Journal on Electronic and Automation Engineering

Vol: 2(1), March 2023

REST Publisher, ISSN: 2583-6951 (Online)

Website: https://restpublisher.com/journals/jeae/

DOI: https://doi.org/10.46632/jeae/2/1/1

Mammogram Image to Detect Breast Cancer Using K-Means Clustering Algorithm

B. Gopinathan, M. Naveena, *M. Shreyas, D. Charan Rohith

Adhiyamaan college of Engineering autonomous, Hosur, Tamil Nadu, India.

*Corresponding author Email: shreyasgowda7867@gmail.com

Abstract. *Breast cancer is the uncontrolled proliferation of a group of cells in the breast and is the second leading cause of death for women in the world. The disease can be cured if detected in the early stages. A lot of research has been done to correctly detect the tumor, but a 100% accurate method has not been found. Research on breast cancer detection using digital image processing is not new, but many new approaches are being considered in this field to accurately predict the tumor area. The current approach consists of detecting the tumor area visually and also finding out in which area the tumor is most concentrated. CC and MLO dualview mammographic screening images are widely used in the diagnostic process. This project presents a method for detecting a tumor area and classifying a normal and oncological patient. Preprocessing operations are performed on the input mammographic image and unwanted parts are removed from the image. Tumor regions are segmented from the image using a morphological operation and are highlighted on the original mammographic image. If the image on the mammogram is normal, it means that the patient is healthy. This work mainly focuses on finding the best algorithms for detecting tumors present in the breast. A number of algorithms were used in the proposed work, but the most suitable for cancer detection is the combination of K-Means clustering algorithm. K-Means classification accuracy is 95% accurate output will be predicted. Keywords: Image Processing, Breast Cancer, K-Means clustering, Dilation, Closing, Edge Detection, Mammography screening images.*

1. INTRODUCTION

Mammography is considered the most important method of breast cancer screening. It can be used to detect disease at an early stage when recovery is possible. The aim of the Computer Aided Diagnosis (CAD) system is to read mammographic images, locate suspected areas of abnormalities and analyze their characteristics. The performance and reliability of CAD depends on the accurate segmentation of lesions and the selection of an appropriate feature. Global lesion segmentation is the most challenging task due to artifacts and healthy tissues present in mammograms. Various algorithms have been developed in the literature for the early detection of breast cancer on mammograms. We have proposed an efficient segmentation algorithm for breast cancer detection in mammograms. This work proposes a shape-based approach for breast cancer detection using mammograms. Since screening mammography is currently the main test for the early detection of breast cancer, a large number of mammograms must be examined by a limited number of radiologists, leading to misdiagnoses caused by human error and eye fatigue. Currently, several image processing methods are proposed for the detection of tumors on mammograms.

2. LITERATURE SURVEY

Techniques used for breast cancer screening. The image produced by mammography is called a mammogram, which consists of the background, breast area, fatty tissue, breast mass, and high-intensity micro calcifications have been diagnosed. [1] Micro calcifications (MCs) are calcium deposits that appear as tiny bright spots on a mammogram. Because MCs and masses appear close to the context on a mammogram, identification and classification is difficult. Because image processing methods play a significant role in the earlier diagnosis of MC. Researchers have developed many strategies to determine the exact location of MCSs and masses.[2] Then, sub regions were obtained by dividing the initial image and bicubic interpolation was used to obtain the local history intensity level. Finally, a difference image is obtained by subtracting the interpolated image from the original image, and an area estimation technique is used to cluster the micro calcifications.[3] One of the the leading cause of death for women worldwide is breast cancer. It caused more deaths than any other disease such as malaria or tuberculosis. The World Health Organization (WHO) cancer research agencies (ie, the International Agency for Research on Cancer (IARC) and the American Cancer Society) report that 17.1 million new cancer cases were registered worldwide in 2018.[4]In developed and developing countries with countries Citizens are changing their lifestyle from traditional to modern, which increases the incidence of breast cancer in women, especially in the age group of 35-55 years. The incidence of breast cancers can be monitored by detecting breast cancers in their early stages.

3. EXISTING SYSTEM

The main difference is that in Fuzzy c-mean (FCM) each point has a weight associated with a specific cluster. Thus, a point is not in a cluster so much as it has a weak or strong association with the cluster, which is determined by the inverse distance to the cluster center. Mammograms aren't perfect. They lack the accuracy of cancer detection. And sometimes a woman will need additional tests to determine whether or not something found on a mammogram is cancer. There is also a small chance of a cancer diagnosis that would never cause any problems if not found during screening.

