Classification Methods to Improve Performance in Breast Cancer Screening

Snehal Mane and Vandana Jagtap

February 17, 2022

Classification Methods to Improve Performance in Breast Cancer Screening

Snehal Mane
Department of Computer Engineering
Maharashtra Institute of Technology, Pune
Pune, India
s30mane@gmail.com

Vandana Jagtap
Department of Computer Engineering
Maharashtra Institute of Technology, Pune
Pune, India
vandana.jagtap@mitpune.edu.in

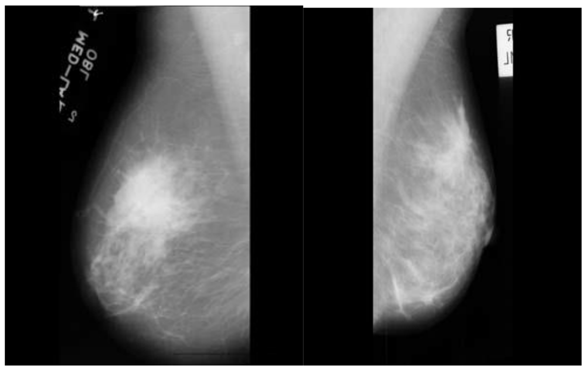
***Abstract*—Breast cancer is a very aggressive type of cancer with a very low median survival. Today the deaths of women in the age group 15-55 are increasing because of malignant cells are increasing in breast. For the death of women it is the main cause. So, the possibility of improvement is only the early diagnosis of patients. Machine Learning (ML) techniques can assist the physicians by expanding tools for detection at initial stage and analysis of breast cancer thus increasing the probability of patient's survival [1]. At present, mammography is the best imaging strategy utilized by radiologist for screening breast tumours. In this paper, author proposes a system using different classification method like Support Vector Machine (SVM), Naive Bayes, Decision tree and MLP (Multi-Layer Perceptron) for early detection of cancer. Propose system extracts the texture based features and shape based features using LBP, GLCM, Otsu, Compactness, Fourier Transform. The main focus of the presented work is on application of MLP for breast cancer classification. In addition real time data has been used to improve accuracy. Proposed system will do the comparative study between both datasets by extracting the feature with and without removing pectoral muscles.**

Keywords— Breast cancer, Mammography, SVM, Decision tree, ANN, Naive Bayes, and MLP, pectoral muscles

I. INTRODUCTION

Today, breast malignant growth is one of the cancer common amongst both men and women. The malignant cells which grow in the breast tissue are describing breast cancer. The dangers of breast disease increase with elements, for example, female sex, obesity, absence of physical exercise, delay in having children, etc. Breast tumor is caused by irregular growth of breast cells. The rapid development of these cells occurs in benign and malignant tumors. Expanding number of cells in benign tumors stops at a characterized stage, but it keeps on developing in malignant tumors until all parts of the body are affected. Prior menstruation in young age, menopause in older age and late marriages increase the possibility of breast cancer. The preventive drugs and hormones, nourishment and way of life are very useful in breast cancer. Early indications of breast malignancy are frequently not perceived or seen by patients. Therefore, many patients already are seeking severe treatment. Breast cancer is characterized by lumps in the breast. The lump can be benign or malignant. One way is the early detection of such breast self-examination should be carried out by women to avoid breast cancer. Technological development, particularly in the field of medical imaging, has made the identification of cancer, like breast cancer. Malignancy can more quickly detect any of them on a mammogram.

