Breast Cancer Detection using Neural Network Mammogram

Renuka Nagpure
Department of Information Technology
Atharva College of Engineering
Mumbai, India
renukanagpure@atharvacoe.ac.in

Sumita Chandak
Department of Information Technology
Atharva College of Engineering
Mumbai, India
sumitachandak@atharvacoe.ac.in

Nileema Pathak
Department of Information Technology
Atharva College of Engineering
Mumbai, India
nileemapathak@atharvacoe.ac.in

2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW) | 978-1-7281-4635-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCDW45521.2020.9318635

***Abstract* - Now-a-days Breast cancer is one of the serious issues for women when compared to all other cancers. In India Breast cancer risk revealed that 1 in 28 women cause to grow and become more mature breast cancer during her lifetime. Breast cancer begins when cells in the breast use to grow or start to grow out of control and it advances from breast tissue. These cells usually form a tumor that can often be seen on a mammogram or felt as a lump. If cells can grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body then the tumor is malignant(cancer). Breast cancer occurs in both women and men. In India the average age is the very great extent of cancer in the age group of 43-46 years unlike in the Western where women aged 53-57 years are more likely to suffer from breast cancer. This paper includes uses of various Data Mining along with neural networks to identify the presence of breast**

- breast cancer, neural network, mammogram, adaptive mean filter, GMM segmentation, PNN classification

I. INTRODUCTION

Breast cancer is one of the leading cancers for women when compared to all other cancers[1][2].Breast cancer is developed from breast tissue. Knowing the early signs of breast cancer can help a person go with exact treatment and in further improve their chances of survival. Risk factors for developing breast cancer include being female, obesity, lack of physical exercise, drinking alcohol, hormone replacement therapy during menopause, ionizing radiation, early age at first menstruation, having children late or not at all, older age, prior history of breast cancer, and family history. This system uses scanned mammograms as inputs for detection of breast cancer. The tests which are conducted to detect breast cancer can be very expensive and may sometimes incur human errors by Doctors. This system will diagnose breast cancer efficiently and hence lead to prompt treatment also detect breast cancer by using a Mammogram and using Machine Learning and Data Mining techniques system can increase the accuracy thereby providing a more efficient solution and further reducing the effort and cost involved. It can be used to increase efficiency using algorithms like Adaptive Mean, GMM Segmentation and PNN Classifier. In this paper, the given system can be used to provide information on breast cancer such as signs and symptoms, treatment methods.

II. LITERATURE REVIEW (SURVEY)

In 2017,U.Ojha, S.Goel, et al. [2][11] have surveyed different data mining algorithms to detect all possible cases of breast cancer that are occurring often by Wisconsin Prognostic Breast Cancer (WPBC) dataset taken from UCI machine learning repository. Also explained different clustering and classification algorithms of data mining techniques to see the performance models used. Four clustering algorithms (K means, EM, PAM and Fuzzy c-means) and four classification algorithms (SVM, C5.0, Naive Bayes and KNN) are selected for this research[11]. In other work , this research is to identify the most useful data- mining algorithm that helps to detect recurrent breast cancer.

In 2017, Deepika Verma; Nidhi Mishra et al. [3][8] they have reviewed the literature of various data mining techniques such as classification, regression, association rules, decision tree, time series and clustering. Accuracy and Classification is done by various data mining tools, also used clustering for the same. Here it uses WEKA tool for the classification of a dataset as it is used for data pre-processing, classification, regression, clustering, association rules, and visualization as well. The WEKA Experimenter interface for determining the accuracy with the use of training set testing mode because it provides a graphical user interface for survey and experiment with machine learning algorithms on datasets is explained in this proposed paper.

In 2014, Vikas Chaurasia, Saurabh Pal et al.[4][5] have explained the Data classification process clearly by using knowledge obtained from known data sets statistics. This paper

978-1-7281-4635-5/20/$31.00 ©2020 IEEE

compares the performance of three classical decision tree classifiers that are suitable for direct interpretability of their results by means of observation and hence a new model is created. Also, this paper investigated the concept obtained by inference from specific cases performance of RepTree, RBF Network and Simple Logistic in order to enhance the prediction models for decision-making systems in the prognosis of breast cancer survival.

