OKU Fen Bilimleri Enstitüsü Dergisi
5(1):258-281, 2022

OKU Journal of The Institute of Science and Technology 5(1):258-281, 2022

Osmaniye Korkut Ata Üniversitesi
Fen Bilimleri Enstitüsü
Dergisi

Osmaniye Korkut Ata University
Journal of The Institute of Science
and Technology

Meme Kanserinin K-Ortalama Kümeleme ve Otsu Eşikleme Segmentasyon Yöntemleri İle Teşhisi*

Aslı Canan KUŞCU[1], Halil EROL[2]**

[1,2]Osmaniye Korkut Ata University, Faculty of Engineering, Department of Electric and Electronics, 80000, Osmaniye

[1]https://orcid.org/0000-0001-8116-467X
[2]https://orcid.org/0000-0001-6171-0362
**Corresponding author: aslikuscu.1994@gmail.com
*This article is derived from the first author's graduate study.

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 12.09.2021
Kabul tarihi:20.12.2021
Online Yayınlanma:08.03.2022

Anahtar Kelimeler:
Mamografi
Görüntü işleme
K ortalama kümeleme
Otsu eşikleme
Performans metrikleri

ÖZ

Meme kanseri, kadınlar arasında büyük oranda artış göstermiştir. Ancak erken teşhisiyle, tedaviye olumlu cevap verilebilmektedir. Araştırmacılar, hastalığı erken ve doğru tespit edebilme adına görüntüleme yöntemlerinde çeşitli çalışmalar yapmaktadır. Bu çalışmada; TCİA görüntü veri bankasından alınan 9 kanserli görüntüde K ortalama kümeleme ve otsu eşikleme yöntemi ile tümör tespiti yapılmıştır. Radyolog tarafından işaretli referans görüntüleri ile (ground truth) ile karşılaştırarak, başarım (performans) metrikleri değerlendirilmiştir. Kümeleme işlemi için sırasıyla TPR (Doğru Pozitif Oranı) 0.89, FPR (Yanlış Pozitif Oranı) 0.14, benzerlik 0.67, doğruluk 0.87, duyarlılık 0.89, hassasiyet 0.86, özgüllük 0.87, F puanı 0.87 bulunmuştur. Otsu için TPR (Doğru Pozitif Oranı) 0.84, FPR (Yanlış Pozitif Oranı) 0.12, benzerlik 0.73, doğruluk 0.84, duyarlılık 0.84, hassasiyet 0.86, özgüllük 0.87, F puanı 0.84 olarak hesaplanmıştır. Bu çalışmada, daha az veri kümesi ile daha kısa sürede, görüntü işleme yöntemlerini kullanarak, piksel tabanlı segmentasyon ile tümör sınırlarının daha doğru belirlenmesi, insana duyulan ihtiyacın azalması ve sağlık alanında sahada görüntülemede kullanılan tıbbi cihazların bilgisayar destekliyazılımlarla geliştirilmesi, mamografik tarama sistemlerinin doğru ve hızlı bir şekilde yapılabilmesi amaçlanmıştır. Sonuç olarak, her iki yöntem de başarılı, sahada kullanılabilir ve birbirine yakın başarım değerleri bulunmuştur.

Diagnosis of Breast Cancer by K-Mean Clustering and Otsu Thresholding Segmentation Methods

Research Article/Reviews

Article History:
Received: 12.09.2021
Accepted: 20.12.2021
Published online:08.03.2022

Keywords:
Mammography
Image processing
K-mean clustering
Otsu threshold
Performance metrics

ABSTRACT

Breast cancer has increased decidedly among women. But with early diagnosis, a positive response to treatment can be given. Researchers are conducting various studies in imaging methods to detect the disease early and accurately. In this study, 9 cancerous images taken from the TCİA image data bank were detected by K-mean clustering and the Otsu threshold method. Performance metrics were evaluated by comparing them with marked reference images (ground truth) by the radiologist. For the clustering process, TPR (True Positive Rate) 0.89, FPR (False Positive Rate) 0.14, similarity 0.67, accuracy 0.87, sensitivity 0.89, exact hit ratio 0.86, specificity 0.87, F Score 0.87 were found, respectively. For Otsu, TPR (True Positive Rate) 0.84, FPR (False Positive Rate) 0.12, similarity 0.73, accuracy 0.84, sensitivity 0.84, exact hit 0.86, specificity 0.87, F Score 0.84 were calculated. The aim of this study is to determine the tumor boundaries more accurately and to use them in imaging devices in the field of health with pixel-based segmentation. As a result, both methods were successful can be used in the field and close to each other.

To Cite: Kuşcu AC., Erol H. Diagnosis of Breast Cancer by K-Mean Clustering and Otsu Thresholding Segmentation Methods.Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2022; 5(1):258-281.

