# Breast Cancer Detection Using KNN Classifier

N. SANTHANA LAKSHMI[1], M. CHITRA EVANGELIN CHRISTINA [2]
[1,2] *Department of ECE, Francis Xavier Engineering college*

***Abstract-Diseases are analyzed by different digital image processing techniques. Early detection of breast cancer can improve survival rates to a great extent. Inter-observer and intra-observer errors occur frequently in analysis of medical images, given the high variability between interpretations of different radiologists. Breast cancer detection involves the steps which includes breast image preprocessing, tumor detection, feature extraction, training data generation, and classifier training. In the breast image preprocessing, denoising and enhancing contrast processes on the original mammogram have been utilized to increase the contrast between the masses and the surrounding tissues. The tumor detection is then performed to localize the tumor ROI. After that, features including morphological features, texture features and density features are extracted from the detected ROI. During the training process, the KNN (K Nearest Neighbour) classifier have been trained with every image from the breast image dataset using their extracted features and corresponding labels. GMM(Gaussian Mixture Model) segmentation used which enhances the contrast of the image to improve its visual quality.The final output in this project finally describes the presence of breast cancer for the given input data.***

***Indexed Terms -- Breast cancer, GMM (Gaussian Mixture Model), KNN (K Nearest Neighbor)***

## I. INTRODUCTION

Breast Cancer is one of the leading cause of death for women from the age group of 15 to 54 years. [1]. Mammography is the standard tool for the early diagnosis of breast cancer. There are various steps involved in the detection of breast cancer which includes the breast image preprocessing followed by the feature extraction, generation of training date, classifier training and tumor detection [2].

This paper is organized as follows. Section II gives a brief overview of related works. Sections III elaborates about the proposed work. Section IV discusses the results. Finally, this paper is concluded with a summary.

## II. PROBLEM IDENTIFICATION

A geometrical rule-based algorithm for the segmentation of the pectoral muscle which supports various types of muscles such as convex, concave or their combination[3].Segmentation of breast regions in mammogram based on density image segmentation means separating the image into similar constituent parts, including identifying and partitioning regions of interest[4].

Here in this method the pectoral muscle has been separated using the geometrical rule based segmentation technique [5].The existing method involves only the segmentation of various types of pectoral muscle type for the given input images of breasts.

## 3. PROBLEM SOLUTION

### 3.1 INTRODUCTION ABOUT THE PROPOSED SYSTEM

The technique used in this proposed system is important as obtaining the seriousness of the breast cancer in defining the type of cancer whether the breast is of malignant, benignor it is of normal type [6].

### 3.2 BLOCK DIAGRAM

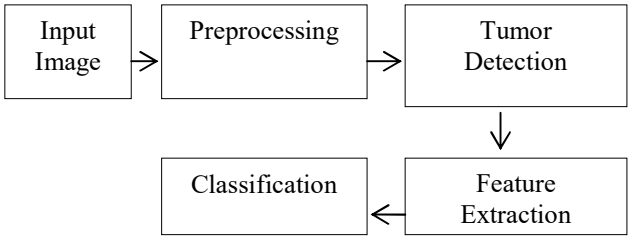


Figure 3.1 Block diagram

### 3.3 PREPROCESSING

In the breast image preprocessing, denoising and enhancing contrast processes on the original mammogram

have been utilized to increase the contrast between the masses and the surrounding tissues.

3.4 TUMOR DETECTION

The tumor detection is then performed to localize the tumor ROI.

3.5 F EATURE EXTRACTION

Features including morphological features, texture features and density features are extracted from the detected ROI.

3.6 C LASSIFICATION

During the training process, the KNN (K Nearest Neighbour) classifier has been trained with every image from the breast image dataset using their extracted features and corresponding labels [7].

GMM refers to the Gaussian mixture model from the colored image is obtained as the result to which several Gaussian random variables contribute [8]. GMM is used to model the original image and to perform the segmentation [9].Three types of severity of the breast cancer has been obtained depending on the given images of benign, malignant or normal type of cancer [10].

## III. RESULT AND DISCUSSION

Thus the breast cancer has been detected from the input received from various sources. First the input image is given which is then followed by the adaptive filtering.GMM segmentation is carried out after which whether the breast is of cancerous or not is detected.KNN classifier mainly classifies whether the cells present in the breast region are normal, malignant or benign type. The portion of the cancerous cells in the breast region is being detected from the input given. Thus the output is obtained which displays the message whether the cells are normal or the cancerous cells.