A mammogram is a device that can deliver two-dimensional, 8-bit gray-scale images are normally captured from X-rays of breast patients. Generally, the specialist can recognize anomalies in the breast by taking a features at the highlights of the mammogram picture. These properties are manifested in the fact that the left and right chests are not equal, there are lumps, and the breast tissue structure is broadened and contains micro-calcifications. From the mammogram picture, the doctor conducts a conventional investigation or straight forwardly determined malignant growth cells to have the unaided eye. But sometime there is an inaccuracy while investigating (because of fatigue or human mistake), as well as losing something important from the eyesight doctor. Based on this background, we will develop strategies that can be utilized to recognize the existence of breast cancer. A portion of the strategies that are created to identify the existence of breast disease are affirming imaging anomalies in mammography and the recognition of breast tumors. Image acquisition, segmentation process and classification are followed in a hierarchy for classifying breast cancer. Figure 1 shows the cleared Mammogram Classification.



(a) (b)

Fig.1 Mammogram Classification

Purpose of the paper is to (1) Discover the region of interest or suspected territory from a mammogram image, (2) Extract the feature from a mammogram image (3) Classification of ordinary and abnormal classes in the mammography image (4) From the classification result the level of exactness is determined (5) The test data is determined from the classification result to the level of accuracy.

II. REVIEW OF LITERATURE

Breast cancer is known as the reason of death in women, and the number of patients diagnosed with breast cancer are increasing. Therefore, early diagnosis can increase the chances of recovery. The paper aimed at extracting features without removing pectoral muscles at the preprocessing stage, using a new effective method. Firstly, Farzam et al applied discrete wavelet transform on the image and approximation matrix is changed into 1-Dimension utilizing zigzag scanning and finally volatile signal features are removed. Limitations of this paper are it requires large number of cases [2].

Clustering classification of breast micro calcification into the Malignant and Benign category is challenging task for the computerized algorithm. In this paper [3] Alan et al used Multi View Classification that is craniocaudal (CC) and mediolateral oblique (MLO) views for the classification of micro calcification and it is implemented using Logistic Regression Classification. They were focusing on the benefits of combining the views. This DDSM dataset includes demographic data of the patients. Limitations are it gives poor accuracy.

In this paper [4],for segmenting a digital mammogram image Farag H. Alhsnony et al proposed a new automated technique and detects the region of pectoral muscle by utilized a bit depth and edge processes systems. For more accuracy CAD system requires.To determine the breast cancer, the 3rd order fitting curve could be used. Limitation is recognition of pectoral muscle region in mammogram image needs a series of mathematical analysis method to be recognized.

Nan Wu et al [5] presented a Deep Convolutional Neural Network for breast cancer screening exam classification. Over 200,000 exams were trained and evaluated. It was shown that a hybrid model is more accurate to predict probability of malignancy than considering the two separately. Deep CNN was used for cancer classification.

In this paper, Y. Ali has discussed about the widely used image processing methods developed for identification of masses and calcifications. There are poor contrast and low visibility in the mammographic images. So, for efficient treatment of the disease, prior identification of breast cancer is a important step. It is a method to classify benign and malignant masses and to detect the microcalcification in mammography[6].

In this paper, for automated breast ultrasound lesions detection, Moi Hoon Yap et al used Convolutional Neural Networks(CNN). For breast lesion identification and investigates three different use of deep learning approaches: a Patch-based LeNet, a U-Net, and a transfer learning approach with a pretrained FCN-AlexNet have been proposed. They were working on two datasets. The limitation of this methods is that they require a training process and negative images in the experiment[7].

M. R. Al-hadidiet al[8] proposed model consists of two parts, the first one is using image processing techniques for feature extraction where the second part was the machine learning algorithms. They are divide in two types of supervised learning algorithms, Logistic Regression (LR) and Back Propagation Neural Network (BPNN). They observed that, the number of features used in LR model was much higher than with the BPNN.

$$P(x) = \frac{1}{1 + e^{-(\beta 0 + \beta 1x)}}$$

Where,

P(x) = is interpreted as the probability of the dependent variable.

X= Design Matrix

β= Shared Parameters

This paper[9] developed and optimized machine learning models for whole image classification mammograms. Xiaofei Zhang et al evaluated 7 different CNN architectures which concluded that combining both data augmentation and transfer learning methods with a CNN is the most effective in improving classification performance. Large training dataset and sufficient memory space for training results in most optimum results.