Introduction

Breast cancer is the most common cancer in women worldwide. According to the American Cancer Society (ACS) report, approximately 2.6 million women have been diagnosed with invasive breast cancer, and approximately 40.000 women have died in 2020 (Cancer Facts and Figures, 2020). 13% of cancers in Canada are breast cancer. An increase of 25% was observed in women and 1% in men (Canadian Cancer Statistics Advisory Committee, 2018; Toronto, 2018). In Chinese women, it is twice as high as the global rate (Sun et al., 2018).

The causes of breast cancer are often explained by uncontrolled malignant growth, although in some cases they remain unclear. Growth usually begins in the cells in the breast tissues, which can be managed or controlled at a certain level, without causing any problems in a healthy and normal state. In the case of breast cancer, it is impossible to control the regeneration and growth of cells. As soon as the repair does not take place, the final mutations proceed with the formation of a cancer tumor. After a cancer tumor has formed, the tumor increases in size and the patient begins to show other complications. The stage of tumor evolution varies from person to person. Other health parameters play an important role for each individual. Also, family history is considered an effective possibility for the genetic transmission of breast cancer. As well as late diagnosis, obesity, early or late menopause, have never given birth, fibrocystic diseases, the presence of abnormal cells, and the possibility of receiving hormone therapy are important factors in the formation of breast cancer.

These lesions typically have a size in diameter due to their very small sizes, microcalcifications can be quite difficult to detect. In general, benign calcifications come in uniform sizes with round or large elliptical shapes, but non-uniform, small, polymorphic, and spreading calcifications with heterogeneous volume and morphology have a higher chance of becoming malignant (Tan et al., 2020). Some anatomical structures, such as fibrous strands, breast borders, or hypertrophic lobules, are also similar to microcalcifications in the mammographic image. Their presence in the chest area can vary, and they can usually be distinguished by their bright color. On the other hand, these lesions can be different in size and shape, and their distribution may vary from patient to patient. Sometimes, because of the difference in density between suspicious spots and the area surrounding these lesions decays, low contrast in their color may be observed. Also, the proximity to the surrounding tissues can cause difficulty in their detection. In dense tissues, suspicious areas cannot be detected due to tissue overlap (Sankar and Thomas, 2010; Rao and Sannapareddy, 2021).

Masses appear as dense zones of different characteristics and volumes. They can be lobular, circular, oval, or non-uniform/speculated. They are well defined and distinctly delimited. Previous studies have shown that, depending on the morphology, masses usually have several chances of malignancy. For example, speculated and ill-defined boundaries are more likely to be malignant (Akay, 2006; Azhar, 2021). The presence of elliptical or circular masses is a sign of benign. Studies show that the large variability of mass appearance is a challenge because it is an obstacle to accurate mammography analysis (Mini and Thomas, 2003).

The normal configuration of the parenchyma is irregular, diffuse, without a visible center or mass, architectural defects are pronounced. It is very difficult to find them because they are very variable (Naranjo and Reymbaut, 2021). In addition to benign vascular calcifications, the classic "popcorn" for involutional fibro adenomas shows two well-defined masses containing calcifications (Gunderman, 2006; Schönenberger and Hejduk, 2021)

Detection of breast cancer in the very early stages is a very important advantage. Early detection with proper medical treatment and assistance can save tens of thousands of women's lives every year. Currently, there is no effective way to prevent breast cancer. However, successful early detection can play an important role in improving treatment options and patient survival before cancer spreads to other parts of the body (Birdwell et al., 2001; Manraj et al., 2021)

There are different imaging methods in the diagnosis of cancer; these are mammography, thermography, ultrasound imaging, and histopathology. Mammography is a traditional technique for diagnosing breast cancer. Image processing, on the other hand, is transforming the image into digital form for various purposes. Different techniques can be used in image processing. K-means clustering and the Otsu thresholding technique were used in this study. The previous studies on this topic are shown in Table 1.

Table 1. A review of the relevant literature on the subject has been conducted and is given in the tabular form.

The researcher	Year	The method he uses	Performance measurements
Podgornova and Sadykov	2019	Segmentation of the Basin, Mean Drift, and k Mean Clustering	In this study, 57.2% of error detection results were found.
Kaur and Singh	2019	K-Mean Clustering for Accelerated Robust Features (SURF) Selection	The average accuracy rates of the three classes using the proposed method, namely, normal, benign, and malignant cancers, were found to be 95%, 94%, and 88%, respectively.
Sadeghi et al.	2018	Histogram Diagram for Calculating the Initial Threshold	A sensitivity of 96.7% and a false positive result of 0.79 were found.
Andrik	2017	Edge-Free Active Contour Models for Investigating the Real Boundary	An accuracy rate of 82.33% has been achieved.
Ciecholewski	2017	A Computer-Aided Method for Segmenting Micro-Calcifications on Mammograms Using Morphological Transformations	A similarity index of 80.5%, an overlap ratio of 75.7%, an overlap value of 70.8%, and a difference of 19.8% were found.