In 2016, AmirEhsan Lashkari ; Fatemeh Pak ; Mohammad Firouzmand,[6] have clarified the automatic technique which has been presented to help physicians in early detection of breast cancer based on various parameters. It also classifies and labeling procedures, supervised learning techniques such as AdaBoost, support vector machine, nearest neighbor, Naïve Bayes and probability neural network analysis. The experimental results gave the best mean accuracy of 88.03% for only using 0° image with combination of mRMR and AdaBoost and for combination of 3 degrees with combination of GA and AdaBoost.

In 2018, Dejun Zhang ; Lu Zou ; Xionghui Zhou ; Fazhi He et al. [7] the goal of this paper is to enhance the performance in cancer prognosis prediction and develop a more generalized outcome classifier. To achieve this, it proposes a more general way of learning features by integrating feature selection and feature extraction methods with several deep learning techniques . Construct an ensemble classifier with a boosting algorithm to strongly predict distant metastasis in breast cancer. Compared with previous classifier learning approaches, the method proposed in this paper demonstrates an unsupervised feature learning and supervised classifier learning mechanism.

In 2018, Mohammed H. Tafish ; Alaa M. El-Halees et al.[9] this paper an advance research of previous technique used to detect breast cancer, it proposed a model to help in resolving the problem faced to determine risk factor and get the use of best practices as well as factors such as short time limit, and costing the objective of advancing well-being, which is based on data collected from hospitals in the Gaza Strip. Also this paper explained models for classification techniques such as Support vector machines, artificial neural networks and k-nearest neighbors on the collected breast cancer data, which in turn predicts the fact of breast cancer[10].

In 2018, Md Faisal Kabir ; Simone A. Ludwig ; Abu Saleh Abdulla et al. [13] this paper explained how to cure breast cancer by taking a reduction of risk factor which is a significant concern. In this paper the main objective is to find invisible or unseen forms of rules from the risk factors data set of breast cancer. The most important work of data mining is to provide rules in concise statements of potentially important information so that it can be easy to understand by the patient. In this paper, association rule mining is used [10], which is a data mining technique to achieve rules from risk factor to start strategies for cancer prevention. In this, the author includes both cancer and non-cancer patients so that it can compare and do analysis of characteristics of that patient which are shown in their generated results.

III. IMPLEMENTATION

In implemented system Fig 1, consists of two main layers as seen from the above block diagram. In the first layer, the first step is taking a single image as input from the dataset. Adaptive Median Filter is applied to it. After that image segmentation is done using the Gaussian Mixture Model (GMM) segmentation. The next step is to check the intensity and extract GLCM features. Once features are extracted then the Probability Neural Network (PNN) classifier is trained and applied on the image to predict if cancer is Benign or Malignant. In the second layer, the input is taken from the database (test set) and again all the steps of applying the adaptive median filter, image segmentation, feature extraction and lastly using PNN classifier to predict the type of cancer is carried out to predict cancer.

The flow of the system includes: First the dataset is trained using the Train PNN function, once the dataset is trained the next step is Browsing the mammogram which needs to be tested. The third step is Adaptive mean Filtering followed by GMM segmentation. The final step is classifying the mammogram into three types: Benign , Malignant and/or Normal.