In this advanced technology new possibilities are coming from a scientist to gathering multimodal data in the different applications such as Medical Imaging, Brain/Body Machine Interface, Bioimaging, and Omics. In this paper [10] Fabian et al provides a deep study on utility with dissimilar biological data and relative study of the deep learning techniques, Reinforcement. Heavy computing memory and power are desired by this method is the main limitation.

Hassan Jouni et al [11] used ANN as it has been proven to be efficient for its ability of learning and generalizing from data. Low-complexity architecture for the back propagation (BP) neural network was proposed. Optimal activation function which minimizes the classification error with less number of blocks was the focus of the study. This results in reduction of the complexity of the implementation with CMOS technology. Implementation in CMOS, the building blocks of the artificial neuron is the next step.

In this paper, the whole study concentrates on different classification methods can be used after preprocessing and subdivision processes and can also be labeled to improve image accuracy results. KamalakannanJ used appropriate methods for mammogram images like decision tree and K-nearest are discussed. Next to it is a fuzzy Bayes, ANN (Artificial Neural Network) collaborative-support vector machine. For every classification, they consider factors like sensitivity, specificity, and accuracy that are selected in accordance with the suitable scenario [12].

III. METHODOLOGY

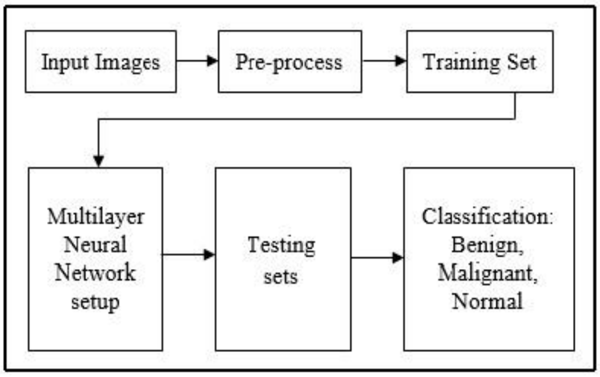


Fig.2. MLP-NN model for Classification of breast malignant growth[1]

A. MLP (Multilayer Perceptron)

The mammogram images can be preprocessed to improve its quality and find the exact region of interest. Medical imaging may contain undesirable noise; therefore, medical imaging is tortured to improve image quality. The improved images are then tagged and entered as training data sets to enter ML-NN for purpose of training.

The main objective of the work presented here is study the operation of MLP to classify breast cancer. The MLP consists of neurons called perceptron. In regard to the weights of neurons in the input nodes and the generation of the output by the employment of the mathematical function of activation is not linear, the linear combination is generated by the perceptron by computing a neuron of output from multiple inputs valued real[1].

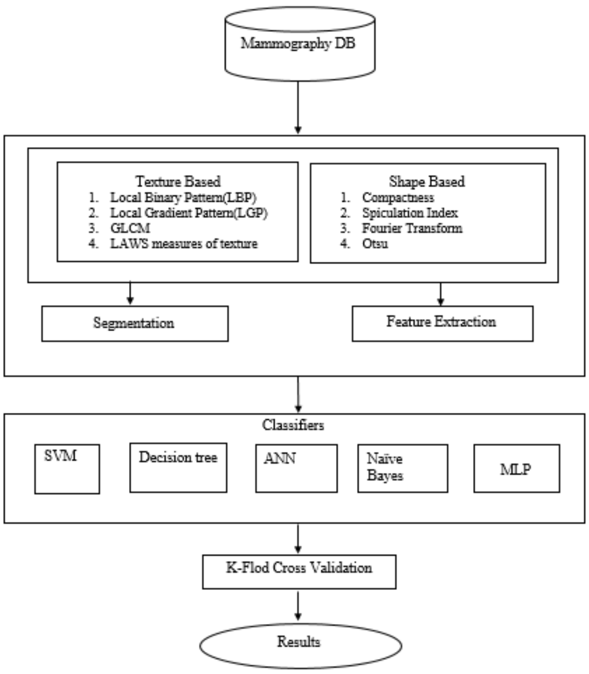


Fig.3 System Architecture

Generally, MLP is adopted to solve problems involving supervised learning. It is indicated that a training dataset with input output combination pairs and tags are using for input. Therefore, the MLP should determine on the basis of training equipment.