Materials and Method

This study aims to design a fully automated, computer-aided diagnosis (CAD) algorithm for manually segmented breast cancer images.

Data Base and Used Programs

The Cancer Imaging Archive (TCIA) is a service that de-identifies and hosts a large publicly available archive of medical images of cancer. TCIA is funded by the Cancer Imaging Program (CIP), a part of the United States National Cancer Institute (NCI), and is managed by the Frederick National Laboratory for Cancer Research (FNLCR).

The imaging data are organized as “collections” defined by a common disease, image modality or type (CT, MRI, etc) or research focus. Dicom is the primary file format used by TCIA for radiology imaging. An emphasis is made to provide supporting data related to the images such as patient outcomes, treatment details, and expert analyses.

Matlab is a programming and numeric computing platform used by millions of engineers and scientists to analyze data, develop algorithms, and create models. Matlab combines a desktop environment tuned for iterative analysis and design processes with a programming language that expresses matrix and array mathematics directly. It includes the Live Editor for creating scripts that combine code, output, and formatted text in an executable notebook. In this study, the editor and workspace pages of the Matlab program were mainly used.

In this study; to distinguish the diseased and healthy breast tissue images from each other quantitatively, 9 cancerous breast images with the least noise were taken from the Tcia (The Cancer İmaging Archive) database. When triple clustering and Otsu were applied to 9 images, the number of analyzed images was 36. 8-bit gray-level images in different pixels were obtained by opening the data in Dicom format in the Matlab environment.

The explanation of the methods used in the study and the performance metrics of the related methods are given below. The method consists of 4 main stages:

In Stage 1, the unnecessary parts were cropped, the contrast was ensured to be in a certain range, the intensity was normalized and the noise generating regions were cleared. Thus, the image was ready for use (Image pre-processing).

In Stage 2, breast tumors were segmented. The specifications used for segmentation could not be specified with strict boundaries. Because the method to be chosen depended largely on the tumor type, class, and subdivision. This difference was also reflected in the image. Image density wasused in the study. Because different tissues had different gray levels.

In the breast mammography image, the hollow structures were black, the filled structures were white, and the gray parts were the pectoral muscles and soft tissues. Since the breast tissue had a hollow structure in general, the parts to be segmented manually were the parts that were displayed as white tumors. Images were fed into Matlab by using the dicomread function of the compiler's image processing library (Khan and Ahmad, 2004; Kaur, 2017). Tumor edge detection was performed with a canny filter. This process was performed for each cell of the breast image and two mask values were obtained. The mask values obtained were passed through the morphology with the imopen and bwareaopen functions, and manually segmented images were obtained. Manually segmented images (ground truth) were binary images because they consisted of mask images.

In stage 3, clustering and Otsu threshold segmentation algorithms were applied. K was determined as K=3, 4, and 5 by mean clustering algorithms. A clustering-independent Otsu thresholdwas then applied to the same images. Groundtruth reference images were compared (Table 3). Normally segmentation is used to analyze regions of different densities, but here it is used to determine tumor presence and to extract tumor location.

In stage 4, the performance metrics of the tumor detection algorithms were calculated on the compared images (Table 13-14). TP, TN, FP, FN, FPR, TPR, similarity, accuracy, sensitivity, precision, specificity, F score were the performance metrics used in our study and measure the success of background and tumor differentiation. TP, TN, FP, FN were in pixels. The algorithm we use has been analyzed whether it is reliable and usable according to these metrics. Now let's tell about the details of these stages.

K -Means Clustering

K-means clustering is a clustering technique that can group large amounts of data with a relatively fast and more efficient processing time (Das, 2008). Similarities or closeness between data is expected. Thus, it can be divided into multiple clusters where a high degree of similarity between cluster points can be achieved (Shokrgozar and Sobhani, 2016). K tools are very simple, easy to measure the distance, and based on iteration termination requirements. K-means clustering is a local optimization, so it is sensitive to the first data point collection from the midpoint of each cluster (Khan and Ahmad, 2004). The purpose of these adjustments is to achieve the best accuracy and fastest convergence. Also, choosing the starting position from the midpoint of a cluster places the K-means clustering algorithm in the optimal position (Kaur, 2017). The K-means clustering method randomly chooses the style from the center to k as the starting point (Yang and Sinaga, 2019). The iteration number with the cluster centroid is affected by the first randomly set cluster centroid (Lin and Ji, 2020). Therefore, it can be fixed to achieve higher performance by identifying the cluster centroid at high baseline data points (Aswathy and Jagannath, 2020). Since K-means clustering is usually applied, the data point {x_1,{x_2, ..., x_n} is grouped into k clusters. It has high-performance computation and can handle multi-dimensional vectors (Çiklaçandir et al., 2019; Bottou and Lin, 2007). Thus, it reduces the extent of distortion, increasing accuracy. X_i (j) is a chosen measure of the distance between the data point and the cluster center, c_j is a measure of the distance between the n data points and their respective cluster centers (Tang, 2019). This correlation is shown in Eq.1.

$$x_i(j) - c_i(j)^2 \tag{1}$$

1. K points are placed in the area represented by the clustered objects and these points represent the first centroids.

2. Each object group is assigned to the category with the closest center.

3. The locations of the k centroids to which all objects will be allocated are recalculated.

4. Steps 2 and 3 are repeated until the centroids move.

This causes the objects to be divided into groups from which the metric to be minimized can be calculated (Katz and Barness, 2015).