4. PROPOSED SYSTEM

The proposed methodology includes several steps including image processing techniques. The first step is to acquire an image from a dataset collected in the Digital Database for Screening Mammography (DDSM), where regular and irregular mammograms are collected. The optical mammograms are then pre-processed with Gaussian filters to reduce noise. The images are further processed using the proposed multi-point K-Means method to extract target breast MCs. Various algorithms were used in the proposed work, but the most suitable for cancer detection is the combination of K Means, Closing and Dilation and the Canny Edge Detection algorithm.

5. MODULES

Login Module: In login module user will first run a project, once program is been run then an welcome page will be appear by clicking next in welcome page an login page will be appear in which user should enter username and password. If user name and password is correct then next page will be appeared if username and password are wrong it will show as username or passwords is wrong.

Image processing module: Image processing is used for transform an image into digital form and perform certain operations on it in order to obtain specific models or to extract useful information from the image that is to extract a pure form of image to get an accurate form of image.

Segmentation module: Input image will be resized for extracting exact image of output, then Edge detection processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in areas such as image processing and to extract pure image. ***Vgg-16:*** VGG architecture is the base of ground-breaking recognition models. Developed as a neural network, the VGG Net also surpasses baseline on many tasks and datasets beyond ImageNet. We will be training with 15 to 20 epochs, for accuracy of breast cancer. Through this epoch training we will get accuracy data that will help doctors to identify it easily.

***K-Mean clustering*:** It is an iterative process of assigning each data point to the groups and slowly data points get clustered based on similar features. The objective is to minimize the sum of distances between the data points and the cluster centroid, to identify the correct group each data point should belong to.

Output Module: In this module it will detect that breast cancer is been identified or not at very earlier stage.

6. ARCHITECTURE DESIGN

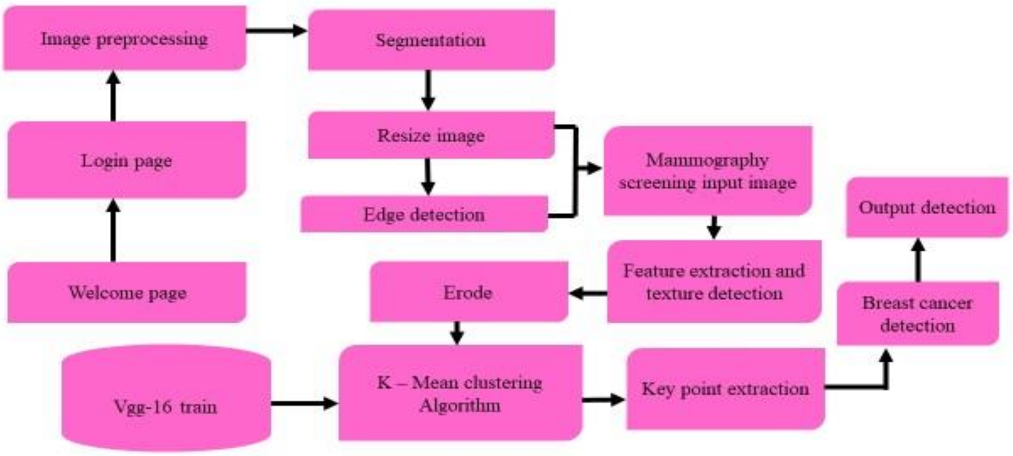

FIGURE 1. Architecture Design

7. SYSTEM FUNCTIONS

K- Mean Clustering Algorithm The K-means clustering algorithm calculates the centroid and iterates until the optimal centroid is found. It is assumed that it is known how many clusters exist. It is also known as K means clustering algorithm. The number of clusters algorithm found from the data by the method is indicated by the letter 'K' in Kmeans. in this method, data points are assigned to clusters in such a way that the sum of the squares of the distances between the data points and the centroid is as small as possible. It is important to note that reduced diversity within clusters results in more identical data points in the same cluster. How the K-Means algorithm works. The following stages will be help and used understand how the K-Means clustering technique works-

Step 1: First, we need to provide the number of clusters, K, to be generated by this algorithm.

Step 2: Next, randomly select K data points and assign each to a cluster. In short, the data based on the number of data points store in database.

Step 3: Repeat the steps below until we find the ideal centroid, which is the assignment of data points to clusters that do not change.