The authors propose to introduce a new technique for early prediction and comparative study of extracting the feature with and without eliminating pectoral muscle in preprocessing phase. In proposed system, results will be compared with the other techniques of classification. The comparison will show the accuracy and performance of system for prediction. As a result, proposed architecture should be able to predict cancer at early stage.

B. Algorithm

Input: I = Mammogram image from medical images dataset

Output: P = Detection of cancer from abnormalities

Begin

1. From the medical image dataset, read image I.
2. Apply LBP algorithm on that images.
3. After that we perform GLCM algorithm. It describes the texture of image by calculating precise pixel values and then removing statistical measures from this matrix. Calculated number of pixels, variance and entropy.
4. For comparison purpose applied compactness algorithm for detecting edge, Fourier Factor/ Fourier Transform.
5. Applied Otsu algorithm. It shows the nucleus in image with the count of pixels in that nucleus.
6. Detect the Growing part in the image.
7. Find out the shape of growing part calculates the area of Growing part.
8. Detect the type of abnormalities.

C. Dataset Used

Two datasets are used to perform the proposed research work. A digitized dataset of the mammographic image analysis society (MIAS) and digital non public which is actual medical images dataset of patients collected from cancer center are the two data sets.

- MIAS Dataset

  Mammograms from this datasets are caught in the Medio Lateral Oblique (MLO) view and scan at a resolution of 0.05 mm pixel size with a thickness. UK Breast Screening Program has a total of 322 digitized images. The 322 images contains data of 161 patients in the MIAS dataset. Therefore, the

MIAS dataset consists of 322 mammogram images that are divided between 208 normal patients, 66 benign and 48 malignant. There are total 48 malignant cases and total 66 benign cases. For a normal breast, there are 3 types of breast arrangements: Fatty, Fatty glandular and dense glandular [26].

- Medical Images Dataset

This dataset is collected from Magnolia Breast Health Center, Pune. There are total 110 images. There are 4 images of every patient. It includes left and right Craniocaudal view (CC), left and right Medio Lateral Oblique view (MLO).

Process:

1) P1= {I};

- Convert image into gray scale.

$$C_{Linear} = \begin{cases} \frac{C_{srgb}}{12.92} & C_{srgb} \leq 0.04045 \\ \left(\frac{C_{srgb} + 0.055}{1.055}\right)^{2.4} & C_{srgb} > 0.04045 \end{cases} \quad (1)$$

Where,

C_{srgb}= Shows any of the three gamma-compressed sRGB primaries (Rsrgb, Gsrgb, and Bsrgb, each in range [0,1])

C_{linear}= linear-intensity value

$$Y_{linear} = 0.2126 R_{linear} + 0.7152\, G_{linear} + 0.0722\, B_{linear} \quad (2)$$

$$Y_{srgb} = \begin{cases} 12.92\, Y_{linear} & Y_{linear} \leq 0.0031308 \\ 1.055\, Y_{linear}^{1/2.4} & Y_{linear} \leq 0.0031308 \end{cases} \quad (3)$$

Where,

Y_{srgb}= inverse of the gamma expansion

2) P2= {I};

Apply OTSU method.

$$\sigma_{\omega}^{2}(t) = \omega_0(t)\,\sigma_0^2(t) + \omega_1(t)\,\sigma_1^2(t) \quad (4)$$

Where,

ω_0= Weights, ω_1= are the probabilities of the two classes separated by a threshold t and σ_0^2 and σ_1^2 are variances of these two classes.