The K-means clustering algorithm is also versatile. There are two known tool clustering algorithms: the first requires a predefined cluster starting number k centroid as a prerequisite parameter for clustering, but generally, without prior knowledge, the best initial clustering number that a dataset can produce is unknown. The other feature is that each point is connected to the nearest cluster (Bottou and Lin, 2007; Tang et al., 2019). The pseudocode of the K-means clustering algorithm is shown in Figure 1.

```
Start: Open image.
       Crop the image.
       Assign K-means number of centers (number of clusters).
       Calculate the number of centers using the mean values.
       Assign the mean value for each center.
End:   Save new image.
```

Figure 1. Pseudo code

Otsu Thresholding

This algorithm is based on the maximum inter-class variance between the background and the target image as the threshold selection rule. It separates the image into foreground and background based on its grayscale properties. When the best threshold is taken, the difference between the two parts is the largest. Since variance is an important measure of the uniform gray distribution, the larger the variance value, the greater the difference between the two parts of the graph. If some targets are erroneously divided into backgrounds, or if some backgrounds are divided into targets, the difference between the two parts becomes smaller. Therefore, as long as the variance between clusters is maximized, the possibility of misclassification will be minimized and thus perfect segmentation of an image will be achieved (Mittal and Saraswat, 2018). The Otsu threshold value of each image of the study is shown in Table 1.

The basic principle of threshold segmentation based on Otsu is as follows: Assuming that the range of grayscale of the image is $i = 0,1, ..., L -1$ and the pixel number with grayscale k is n_k, then the total number of pixels k in an image is shown in Equation 2.

$$N = \sum_{k=0}^{L-1} n_k = n_0 + n_1 + \cdots + n_{L-1} \quad (2)$$

The probability of occurrence of gray level k is shown in Equation 3.

$$P_k = \frac{n_k}{N} = \frac{n_k}{\sum_{k=0}^{L-1} n_k} \quad (3)$$

The gray level threshold t can be used to divide the gray level of an image into two parts: $C_0 =$ (0,1,2,…, t), $C_1 =$ (t + 1, t + 2,…, L -1), then the probability and mean of the class C_0 and C_1 are as follows:

The probability P is calculated separately for each pixel value, as in Eq. 4. There is the following relationship between them:

$$u_t = \sum_{i=0}^{L-1} iP_k \quad (4)$$

For any value of *t*, equation is expressed as in 5,6.

$$w_0u_0 + w_1u_1 = u_t \quad (5)$$

$$w_0 + w_1 = 1 \quad (6)$$

w0 = Probability of class 1(separated by threshold), w1=Probability of class2 (separated by threshold), u0= class mean u0, u1= class mean u1

Eq.5, compute sigma variance (between class)

Eq.6, the desired threshold corresponds to the maximum variance of between classes.

When summing the variances of C_0 and C_1, equation is expressed as in 7,8.

$$\sigma_0^2 = \sum_{i=1}^{t} (i - u_0)^2 P_k / w_0 \quad (7)$$

$$\sigma_1^2 = \sum_{i=t+1}^{L-1} \frac{(i-u_1)^2 P_k}{w_1} \quad (8)$$

The inter-class variance is defined Eq.9.

$$\sigma_{w_1}^2 = w_0\sigma_0^2 + w_1\sigma_1^2 \quad (9)$$

w_0 and w_1 = They are probabilities of two classes divided by a threshold.

The population's inter-class variance is defined Eq.10.

$$\sigma_T^2 = \sigma_B^2 + \sigma_w^2 \quad (10)$$

Introduction of decision criteria on *t* is defined Eq.11.

$$w_0 = P_r(C_0) = \sum_{r=0}^{t} p_k = w(t) \quad (11)$$

$$w_1 = P_r(C_1) = \sum_{i=t+1}^{L-1} p_k = 1 - w(t) \quad (12)$$

$$u_0 = \sum_{i=0}^{t} ip_k / w_0 = u_t / w(t) \quad (13)$$

$$u_1 = \sum_{i=t+1}^{L-1} \frac{ip_k}{w_1} = \frac{u_T - u(t)}{1 - w(t)} \quad (14)$$

$$\partial(t) = \frac{\sigma_B^2}{\sigma_w^2} \quad (15)$$

$$\sigma(t) = \frac{\sigma_B^2}{\sigma_T^2} \quad (16)$$

$$K(t) = \frac{\sigma_T^2}{\sigma_w^2} \quad (17)$$

Through analysis, it is not difficult to see that the above three criteria are equivalent to each other. They all regard the best value t separated from the class C_0 and C_1 as the best threshold value. Therefore, $\partial(t), \sigma(t), K(t)$are recognized as the maximum judgment criterion. Because,σ_B^2 is the statistical characteristics based on the first order, while,$\sigma_w^2$and$\sigma_B^2$are the functions of the threshold value t, so it is the simplest to choose $\sigma(t)$of the three as the criterion, and the best threshold value t* can be obtained as a range with in shown Eq.19.