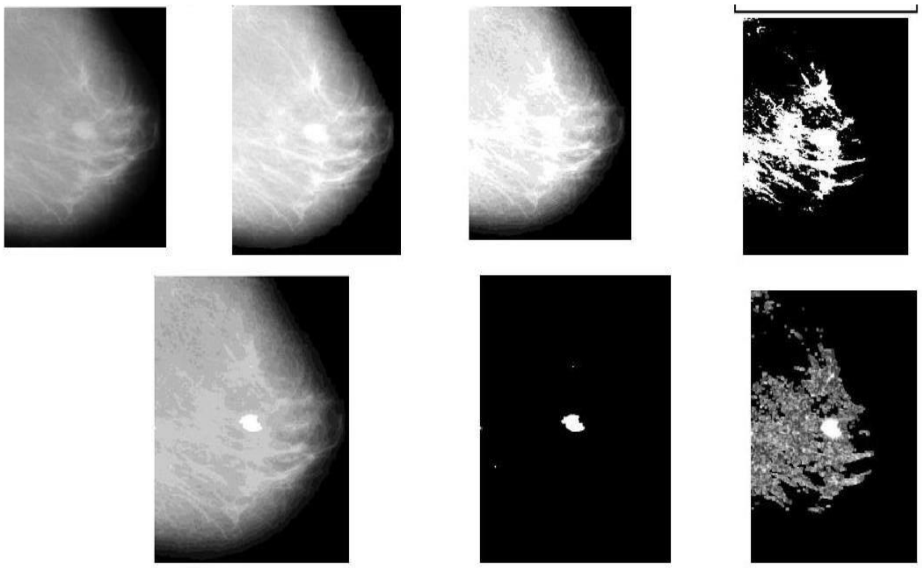

FIGURE 2. K-means clustering algorithm

VGG-16

TRAIN VGG16 is the VGG model, also called VGG Net. It is a convolution neural network (CNN) model supporting 16 layers. VGG16 is therefore a relatively large network with a total of 138 million parameters - huge even by today's standards. However, its main attraction is the simplicity of the VGGNet16 architecture. The VGG Net architecture incorporates the most important features of a convolutional neural network.

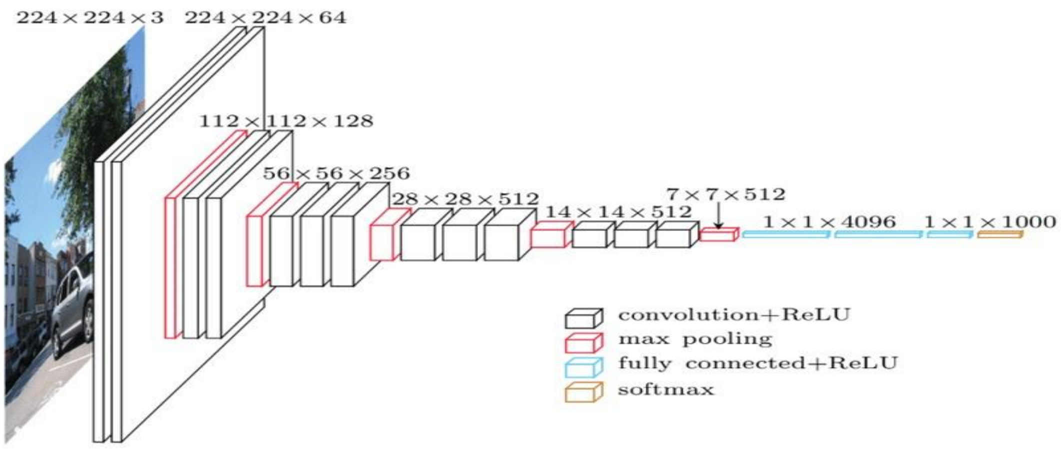

FIGURE 3. VGG-16

IMAGE PROCESSING

Image processing is a method for performing some operations on an image to obtain an improved image or to extract some useful information from it. It is a type of signal processing in which the input is an image and the output can be an image or characteristics/properties associated with that image. Nowadays, image processing is one of the rapidly developing technologies. It forms a major research area within engineering and computer science disciplines. Image import using image acquisition tools;

Image analysis and manipulation;

Output, which can result in an altered image or a report based on image analysis.

They 2 types of methods used in image processing, namely analog with digital image processing. Analog image processing can be used for hard copies such as prints and photographs. Image analysts use different bases of interpretation when using these visual techniques. Digital image processing techniques help in manipulating digital images using computers.