3) P3= {P2};

- Calculate value of pixel (x_c, y_c)

$$LBP = \sum_{p=0}^{p-1} s(g_p - g_c) 2^p \quad (4)$$

$$S(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & otherwise \end{cases} \quad (5)$$

4) P3= {I};

- Perform Fourier Transform.

$$f(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi}\, dx, \quad (6)$$

Where,

f=Fourier transform of the function

$f(x) = e^{-2\pi i x \xi}$=a signal with zero initial phase

ξ_0= frequency

5) P4= {I};

- Compactness Algorithm

$$\Psi = \frac{(Surface\ area)^{1.5}}{(Volume)} \quad (7)$$

6) 6) P5= {P3,P4};

- GLCM feature

$$Energy = \sum_{i,j=0}^{N-1} (P_{ij})^2 \quad (8)$$

$$Entropy = \sum_{i,j=0}^{N-1} -ln\,(P_{ij}) P_{ij} \quad (9)$$

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij} (i-j)^2 \quad (10)$$

Where,

P_{ij}=Element i,j of the normalized symmetrical GLCM

N = Determines the number of gray levels in the image which is as given by the number of levels in under Quantization on the GLCM texture page of the variable properties dialog box.

IV. RESULTS AND DISCUSSION

A. *Experimental Setup*

Hardware and software of proposed system given below:

- Software Technology:

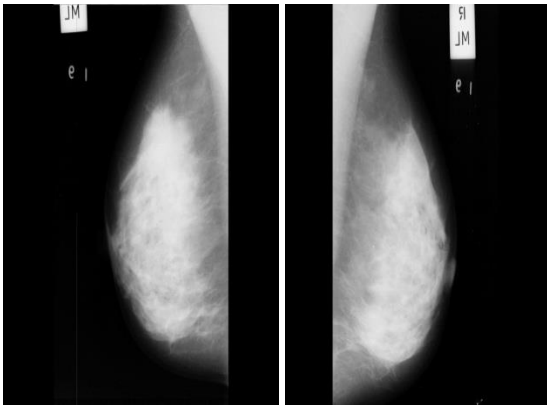
1. Technology: Core Python
2. Tools: Jet- JetBrains, Paycharm
3. Operating System: Windows 7

- Hardware Technology

1. Processor: 1.0 GHz
2. RAM: 1 GB
3. Hard Disk: 730 GB

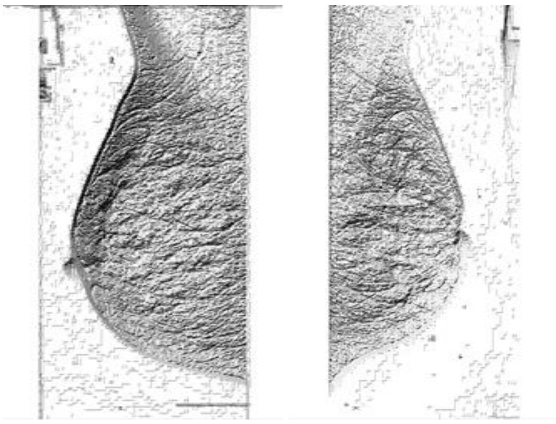
B. *Experimental Results*

Successfully completed some preprocessing and classification steps on MIAS dataset. Fig 4 contains original dataset images. On that data our novel approach that is Local Binary Pattern (LBP) method helps to convert it to binary format that is 0(black) or 1(white). It is shown in fig 5. Next GLCM (Gray Level Co-occurrence Matrix) method is used, shown in fig 6. It shows features like entropy, correlation, variance and angular moment. Hence LBP and GLCM both methods are finding texture features of images. After that Otsu, watershed compactness and Fourier transform algorithm is applied for finding the shape based features, shown in fig 7, fig 8, fig 9. Otsu method shows nucleus part and it gives count of pixels included in that nucleus.