$$t *= \arg_{0\leq t\leq L-1} \max \sigma(t) \tag{18}$$

From the above deduction, it can be seen that when σ_B^2 is the maximum value, the best threshold value t∗ of t can be obtained.

Performance Metrics

Performance analysis is the determination of the accuracy of tumor detection algorithms.

TP (True Positive): It is the result found if the tumor region is labeled as “Tumor” in the segmentation process. It is shown in white in our study.

TN (True Negative): It is the result found if the non-tumor region is not labeled as “Tumor” in the segmentation process. It is shown in black in our study.

FP (False Positive): It is the result found if the tumor region is not labeled as “Tumor” in the segmentation process. It means the wrong guess. It is shown in green in our study.

FN (False Negative): It is the result found if the non-tumor region is labeled as “Tumor” in the segmentation process. It means the wrong guess. It is shown in pink in our study.

This information alone may not be meaningful. Performance metrics are calculated using their values. Values ranging from [0, 1] are used to quantify the performance of the algorithms. Formulas for some performance metrics are given in shown Eq.20-27.

$$\text{TPR(True Positive Rate)} = \frac{\text{TP}}{\text{TP+FN}} \tag{19}$$

$$\text{FPR(False Positive Rate)} = \frac{\text{FP}}{\text{FP+TN}} \tag{20}$$

$$\text{Similarity(Jaccard)} = \frac{\text{TP}}{\text{TP+FP+FN}} \tag{21}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{22}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{23}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{24}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP+TN}} \tag{25}$$

$$\text{F Score} = \frac{2*\text{TP}}{2*\text{TP+FP+FN}} \tag{26}$$

Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the actual values are known. The values TP, TN, FP, and FN are transferred to a table defined as a Confusion Matrix or “Error Matrix”. This table contains the accuracy of the prediction results given by a given classifier on a given data set in a two-class data set. It is divided into 2 groups, namely, what actually happened and what was predicted. The predicted class is with tumors, the real class is without tumors. These statements are shown in a table in Table 2.

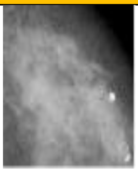
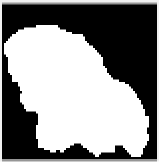
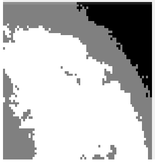
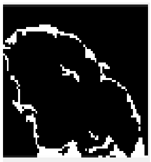
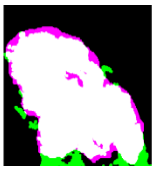
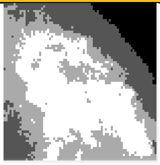
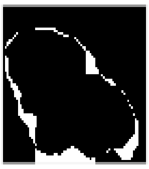
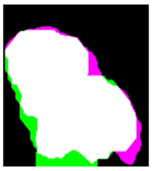
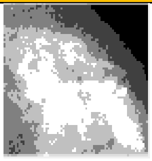
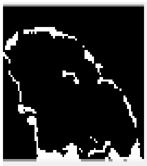
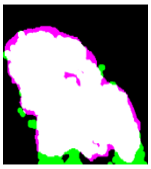
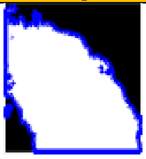
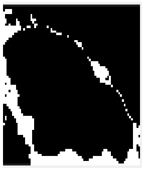
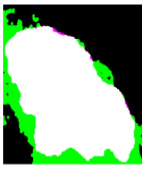
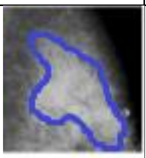
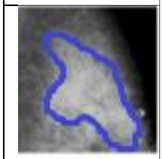
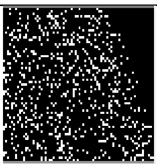
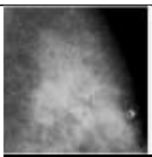
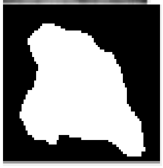
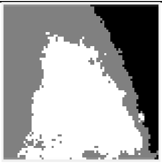
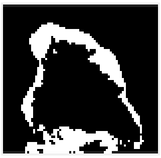
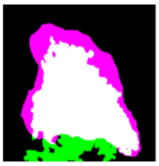
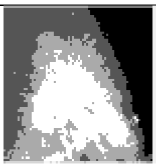
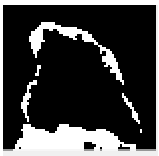
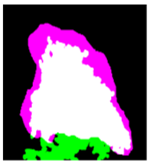
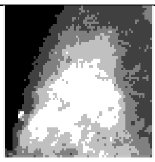
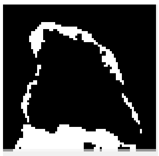
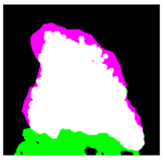
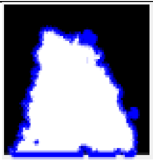
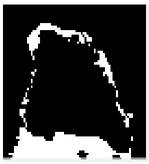
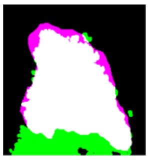
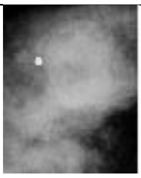
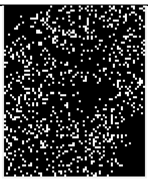
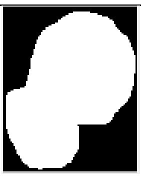
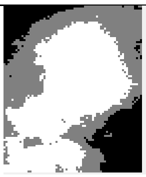
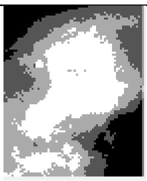
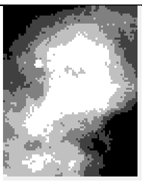
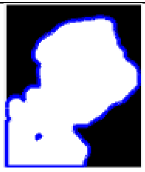
Table 2. Confusion Matrix