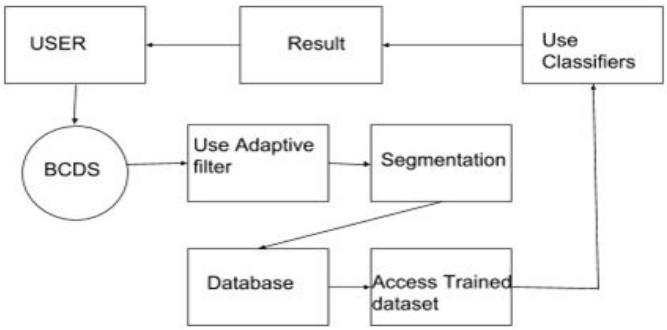

Fig. 1 Block Diagram of the System

Steps / WorkFlow of the System:

Step 1:The first step as mentioned is browsing a new input mammogram which the user needs to test and classify.
Step 2:The user needs to click on the browse function in the application to start the browsing process.
Step 3: Once the user clicks on browse a window appears and the user can then navigate to the site where the user has stored the new mammogram.

Step 4: After going to the location of the new mammogram the user click on the open button, On clicking the open button the Users' mammograms are displayed on the display axis.
Step 5: While selecting the mammogram image the user must make sure that the mammogram image is of proper resolution so that the detection process carries on seamlessly.
Step 6: Next step after browsing of the mammogram is Adaptive Mean Filter.It does the job of removing unwanted data and noise from the initial mammogram, generally called as salt and pepper noise and flattens the image to the same same intensity level by converting each pixel intensity to the same value.
Step 7: Adaptive Mean Filter achieves this by algorithm which works by first moving through each pixel and taking their intensity into consideration.
Step 8: Once the process of adaptive mean filter is complete the system displays a message and the output mammogram of the adaptive mean filter is displayed on the second display axis, which is cleaned of all the noise which was present before.
Step 9: After the completion of Adaptive Mean Filter, The user uses the GMM Segmentation function. GMM Segmentation is the last step before the classifier is applied and the result is given.
Step 10: GMM Segmentation is a combination of various functions which work together to give the output.GMM Segmentation is basically used to outline the cancerous part/pixel in the mammogram.
*Step 11:*Once the GMM Segmentation is completed the last and final step is to Classify the input into one of the three types which are Malignant, Benign and/or Normal using the PNN classifier.
*Step 12:*The cancerous part which is detected in the previous step i.e GMM Segmentation. This part is then checked with a pretrained dataset, the new image is checked to see if there are any similarities amongst the categories of pretrained data.
Step 13: If the new image shows similarities to any pre trained data that is Benign, Malignant or Normal the mammogram is classified into that category.
*Step 14:*The output is shown by mentioning the number against the following category in an output. For example, 1 specifies that the mammogram has been classified as Benign , 2 means classified as Malignant, 3 means classified as Normal part. Some of the border detection operators are as follows: Roberts operator, sobel operator, log operator, syntax: edge (image, operator).
Step 15: Once the border of the cancer part is detected, the function just highlights that part and blackens out the background which is not important. The output is then displayed on the second display axis.

IV. EXPERIMENTS AND RESULTS

The user who buys this system or product will have their unique username and password. This user name and password will be used to login into the system as shown in Fig. 2.The valid username and password will gain access to the system. However, in case of invalid username and password, a pop-up will appear if the password or username is wrong and the access into the system will be denied.

One can see the application has many functionalities and each of these functions must be executed in a particular order to get the desired and correct result, each of these functions have individual importance and failure of any one of the functions may result in incorrect output or the failure of the entire system itself. Therefore each function must be operated correctly.

The implementation is as follows: First the dataset is trained using the Train PNN function; once the dataset is trained the next step is browsing the mammogram which needs to be tested. The third step is Adaptive mean Filtering followed by GMM segmentation.