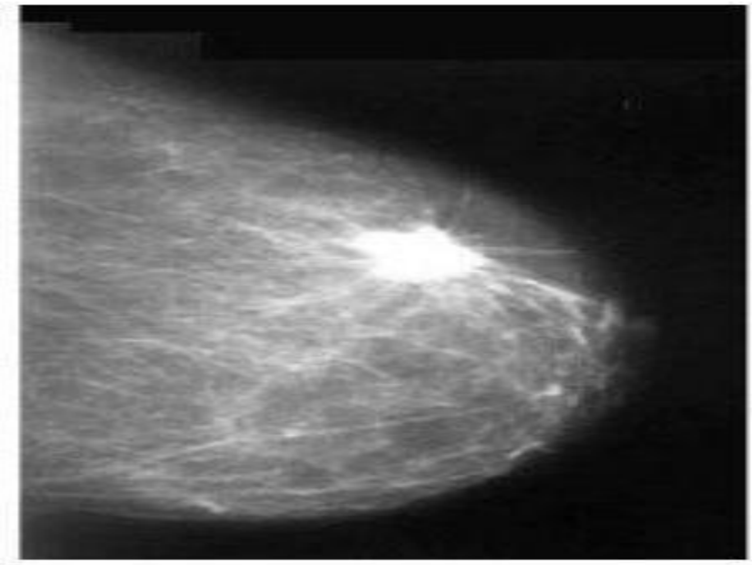

FIGURE 4. Image processing

SEGMENTATION

Edge detection is the process of locating edges in an image, which is a very important step in understanding image properties. Edges are assumed to consist of meaningful elements and contain significant information. It significantly reduces the size of the image to be processed and filters out information that may be considered less relevant, while preserving and focusing only on the important structural features of the image for the business problem. Edge-based segmentation algorithms work to detect edges in an image based on various discontinuities in gray level, color, texture, brightness, saturation, contrast, etc. To further improve the results, additional processing steps must follow to merge all edges into an edge. Chains that better match the edges in the image.

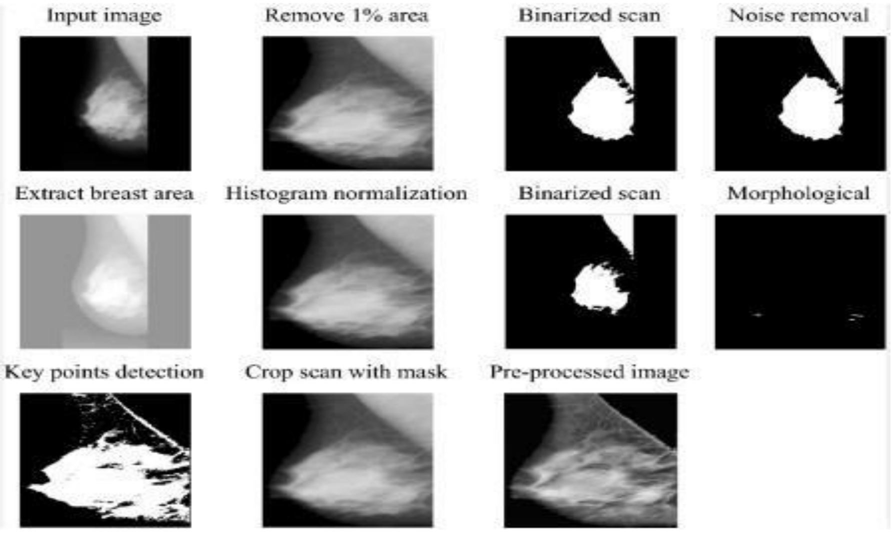

FIGURE 5. Segmentation processing

EDGE DETECTION

Edge detection is an image processing technique used to identify points in a digital image with discontinuities, simply put, sharp changes in image brightness. These points where the brightness of the image changes sharply are called the edges (or boundaries) of the image. It is one of the most important steps in image processing, image pattern recognition and computer vision. When we process digital images with very high resolution, convolution techniques come to our aid. Let's understand the convolution operation (shown in the image below by *) with an example - For this example we are using a 3*3 Prewitt filter as shown in the image above. We would continue the above process to get the processed image after edge detection. But in the real world we are dealing with very high resolution images for AI applications. That's why we decided on an algorithm to perform convolutions and even use Deep Learning to decide the best filter values.

