International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

MACHINE LEARNING TECHNIQUES TO IMPROVE BREAST CANCER EARLY DETECTION ON SCREENING MAMMOGRAPHY

Bhoomika M P[1], K R Sumana[2]

[1]PG Student, Department of MCA, The National Institute of Engineering, Mysuru, Karnataka, India
[2]Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India

---

Abstract - *Cancer is the uncontrolled growth of cell groups in specific parts of the body and is the second largest disease in the world that causes the death of women. If the disease is detected early, it can be completely cured. Much research has been done to find the tumor accurately, but no exact method has been found. Research for breast cancer detection using digital image processing is not new, but many new approaches are being considered in this area to accurately predict tumor sites. The current approach is to visually detect the tumor area and identify the area where the tumor is predominantly concentrated. This work is mainly focused on finding the best algorithm for detecting tumors in the breast. Although a machine learning algorithm is applied to the proposed work, it is an adaptive median filter, GMM algorithm, and GLCM feature extraction method that are optimal for cancer detection.*

Key Words: Breast Cancer, Tumor, Digital Image Processing, Machine Learning, Adaptive Median Filter, GMM.

1. INTRODUCTION

When healthy cells in the breast change and enlarge out of control, you have a tumor. This is a mass or sheet of cells. Tumors are likely to be malignant or non-cancerous. Malignant tumors are tumors that can grow and spread to other areas of the body. The term 'benign tumor' refers to a tumor that can develop but cannot spread. Breast cancer invades nearby organs and other parts of the body, or enlarges for breast cancer cells to migrate through blood and lymphatic vessels to other parts of the body. This is called metastasis. This guide describes non-invasive (stage 0) early and locally advanced invasive breast cancer (stages 1, 2, 3) and non-invasive (stage 0). The stage of breast cancer indicates how far the cancer has grown and whether or where the cancer has spread.

2. LITERATURE REVIEW

This treatise presents a new method for detecting breast cancer using machine learning techniques. The author performed an experimental analysis of the dataset to assess performance. The proposed method gave very accurate and efficient results [1].

We train and evaluate a set of powerful neural networks on mammography datasets labeled with huge size histological examinations proven by medical imaging standards, let alone breast cancer screening. Using two free label types, a breast-level label indicating whether each breast has benign or malignant findings and a pixel-level label indicating the location of the biopsied malignant and benign findings [2] .. Through a breast tissue biopsy, a pathologist can histologically assess factors with the microstructure of the tissue. Histopathology aims to evaluate the prognosis by distinguishing between normal tissue, non-malignant (benign) and malignant lesions (cancer) [3].

3. PROPOSED SYSTEM

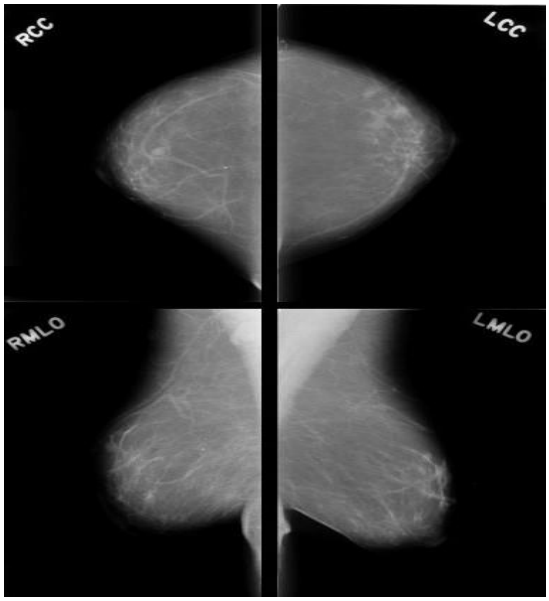

Fig-1 Multiview breast Mammogram of a patient.

We use image processing applications to automate breast cancer diagnostic systems. MATLAB tools are used to train and test medical images that detect breast cancer. Feature extraction methods such as GLCM are also used to extract features of medical images, and partitioning techniques such as the GMM clustering algorithm are used to cluster input images into different groups. Adaptive median filters are used to remove noise and blur in the input image.

4. IMPLEMENTATION

4.1 Datasets

The dataset has 32 parameters. All parameters can be useful for the categorization of cancer. If these parameters have a relatively large value, it may be a marking of the malicious organization. The first parameter is ID and is the number used for identification. The second parameter is the diagnosis of the membrane and there are two diagnoses of the organization: malicious and positive. For other types of cancer, it is necessary to determine the exact diagnosis of the fabric if two membranes have different therapies.

4.2 Pre-Processing Techniques

The **Adaptive Median Filter** is designed to eliminate the drawbacks encountered by the standard median. The main advantage of the adaptive median filter is that the size of the kernel around the damaged image can be changed, thus achieving better results. The other main advantage of the adaptive filter is that unlike the median filter, it does not replace all pixel values with the average. The operation of the adaptive filter is a two-step process; in the first step it finds the average value for the kernel and in the second step it checks if the current pixel value is a pulse (salt and pepper noise). If the pixel value is corrupted, it will change its value with the average or it will keep the pixel value of the gray scale. In this way, it can be ensured that only pixels with pulse noise will be changed while all other pixel values remain the same.

A **Gray Level Co-Occurrence Matrix (GLCM)** is a matrix whose rows and columns are equal to the number G of gray levels in the image. Color-oriented shape-oriented various shapes are extracted using MATLAB commands. These features play an important role in determining the breast area. Various features such as mean, standard deviation, and variance are extracted to know the health status of the breast.