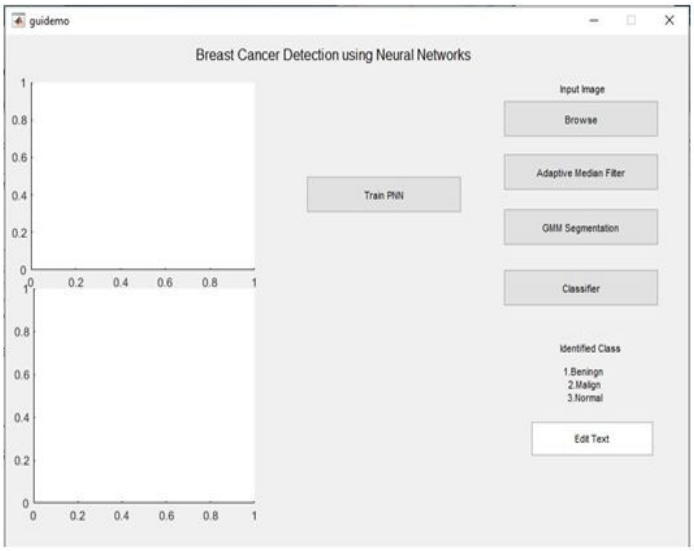

Fig. 2 Start Page of Program

The final step is classifying the mammogram into three types: Benign , Malignant and/or Normal.

The browsing of a new input mammogram in which the user needs to test and classify. The user needs to click on the browse function in the application to start the browsing process.Once the user clicks on browse a window appears and the user can then navigate to the site where the user has stored the new mammogram.

After going to the location of the new mammogram the user clicks on the open button. On clicking the open button the users mammogram is displayed on the display axis as shown in the Fig. 3 below. While selecting the mammogram image the user must make sure that the mammogram image is of proper resolution so that the detection process carries on seamlessly. Preferably the user should choose an image which has the same rows and columns resolution for example 256x256 or

320x320. Uploading this type of images makes it easier for the system to process the image at a faster speed.

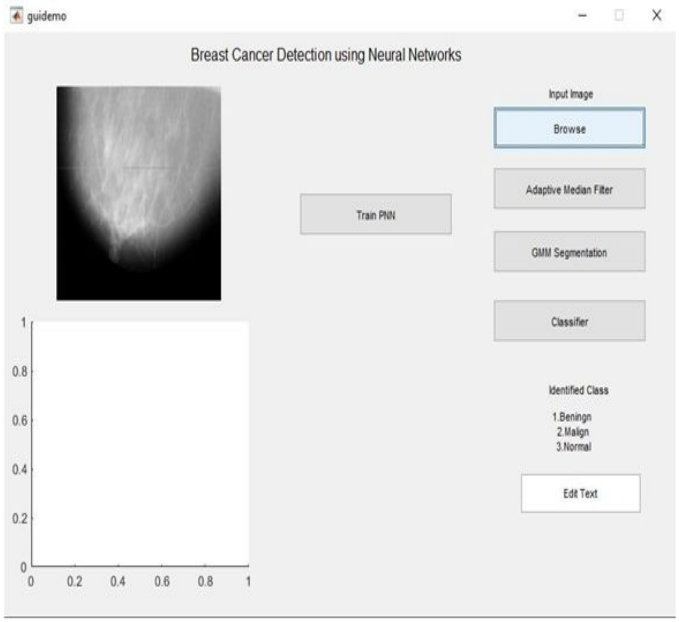

Fig. 3 Browse new Mammogram

Next step after browsing the mammogram is Adaptive Mean Filter. Adaptive Mean Filter does the job of removing unwanted data and noise from the initial