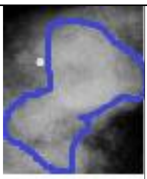
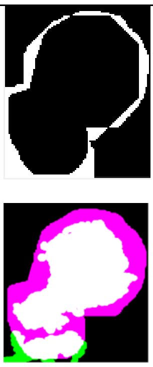
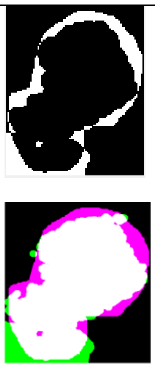
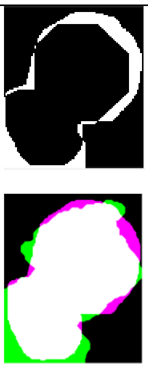
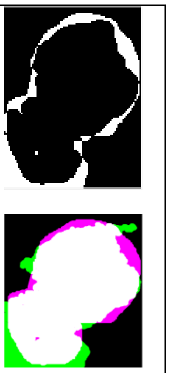
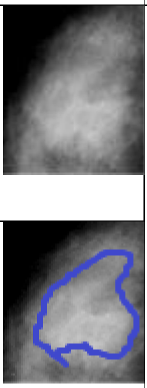
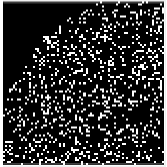
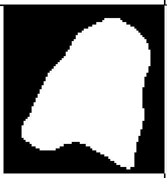
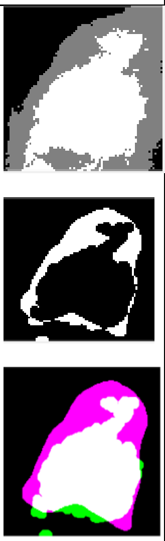
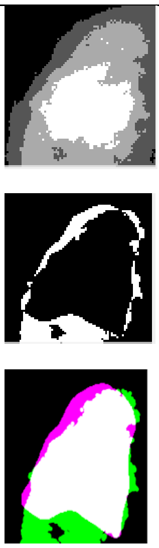
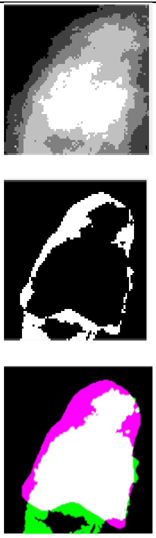
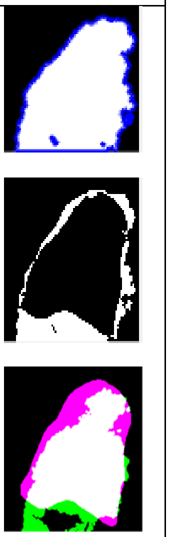
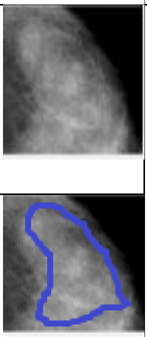
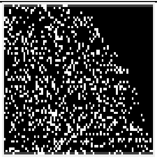
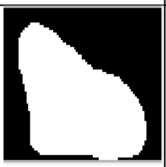
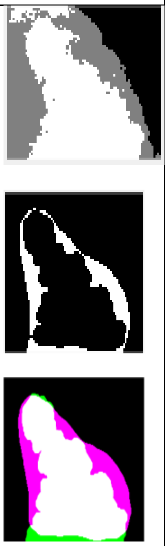
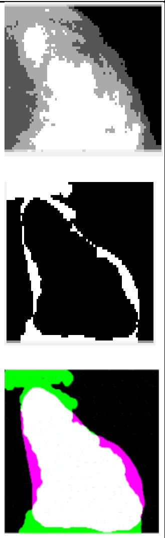
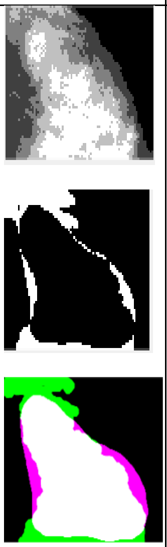
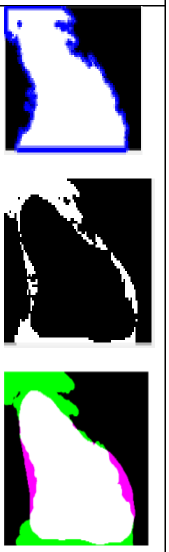
Confusion Matrix		Estimated	
		No	Yes
Real Value	No	True Negative, TN	False Positive, FP
	Yes	False Negative, FN	True Positive, TP