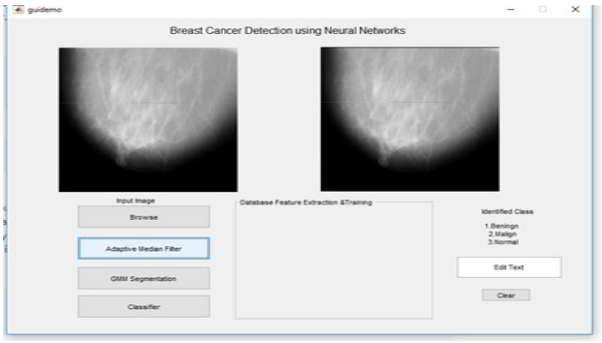


Figure 4.1 Input and Filtering

Figure 4.1 represents the input image which is then filtered using the adaptive median filter which performs spatial processing and to find which pixels in an image has been affected by the impulse noise.It compares each pixel in the image to its surrounding neighbor pixels.

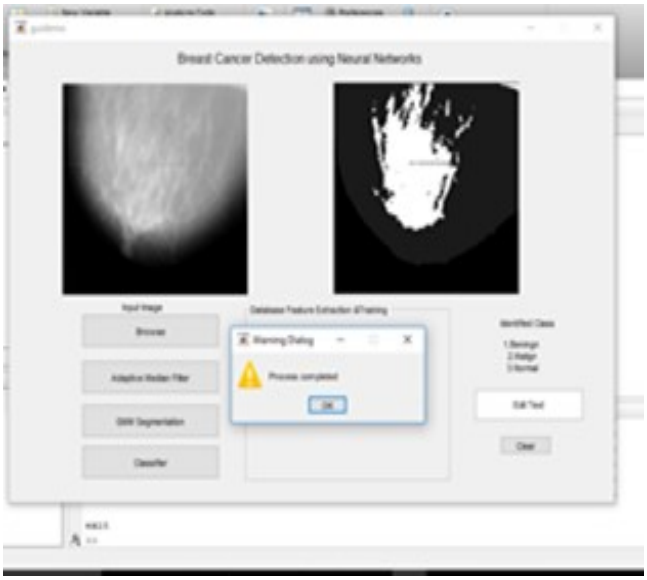


Figure 4.2 GMM Segmentation

Figure 4.2 represents the GMM which is a general estimation principle. The segmentation of color image is an important research field of image processing and pattern recognition. A color image could be considered as the result from Gaussian mixture model to which several Gaussian random variables contribute.

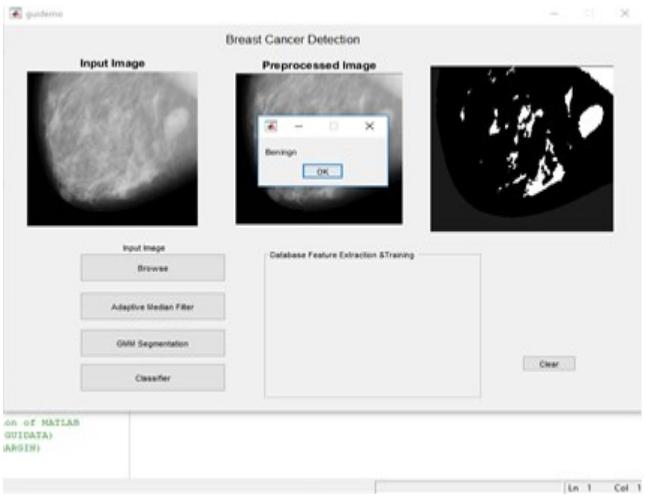


Figure4.3 Output1-Benign

Figure 4.3 represents the benign type of tumor which does not invade nearby tissue or spread to other parts of the body.


Figure 4.4 Output 2-Malignant

Figure 4.4 represents the malignant type of tumor which can invade neighboring tissue and organs through blood or lymph system.