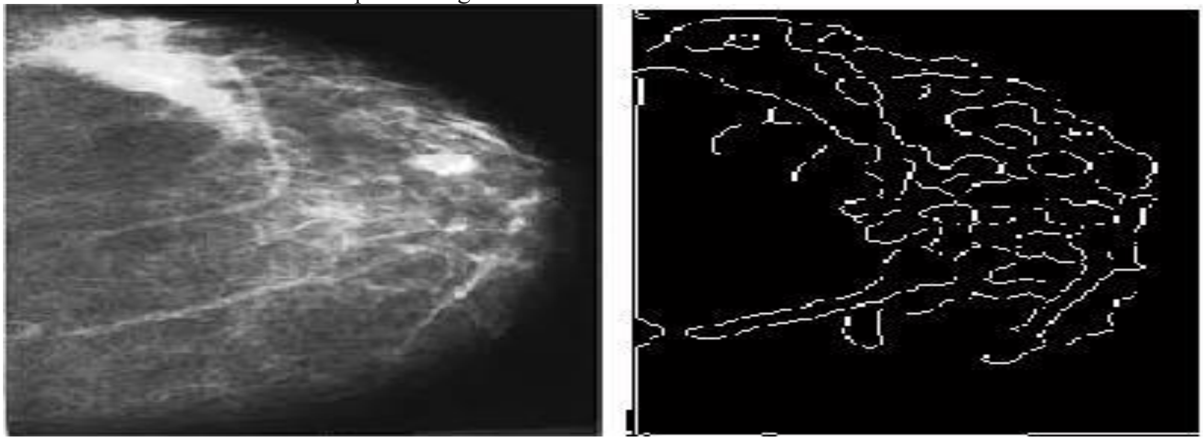

FIGURE 6. Edge detection

MAMMOGRAPHY SCREENING INPUT IMAGE

An image of the breast is known as a mammogram. The background of the image will be black and the breasts will be displayed in gray and white. Tissue that is denser, including connective tissue and glands, will appear white. This can make it difficult to detect abnormalities on a mammogram because the tumor is made up of dense tissue and also appears white. Breasts tend to shrink with age. Less dense tissue, such as fat, appears gray on a mammogram. A standard mammogram will be mostly gray, with some white areas showing healthy dense tissue. Whiter in the picture does not always mean a health problem. Everyone's breasts are different, so no two mammograms will be the same. Healthy mammograms can still vary in appearance. A doctor who reviews imaging tests such as X-rays or MRIs is called a radiologist.

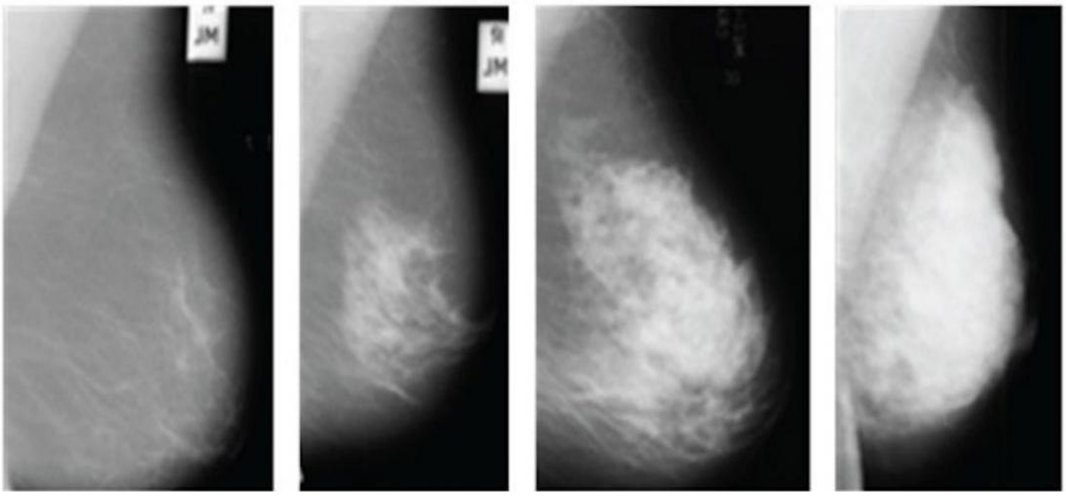

FIGURE 7. Mammography screening input image

FEATURE EXTRACTION AND TEXTURE DETECTION

Features extraction and texture detection the parts or patterns of an object in an image that help identify it. For example: A square has 4 corners and 4 edges, these can be called elements of a square and help us humans to identify that it is a square. Features include features such as corners, edges, regions of interest, ridges, etc. As shown in the image below, the yellow dots represent features detected using a technique called Harris detection.