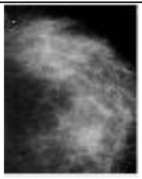
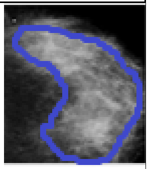
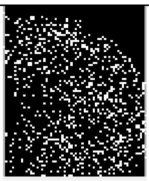
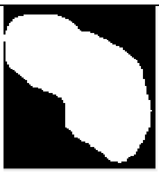
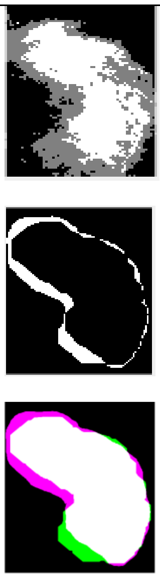
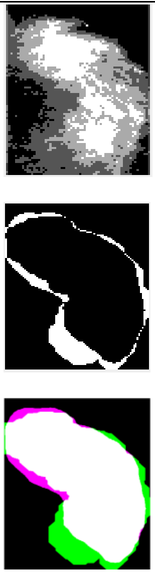
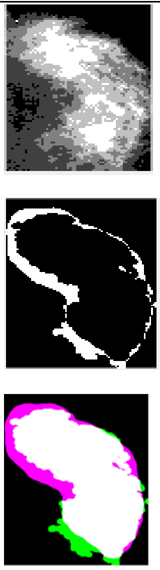
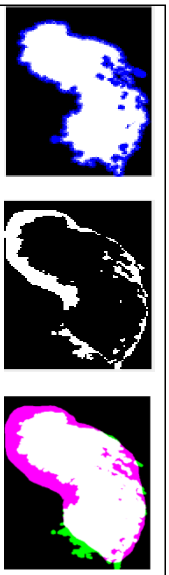
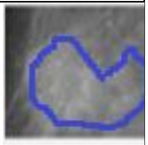
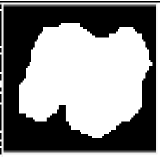
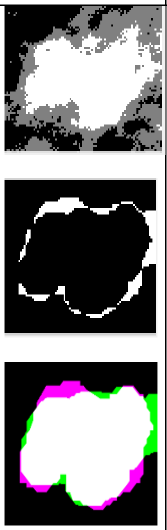
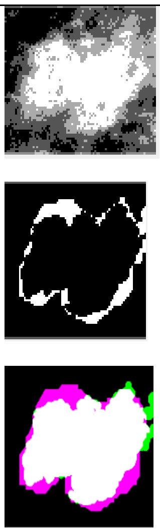
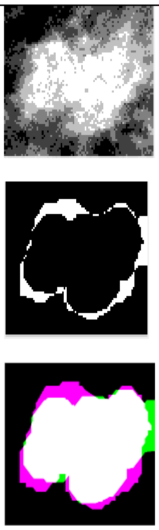
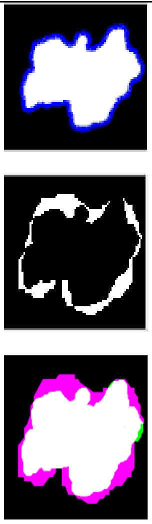
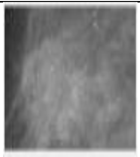
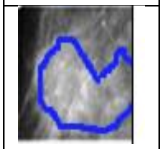
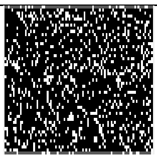
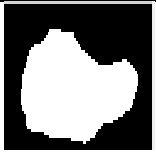
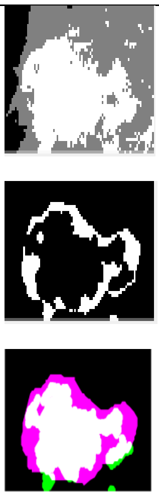
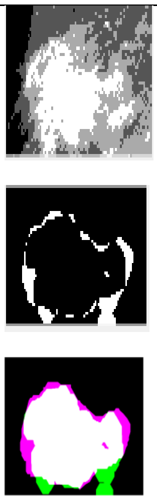
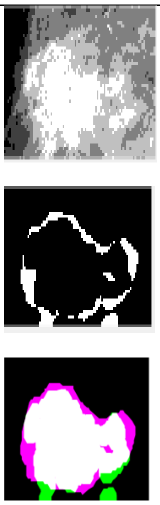
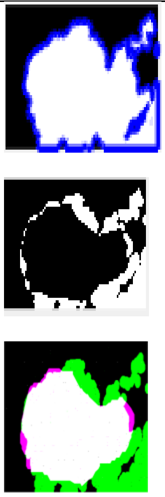
Results