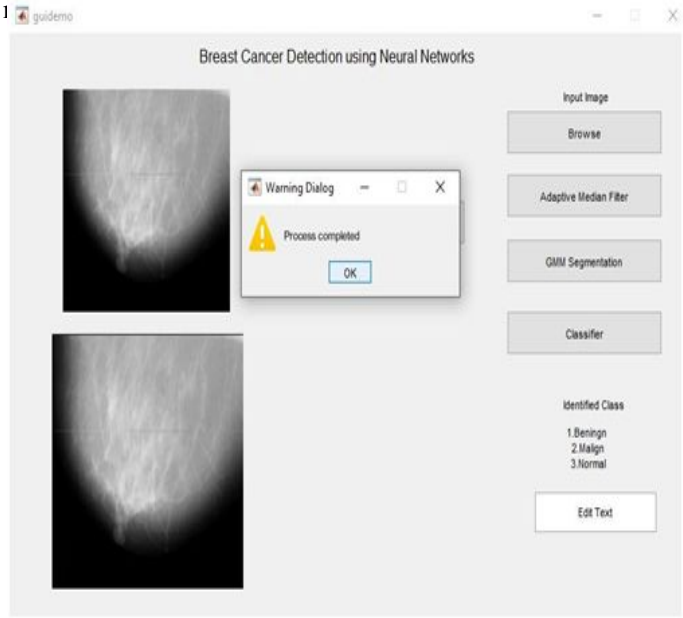

Fig. 4 Apply Adaptive Mean Filter

generally called as salt and pepper noise and flattens the image to the same same intensity level by converting each pixel intensity to the same value. Once the process of adaptive mean filter is complete the system displays a message indicating that the process is completed as seen in Fig.4 and the output mammogram of the adaptive mean filter is displayed on the second display axis, which is cleaned of all the noise which was present before.

After the completion of the Adaptive Mean Filter, The user uses the GMM Segmentation function. GMM Segmentation is basically used to outline the cancerous part/pixel in the mammogram. This outlined or highlighted area which can be seen in Fig. 5 compared to the pre-trained dataset to give the final output.

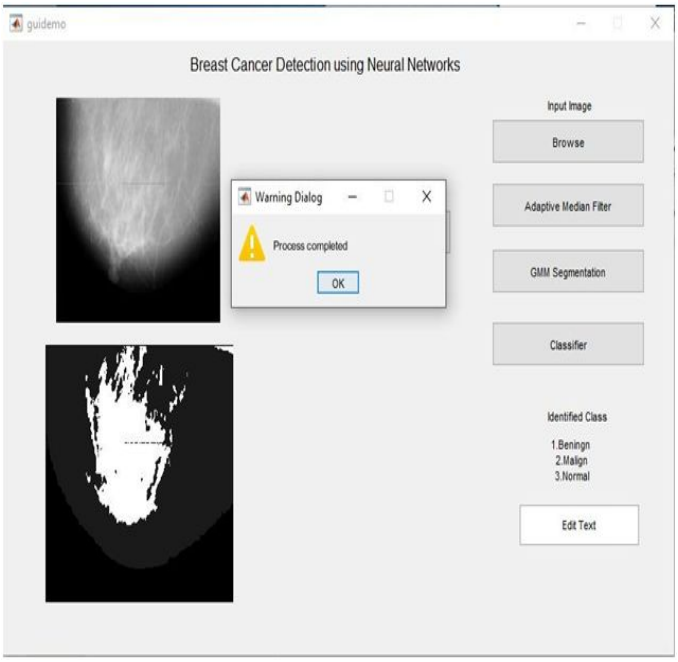

Fig. 5 Apply GMM Segmentation.

Once the border of the cancer part is detected, the function just highlights that part and blackens out the background which is not important. The output is then displayed on the

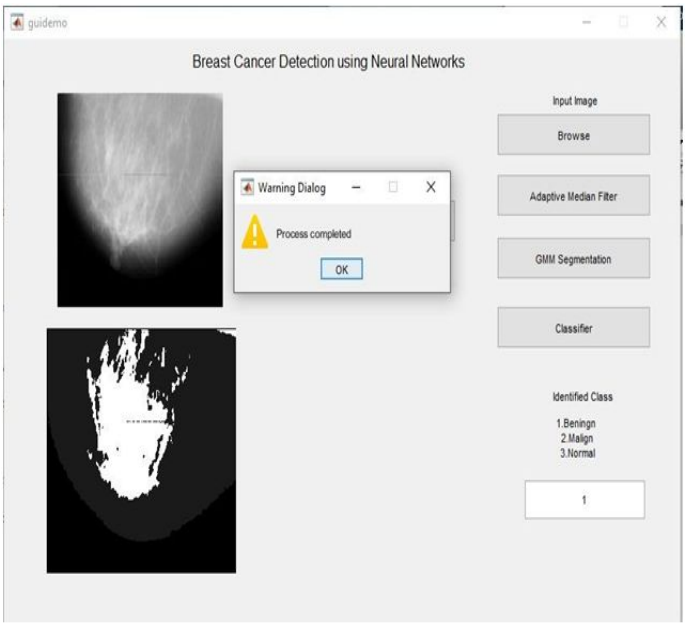

Fig. 6 Applying Classifier (Benign)