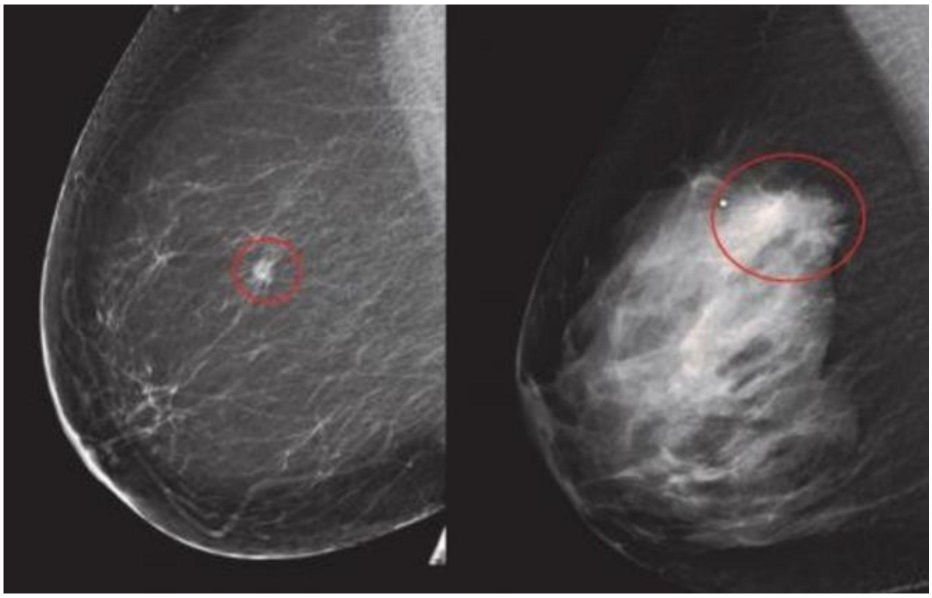

FIGURE 8. Feature extraction and texture detection

DESCRIPTOR OF KEY POINT

The local region (key point) SIFT descriptor is a 3D spatial histogram of image gradients. The gradient at each pixel is considered to be a sample of a three-dimensional vector of elementary features, formed by the location of the pixel and the orientation of the transition. Binary image descriptors encode the appearance of a patch using a compact binary string. The Hamming distance in this space is designed to track the desired degree of image similarity, which usually tries to be invariant to scene illumination and viewpoint changes. A feature descriptor is an algorithm that takes an image and outputs feature descriptors/vectors. Feature descriptors encode interesting information into a series of numbers and act as a sort of numerical "fingerprint" that can be used to distinguish one feature from another.

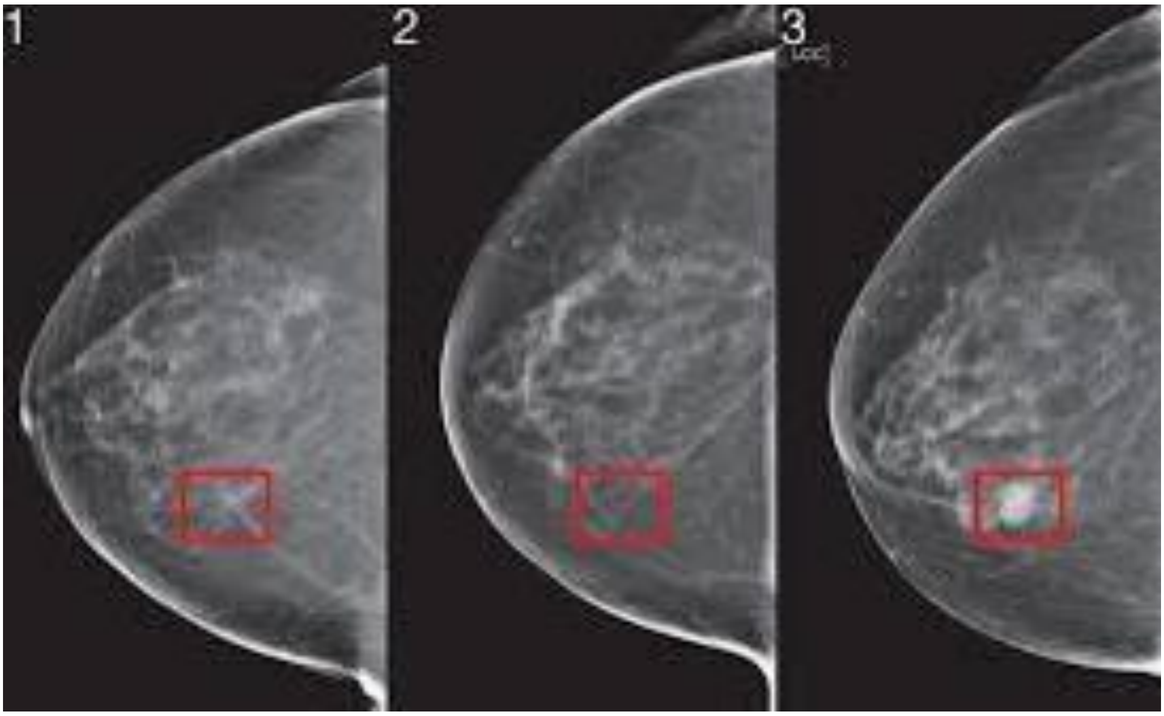

FIGURE 9. key point description

8. RESULT

The results obtained in this study are in two phases: tumor classification without augmentation and with augmentation of the original database.