mdb003 mdb004

Fig.4 Original data



mdb003 mdb004

Fig.5 Performing LBP

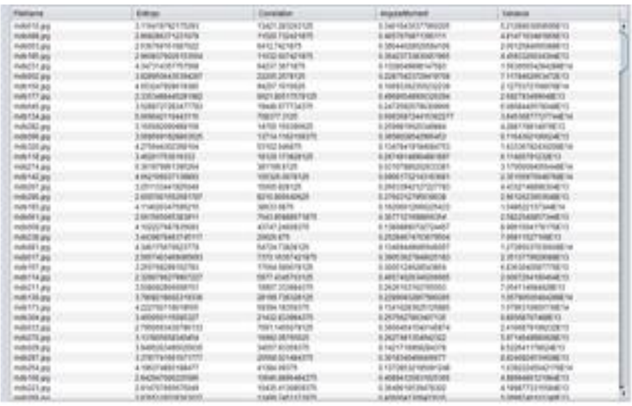
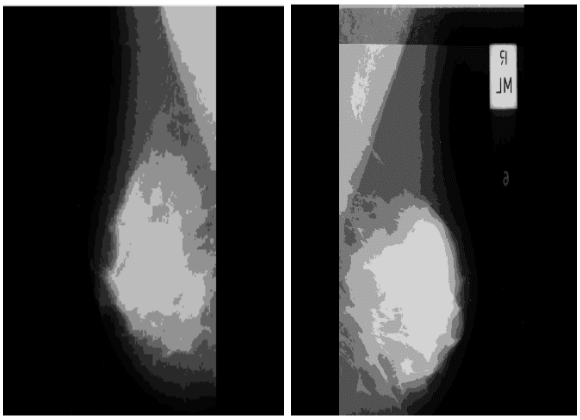


Fig.6 Performing GLCM

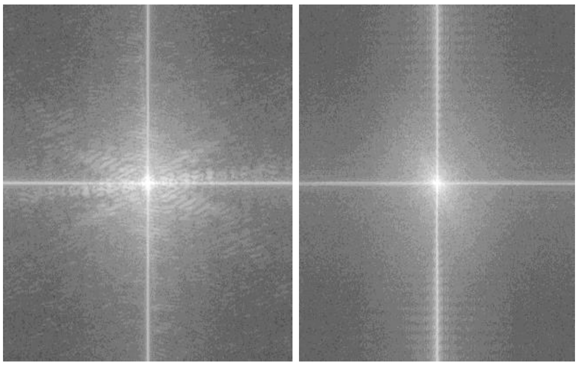


mdb003 mdb004

Fig.7 Performing Compactness

mdb003 mdb004

Fig.8 Otsu method



mdb003 mdb004

Fig.9 Fourier Transform

After applying all this feature extraction methods in preprocessing stage, classification algorithms like SVM, Naïve Bayes, Decision Tree, MLP are applied on both dataset.

Table II shows comparision between different preprocessing methods with different machine learning algorithm in terms of accuracy on MIAS dataset. Proposed MLP with Fourier Transform, Otsu gives more accuracy than other combinations with 94% and 97% accuracy respectively.

TABLE II. ACCURACY COMPARİSON OF PREPROCESSİNG METHODS WİTH MACHİNE LEARNİNG ALGORİTHM ON MIAS DATASET

Algorithm	**SVM**	**Naïve Bayes**	**Decision Tree**	**MLP**
LBP	80	80.76	81	80
Fourier Transform	87.39	84	82	84
Watershed Compactness	75.22	78.61	84	80
Otsu	47.69	83	84.33	88.66

Table III shows output on Real-time dataset. It also shows same algorithms like MIAS dataset. Proposed MLP with Otsu gives more accuracy than other combinations with 98.5% accuracy respectively. Real time dataset gives more accuracy than MIAS dataset.