second display axis. Once the GMM Segmentation is completed the last and final step is of Classifying the input into one of the three types which are Malignant, Benign and/or Normal using the PNN classifier.

The cancerous part which is detected in the previous step i.e GMM Segmentation. This part is then checked with a pretrained dataset, the new image is checked to see if there are any similarities amongst the categories of pretrained data.

If the new image shows similarities to any pre trained data that is Benign, Malignant or Normal the mammogram is classified into that category. The output is shown in Fig. 6 by mentioning the number against the following category in an output. For example, 1 specifies that the mammogram has been classified as Benign , 2 means classified as Malignant, 3 means classified as Normal.

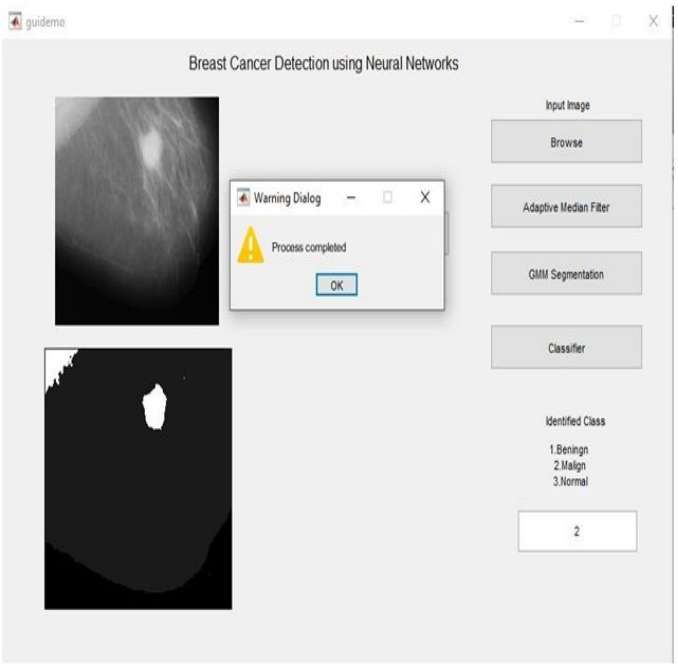

Fig.7 Applying Classifier (Malignant)

The above Fig. 7 shows the output of a Malignant mammogram. The steps before the classification process remain the same. Each and every step is followed in exactly the same way as before starting with browsing of mammograms, then Adaptive Mean Filter and GMM segmentation and then ending with PNN classification.

The below Fig. 8 shows the output of a Normal mammogram. The steps before the classification process remain the same.

