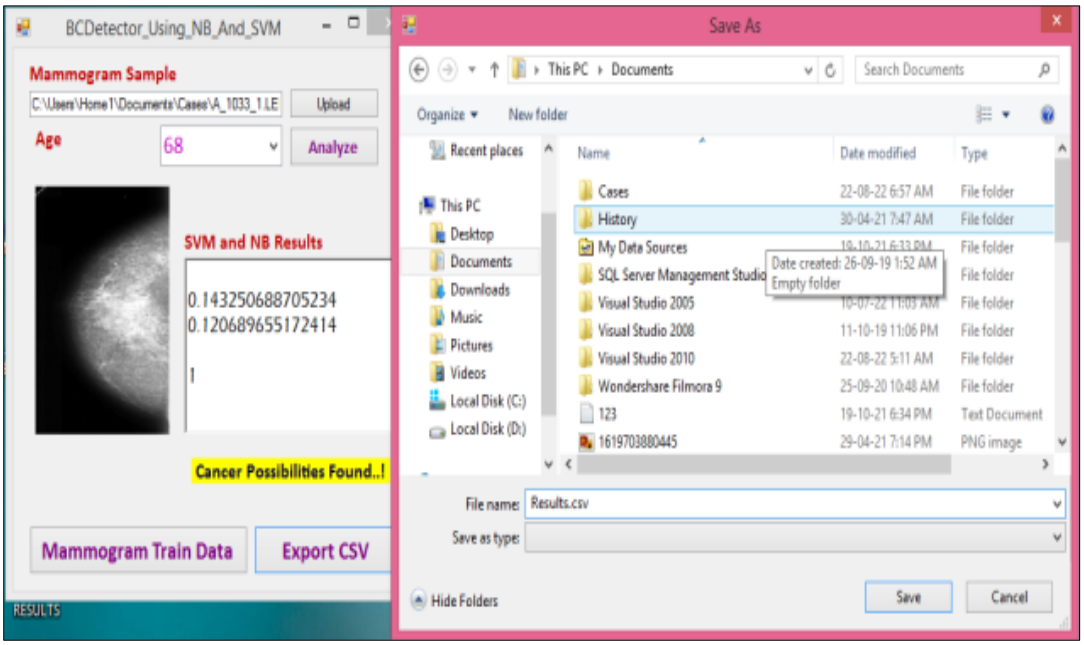

Figure 7 CSV Export utility

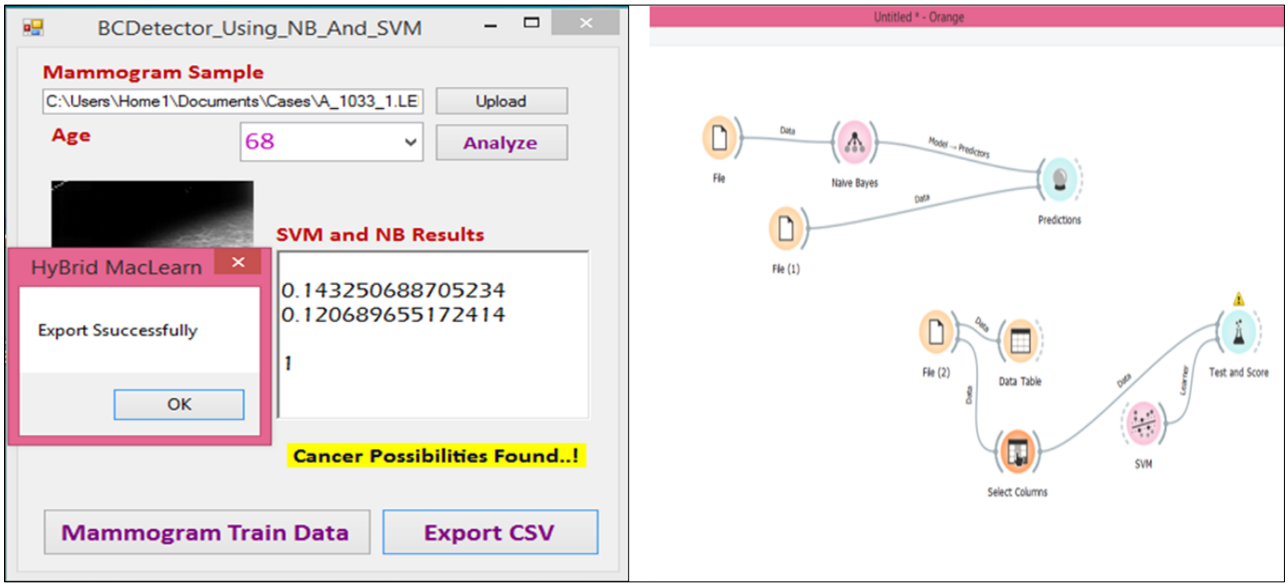

Figure 8 Prediction results using Utility Tool

6. Conclusion

The HMLA Approach proposed a hybrid model of breast cancer prediction using the Support vector machine and naïve bayes machine learning algorithms with in-depth pixel matching process in mammogram images. The implementation process includes a large dataset of Electronic Health record of patients with their mammogram images. The health records are arranged with the categories of normal breast, benign and malignant. The proposed research work permits the new predictions by uploading the mammogram image.

The proposed approach performs the prediction by using the following stages: Train Dataset, Data Selection, Data Preprocessing, Data Transformation, Pattern Matching and Prediction. The model implements the SVM and naïve Bayes algorithms to perform and calculate the accuracy and prediction results. Both the SVM and naïve Bayes results are verified separately at every matching process. The SVM and Naïve bayes are utilized because of its accuracy in earlier researches. The results are also verified with the utility tool. By combining the hybrid algorithm we can get the high accuracy results in new breast cancer predictions. The proposed model is tested with totally 150 instances and that are

equally divides into the normal breast, benign and malignant. The proposed approach performs the matching process in 0.36 Seconds and provides the 97.876 % accuracy. The result also verified with the utility tool.

Compliance with ethical standards

Acknowledgments