TABLE III. ACCURACY COMPARİSON OF PREPROCESSİNG METHODS WİTH MACHİNE LEARNİNG ALGORİTHM ON REAL-TİME DATASET

Algorithm	**SVM**	**Naïve Bayes**	**Decision Tree**	**MLP**
LBP	78.26	82.25	84	86.60
Fourier Transform	82.60	85.3	85	86.6
watershed compactness	69.56	83.12	86	88.7
Otsu	89.2	83.12	88.9	91.66

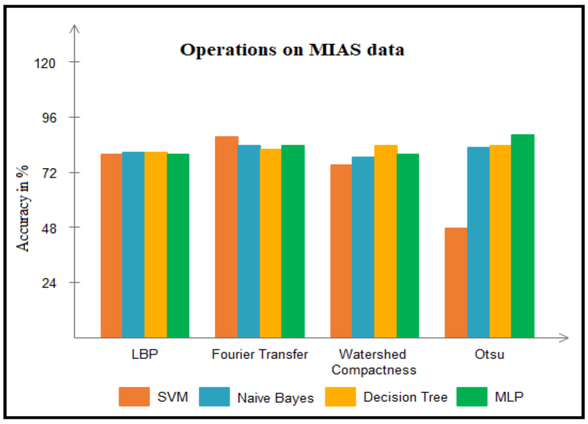


Fig. 10 Time Comparison Graph of preprocessing methods with Machine Learning Algorithms on MIAS dataset

Fig. 10 shows time comparision graph of different preprocessing methods with different machine learning algorithms on MIAS dataset. Proposed system gives more accuracy for Otsu, watershed compactness with MLP.

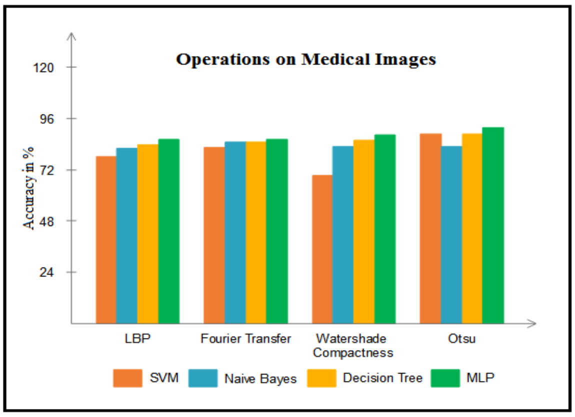


Fig. 11Time Comparison Graph of preprocessing methods with Machine Learning Algorithms on Real-time dataset

Fig. 11 shows time comparison graph of real time data. Proposed system gives more accuracy for Otsu, watershed compactness with MLP.

V. CONCLUSION

This paper includes the new method for early diagnosis of breast cancer. The main goal is to develop techniques which can be helpful for early detection and comparative study of extracting the feature with and without eliminating pectoral muscle in preprocessing phase using new method. The main contribution is to perform strong preprocessing and powerful feature extraction and selection before classification. For that this paper compares the combination of preprocessing method and machine learning algorithm. From study it was found that for texture feature-LBP and GLCM and for shape-Compactness, Otsu and Fourier Transform methods are available. But Otsu with MLP combination gives best results in both datasets. In that also Real-time dataset gives more accuracy than MIAS dataset. Hence MLP with recurrent neural network provides more accuracy as compared to other classifiers. Proposed model is used to predict the cancer at early stage whether it is benign, malignant or normal patient. Results can be improved using without removal of pectoral muscle from mammograms which minimizes the computation. For that Real-time dataset with pectoral muscles has been used.

REFERENCES

[1] Snehal Mane, Vandana Jagtap, “Early Prediction of Breast Cancer from Mammogram Images using Classification Methods: A Comparison”, The Internet of everything: Advances, Challenges and Applications, Web of Science, Scopus indexed, 2019.

[2] F. K. Nezhadian, and SaeidRashidi, “Breast cancer detection without removal pectoral muscle by extraction turn counts feature”, Artificial Intelligence and Signal Processing(AISP), 2017.