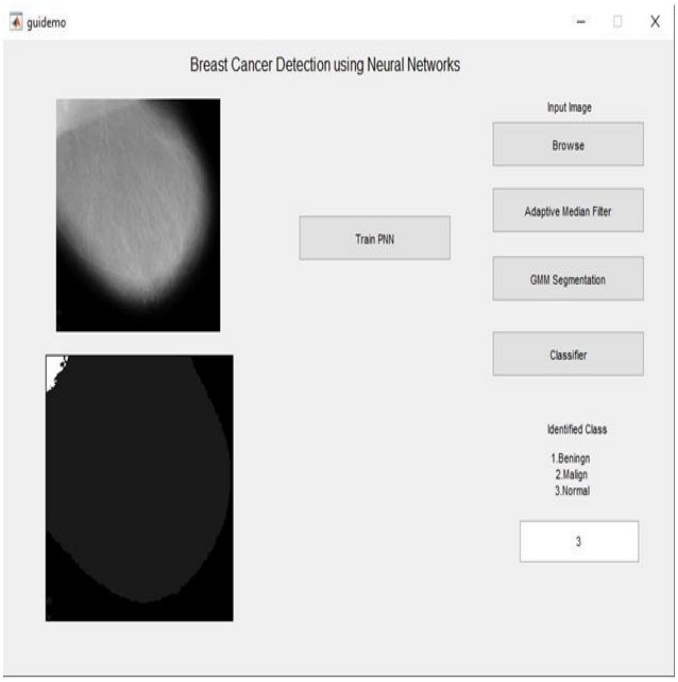

Fig. 8 Applying Classifier (Normal)

Each and every step is followed in exactly the same way as before starting with browsing of mammograms, then Adaptive Mean Filter and GMM segmentation and then ending with PNN classification.

V. FUTURE SCOPE

This system could be further enhanced if completely deployed on the cloud which will allow easy and faster access. Furthermore, time will be saved if the code is made more efficient and results are provided faster.The system could be further enhanced if the entire process of detection is done on a single click instead of using separate buttons for each step. The UI could be made more user-friendly and attractive as MATLAB only allows the development of a basic and simple UI. The mammograms provided in the current systems have to be in the specified format. If these restrictions are removed, the system can make the tasks easier and help fasten the entire process.

VI. CONCLUSION

In this paper system used various algorithms such as Adaptive mean, GMM segmentation and PNN classifier which helps to predict whether the given mammogram has Benign cells or Malignant cells which can help the patient to detect the disease at a faster rate and take appropriate actions.

REFERENCES

[1] A. Gupta,K. Shridhar and P.K. Dhillona, "A review of breast cancer awareness among women in India: Cancer literate or awareness deficit?" Eur J Cancer. 2015 Sep; 51(14): 2058–2066.

[2] U.Ojha, S.Goel "A study on prediction of breast cancer recurrence using data mining techniques" IEEE 2017.

[3] Deepika Verma; Nidhi Mishra "Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques", IEEE 2017.

[4] Vikas Chaurasia, Saurabh Pal "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability",IJCSE 2014.

[5] Shelly Gupta , Dharminder Kumar , "Data Mining Classification Techniques Applied for Breast Cancer Diagnosis" ,IJCSE 2011.

[6] AmirEhsan Lashkari ; Fatemeh Pak ; Mohammad Firouzmand,"Breast thermal images classification using optimal feature selectors and classifiers", The Journal of Engineering 2016.

[7] Dejun Zhang ; Lu Zou ; Xionghui Zhou ; Fazhi He, "Integrating Feature Selection and Feature Extraction Methods With Deep Learning to Predict Clinical Outcome of Breast Cancer", IEEE Access 2018.

[8] Mr. Chintan Shah ; Anjali G. Jivani, "Comparison of data mining classification algorithms for breast cancer prediction", IEEE ICCCNT 2013.

[9] Mohammed H. Tafish ; Alaa M. El-Halees,"Breast Cancer Severity Degree Prediction Using Data Mining Techniques in the Gaza Strip" ,ICPET 2018.

[10] Umesh D R ; B Ramachandra ,"Association rule mining based predicting breast cancer recurrence on SEER breast cancer data", ICERECT 2015.

[11] Runjie Shen ; Yuanyuan Yang ; Fengfeng Shao,"Intelligent Breast Cancer Prediction Model Using Data Mining Techniques", Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics, 2014.

[12] Mohammad Ashraf ; Kim Le ; Xu Huang, "Iterative weighted k-NN for constructing missing feature values in Wisconsin breast cancer dataset",The 3rd International Conference on Data Mining and Intelligent Information Technology Applications,IEEE Xplore December 2011

[13] Md Faisal Kabir ; Simone A. Ludwig ; Abu Saleh Abdulla on "Big Data (Big Data) Rule Discovery from Breast Cancer Risk Factors using Association Rule Mining", IEEE International Conference ,2018