Form
MAMMOGRAM IMAGE TO DETECT BREAST CANCER
USING K-MEANS CLUSTERING ALGORITHM
Next
sheknows

FIGURE 10. Welcome Page

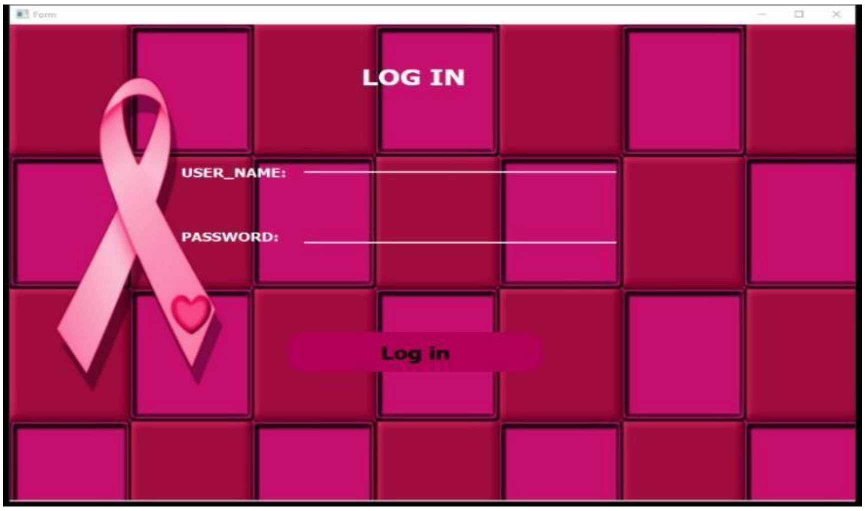

FIGURE 11. Login Page

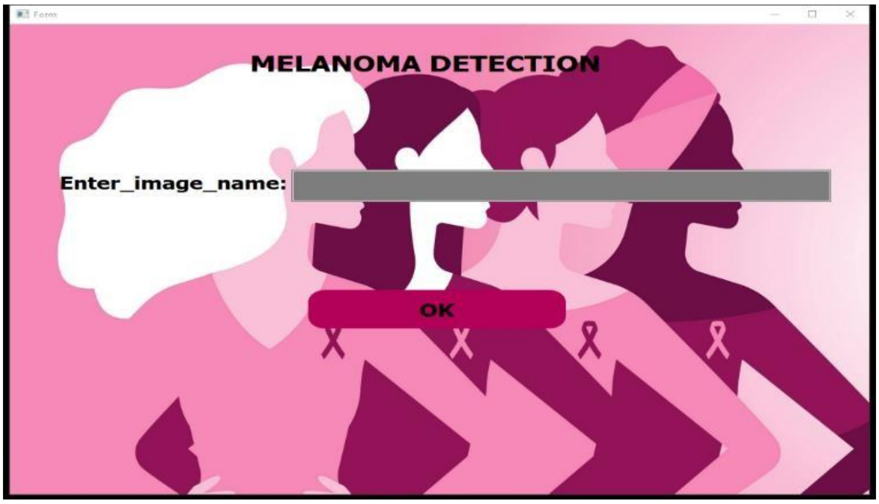

FIGURE 12. Melanoma Detection Page

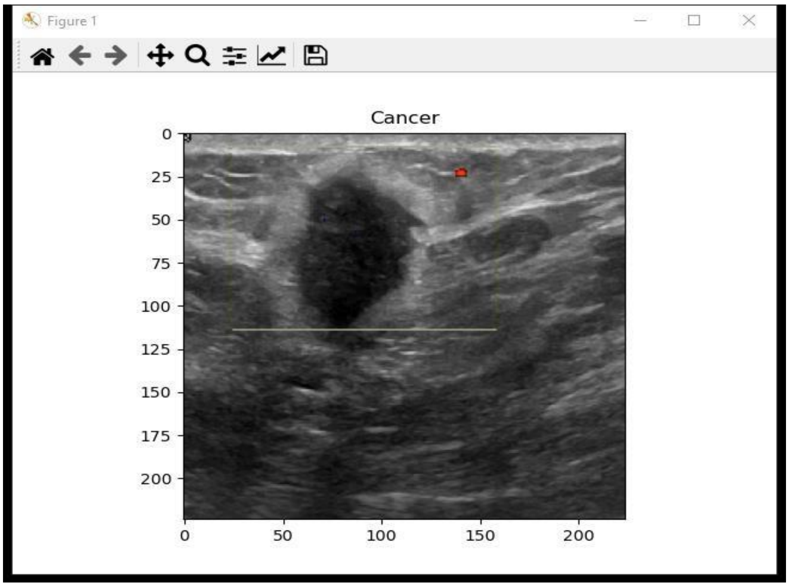

FIGURE 13. Cancer Identification

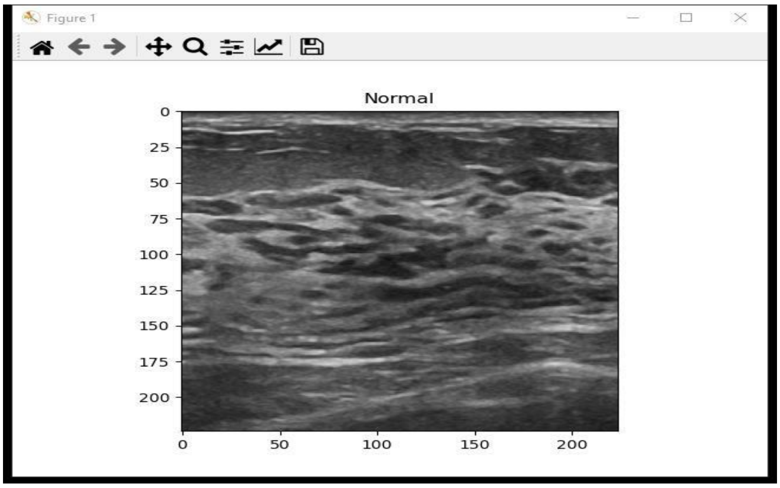

FIGURE 14. Normal Mammogram Image

9. CONCLUSION

Processing techniques, K-mean clustering algorithm for earlier detection of breast cancer. Image pre-processing, image segmentation, feature extraction, classification, Vgg-16, K-mean clustering are used for this. The resulting classification process is performed using the ensemble learning method and determines whether the tumour is normal, benign or malignant with the output image of the segmented part of the breast. The project thus helps in detecting a cancerous tumour before it spreads to other parts of the body and increases the chances of a successful diagnosis.

REFERENCES

[1]. Detection of breast cancer tumor using mathematical morphology and wavelet analysis", Mohiy Hadhoud, ohamed Amin, Walid Dabbour in GVIP 05 conference, 19- 21 December 2005, CICC, Cairo, Egypt.

[2]. Characterization of micro-calcifications and mass on female breast using processing in full field digital mammography (FFDM)", Kanaga, K.C.1, Anandan, S.2, Chin, M.Y.1 & Laila, S.E.1, symposium sainskesihatankebangsaanke 7, hotel legend, Kuala Lumpur, 18-20 June 2008: 183-187.