[3] A. J. Bekker, M. Shalhon, H. Greenspan and J. Goldberger, “Multi-view probabilistic classification of breast microcalcification.” IEEE Transaction on Medical Imagine, 2015.

[4] Farag H. Alhsnony, Maher G. M. Abdolrasol, and Samei G. M. Abadelrsool, “Auto-Identification of Pectoral Muscle Region in Digital Mammogram Images”, International Journal of e-Education, e-Business, e-Management and e-Learning, February 2014.

[5] Nan Wub, Jason Phangb, Jungkyu Parkb et al, ” Deep Neural Networks Improve Radiologists’ Performance in Breast Cancer Screening”, Cornell University, March 2019.

[6] Yadollahpour Ali, Shoghi Hamed, “Early Breast Cancer Detection using Mammogram Images: A Review of Image Processing Techniques”, biosciences biotechnology research ASIA, Vol. 12(Spl. Edn. 1), March 2015.

[7] Moi Hoon Yap, Gerard Pons, Joan Marti, Sergi Ganau, Melcior Sentis, Reyer Zwiggelaar, Adrian K. Davison and Robert Marti, “Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks”, IEEE journal of biomedical and health informatics, vol. 22, no. 4, july 2018.

[8] Moh’d Rasoul Al-hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah, “Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm”, 9th International Conference on Developments in eSystems Engineering, 2016.

[9] Xiaofei Zhang, Yi Zhang, Erik Y. Han, “Whole Mammogram Image Classification With Convolutional Neural Networks”, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017.

[10] E. Romero, J. Tarquino and F. Naraez “Applications of Deep Learning and Reinforcement Learning to Biological Data”, IEEE Transactions on Neural Networks and Learning Systems, 2018.

[11] Hassan Jouni, Mariam Issa, Adnan Harb, Gilles Jacquemod, Yves Leduc, “Neural Network Architecture for Breast Cancer Detection and Classification”, IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), 2016.

[12] Kamalakannan J and Abinaya Vaidhyanathan, “Study on Different Classification Technique for Mammogram Image”, IEEE, International Conference on Circuits, Power and Computing Technologies, July 2015.

[13] Qiao Pan,Yuanyuan Zhang, Dehua Chen, Guangwei Xu, “Character-based Convolutional Grid Neural Network for Breast Cancer Classification” IEEE 2017 International Conference on Green Informatics, 15-17 Aug, 2017.

[14] Narain Ponraj, Poongodi Merlin Mercy, “Texture Analysis of Mammogram for the Detection of Breast Cancer using LBP and LGP: A Comparison”, IEEE Eighth International Conference on Advanced Computing, 2016.

[15] Michelle Dcruz, Dr. TanujaSarode, “Feature Exraction in Mammograms Using NSCT and LAWS Texture Analysis Approach”, International Journal of Engineering Research and Application, 2248-9622, Vol. 7, Issue 8, (Part-6) August 2017.

[16] D. Saranyaraj, M. Manikandan, “Medical Image Processing to Detect Breast Cancer – A Cognitive-Based Investigation”, IEEE 4th International Conference on Signal Processing, Communications and Networking (ICSCN -2017), March 16-18, 2017.

[17] Pengcheng Xi, Rafik Goubran and Chang Shu “Abnormality Detection in Mammography using Deep Convolutional Neural Networks”, arXiv:1803.01906v1 [cs.CV], 5 Mar 2018.

[18] Ibrahima Faye, Muhammad Awais, Fabrice Meriaudeau and Syed Jamal Safdar Gardezi, “Mammogram classification using Deep learning features”, IEEE 2017.

[19] Abu Baker, Ayman A., R.S .Qahwaji, Musbah J. Aqel, Hussam Al-Osta, and Mohmmad H. Saleh. "Efficient Pre-processing of USF and MIAS Mammogram Images", Journal of Computer Science, 2007.