The authors would like to acknowledge Department of Science and Technology, New Delhi for the financial support in general and infrastructure facilities sponsored under PURSE 2nd Phase programme (Order No. SR/ PURSE Phase 2/38 (G) dated: 21.02.2017).

Disclosure of conflict of interest

No conflict of interest.

References

[1] Ch. Shravya, K. Pravalika, Shaik Subhani, 'Prediction of Breast Cancer using supervised Machine Learning techniques', 2019

[2] David Oyewola, Danladi Hakimi, Kayode Adeboye, Musa Danjuma Shehu, 'Using Five Machine learning for breast cancer Biopsy Predictions based on mammographic Diagnosis', 2016

[3] Wan Nor Liyana Wan Hassan Ibeni, Mohd Zaki Mohd Salikon, Aida Mustapha, Saiful Adli Daud, Mohd Najib Mohd Salleh, "Comparative Analysis on Bayesian Classification for breast cancer problem" 2019.

[4] Apoorva, Yogish H K, Chayadevi M L, "Breast Cancer Prediction Using Machine Learning Techniques", 2021

[5] Roma Mitra DEBNATH, Apaar DHINGRA, "Health sector and application of big data: A case study of India", 2022

[6] Md. Ehtisham Farooqui, Dr. Jameel Ahmad, "Disease Prediction System using Support Vector Machine and Multilinear Regression", 2020

[7] Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan, "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors", 2017

[8] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika , EL Habib Benlahmar , Rachida Ait Abdelouhahid , Olivier Debauche, "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis", 2021

[9] Chandrasekhar Rao Jetti , Rehamatulla Shaik , Sadhik Shaik , Sowmya Sanagapalli, "Disease Prediction using Naïve Bayes - Machine Learning Algorithm", 2021

[10] 'WHO | Breast cancer', WHO. http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/ (accessed Feb. 18, 2020)