Gaussian mixture model (GMM) is also a kind of clustering algorithm. As the name suggests, each cluster is modeled according to a different Gaussian distribution. This flexible and probabilistic approach to data modeling means assigning in software rather than hard assigning from clusters like k- means. This means that each data point is generated by a random distribution with that probability. In fact, each distribution has a "responsibility" for generating certain data points.

5. RESULT ANALYSIS

5.1 Benign

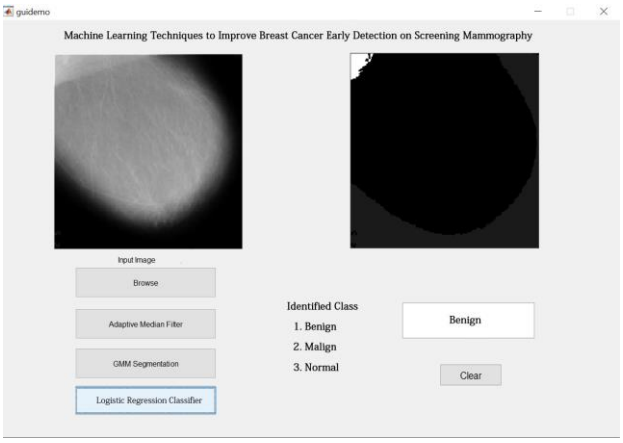

Fig –2 Infected Result 1

Analysis: Breasts infected with benign tumors were loaded from the database. Contrast enhancement and image preprocessing are done in the second stage. In the image segmentation column, one of the clusters is loaded. Figure 2 above shows the disease classified as mild. Additionally, the area of the affected area is displayed.

5.2 Malign

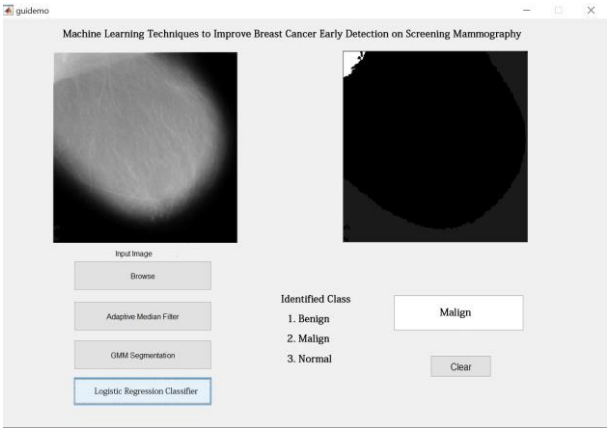

Fig -3 Infected Result 2

Analysis: Malign infected breasts are loaded from the database. Contrast enhancement and image preprocessing are done in the second stage. In the image segmentation column, one of the clusters is loaded. Figure 3 above shows the disease classified as malignancy. Additionally, the area of the affected area is displayed.

5.3 Normal

Analysis: When images of healthy breast quality were taken, the results were disease-free and were successfully classified with 96% accuracy. Figure 4 below shows how the disease is successfully classified.

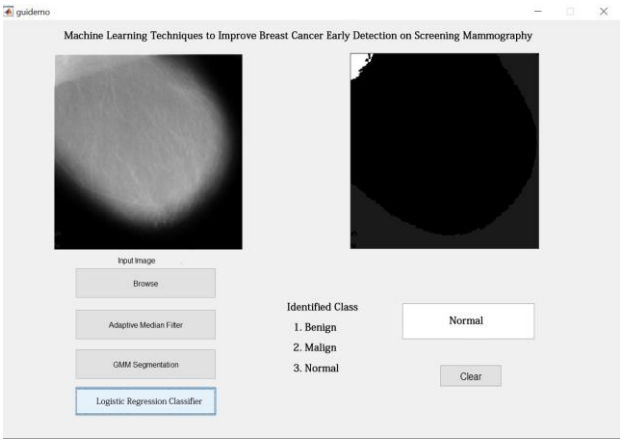

Fig - 4 Infected Result 3

6. CONCLUSIONS

Machine learning techniques play an important role in the medical industry and other fields. Image processing algorithms are easy to implement, we can solve complex problems using these techniques. Using image processing techniques and machine learning algorithms, we were able to identify the affected area in the breast. Various features of the image are extracted along with the numerical values. The algorithm used here is very efficient, the best case is the complexity of the construction. The average time it took to group and process 90% of the images was less than 20 seconds. In addition, the proposed algorithm was tested on the mini-MIAS database. The results obtained were reported to be 98% and 97%, respectively, compared to the completeness and accuracy of pectoral muscle resection. Summing up, these results suggest that the proposed method is very suitable for improving the quality of mammogram images in Auto CAD system.

REFERENCES

[1] Anji Reddy V., B. Soni and Sudheer Reddy K., "Breast cancer detection by leveraging Machine Learning", Year: 2020.

[2] Alexander Rakhlin1, Alexey Shvets2, Vladimir Iglovikov3, and Alexandr A. Kalinin4, "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis", Year: 2018.

[3] Nan Wu2, Jason Phang, "Deep Neural Networks Improve Radiologists Performance in Breast Cancer Screening", Year: 2019.

[4] https://www.kaggle.com/ashrafkhan94/mammography-breast-cancer