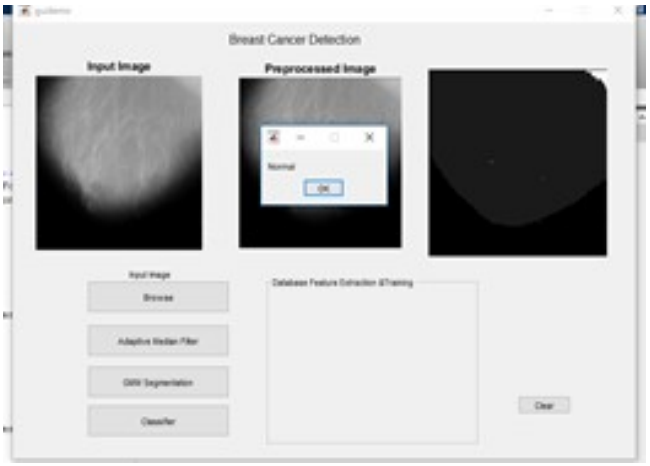


Figure4.5 Output 3-Normal

Figure 4.5 represents the normal type which is non cancerous and not treatable to the body.

## IV. CONCLUSION AND FUTURE WORK

A framework which helps in the identification of breast cancer is proposed with the advantages of finding the classification of breast cancer types which includes malignant, benign and normal type. Thus it helps the patients to know about the severity of thecancer type and continue with the medical proceedings.

The proposed system is versatile enough to be applied to extensive varieties of breast types. Thus depending on the input images of the breast the process of filtering, segmentation and classification of the cancer type has been identified. The accuracy obtained by our method is almost >90%.

Future work may include the reduction of time consumption during the entire process and to use various other algorithms with very effective accuracy and detection at an early stage of cancer to carry out the medical examinations which could save many lives.

## REFERENCE

[1] Seyyid Ahmed Medjahed, Tamazouzt Ait Saadi, Abdelkader Benyettou “Breast Cancer Diagnosis by using k-Nearest Neighbor with Different Distances and Classification Rules” International

Journal of Computer Applications (0975 - 8887) Volume 62 - No. 1, January 2013.

[2] Roar Johansen MD ,Line R. Jensen MS ,Jana Rydland MD ,Pål E. Goa PhD ,Kjell A. Kvistad MD, PhD ,Tone F. Bathen PhD ,David E. Axelson PhD ,Steinar Lundgren MD, PhD ,Ingrid S. Gribbestad PhD**"**Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI"

[3] Ville Hautamaki, ¨ IsmoKarkk ¨ ainen ¨ and PasiFranti "Outlier Detection Using k-Nearest Neighbour Graph"

[4] Moh'd Rasoul Al-hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahnah, AlBalqa, Salt, Jordan "Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm" 2016 9th International Conference on Developments in eSystems Engineering.

[5] J. S. Snchez, R. A. Mollineda, and J. M. Sotoca. "An analysis of how training data complexity affects the nearest neighbor classifiers" Pattern Analysis and Applications, 10(3), 2007.

[6] M. Raniszewski. Sequential reduction algorithm for nearest neighbor rule.Computer Vision and Graphics, 6375, 2010.

[7] A. Mert, N. Kilic, and A. Akan. Breast cancer classification by using support vector machines with reduced dimension. ELMAR Proceedings, 2011.

[8] M. Martn-Merino and J. De Las Rivas.Improving k-nn for human cancer classification using the gene expression profiles.Computer Science Advaces in Intelligent Data Analysis VIII, 5772/2009, 2009.

[9] L. Li and C. Weinberg. Gene selection and sample classification using a genetic algorithm and k - nearest neighbor method. A Practical Approach to Microarray Data Analysis, 2003.

[10] D. Coomans and D.L. Massart. Alternative k-nearest neighbour rules in supervised pattern recognition. AnalyticaChimicaActa, 136, 1982.

[11] M. F. Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems with Applications, 2(36), 2009.

[12] D. Bremner, E. Demaine, J. Erickson, J. Iacono, S. Langerman, P. M., and Godfried. Output-sensitive algorithms for computing nearest-neighbour decision boundaries. Discrete and Computational Geometry, 33(4), 2005.

[13] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines.Machine Learning, 46(1-3), 2002.

[14] O. L. Mangasarian and W. H. Wolberg.Cancer diagnosis via linear programming. SIAM News, 5(23), Sep 1990.

[15] R. Mallika and V. Saravanan.Ansvm based classification method for cancer data using minimum microarray gene expressions. World Academy of Science, Engineering and Technology, 62, 2010.