[3]. M. Bhattacharya & A. Das, "Fuzzy logic based segmentation of Micro calcification in Breast using Digital Mammograms considering Mutiresolution. [4]. M.J. Bottema, G.N.Lee and S.Lu, "Automatic image feature extraction for diagnosis and prognosis of breast cancer," Artificial intelligence techniques in breast cancer diagnosis and prognosis, Series in machine perception and artificial intelligence, Vol.39, World Scientific Publishing Co.Pte.Ltd, 2000, pp. 17-54.

[5]. T. C. Wang and N. B. Karayiannis, " Detection of microcalcifications in digital mammograms using wavelets", IEEE Transaction on Medical im aging, vol. 17, no. 4, pp. 498-509,AUGUST 1998. GVIP 05 Conference, 19-21 December 2005, CICC, Cairo, Egypt.

[6]. Aijuan Dong and Baoying Wang "Feature selection and analysis on mammogram classification, Communications, Computers and Signal Processing, 2009. Pac Rim 2009. IEEE Pacific Rim Conference on 23-26 Aug. 2020.

[7]. Thangavel, K.Mohideen, A.K. Dept. of Comp.Sci., Peri Tobias Christian cahoon, melanie a. sutton and james c. bezdek, "Breast cancer detection using image processing technique", IEEE conference , 2019.

[8]. P. Saha, J. Udupa, E. Conant et al., "Breast tissue density quantification via digitized mammograms," IEEE Trans. Med. Imaging, vol. 20, no. 8, pp. 792–803, 2021.

[9]. P. Saha, J. Udupa, E. Conant et al., "Breast tissue density quantification via digitized mammograms," IEEE Trans. Med. Imaging, vol. 20, no. 8, pp. 792–803, 2018.

[10]. X. Liu and D. Wang, "Image and Texture Segmentation Using Local Histograms", IEEE Trans. Med. Img., vol.15, pp. 30663076, 2020.

[11]. American Cancer Society, "Breast cancer facts and figures 2007-2008", Atlanta, Georgia: American Cancer Society, Inc.2021.