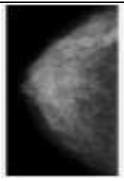
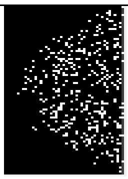
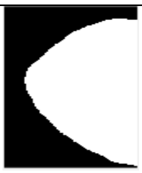
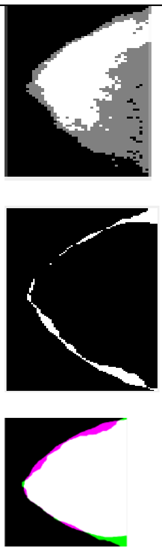
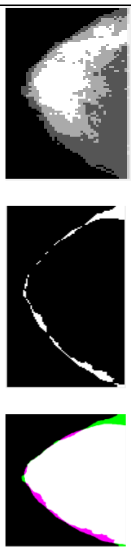
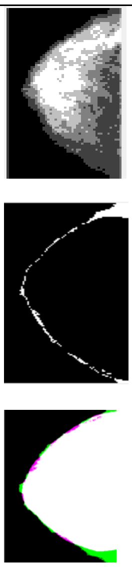
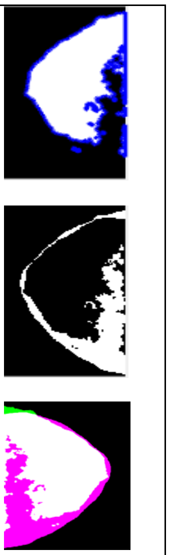
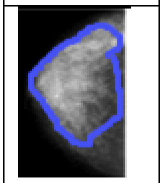
Clustering and Otsu thresholding were applied to the images. The 36 images and difference areas of the processes are shown in Table 3.The lesion sample marking via mammography image was shared as a binary image by the radiologist in the column titled "ground truth”.

Table 3. Area representation of images

Image No	Original Image	Canny filter	Ground Truth	K=3	K=4	K=5	Otsu Threshold
1 Calc-Test_P_00100_R211				  	  	  	   Otsu Threshold Value: 0.400
2 Calc-Test_P_00202_RIGHT	 		 	  	  	  	   Otsu Threshold Value: 0.384
3 Calc Training_P_002 14_RIGHT_CC							

							 Otsu Threshold Value: 0.376
4 Calc-Test_P_00368_LEFT_CC							 Otsu Threshold Value: 0.329
5 Calc-Test_P_00608_LEFT_MLO							 Otsu Threshold Value: 0.298

6 Calc-Test_P_00678_LEFT_CC	 						 Otsu Threshold Value: 0.305
7 Calc-Test_P_00876_LEFT_MLO	 						 Otsu Threshold Value: 0.545
8 Calc-Test_P_00906_LEFT_MLO	 						 Otsu Threshold Value: 0.415

9 Calc-Test_P_00974_RIGHT_MLO							 Otsu Threshold Value: 0.223
							

According to the images obtained after clustering and Otsu thresholding, it was observed that the tumor borders became clearer as the number of clusters increased. The white region in the table represents TP, the black region represents TN, the pink region represents FN, and the green region represents FP. These statements are shown in Tables 13,14.

Table 4. Image 1 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted /Actually Happening	Positive	Negative	Predicted /Actually Happenig	Positive	Negative
Positive	TP 584287	FN 38193	Positive	TP 616683	FN 65522	Positive	TP 604262	FN 54167	Positive	TP 658236	FN 30394
Negative	FP 77458	TN 584062	Negative	FP 45062	TN 556733	Negative	FP 57483	TN 568088	Negative	FP 3509	TN 445311

Table 5. Image 2 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 588778	FN 119194	Positive	TP 699789	FN 180354	Positive	TP 764792	FN 221313	Positive	TP 737356	FN 206171
Negative	FP 235191	TN 1009638	Negative	FP 124180	TN 948478	Negative	FP 59177	TN 907519	Negative	FP 86613	TN 922661

Table 6. Image 3 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 300444	FN 17147	Positive	TP 394900	FN 38760	Positive	TP 428464	FN 56404	Positive	TP 418954	FN 52831
Negative	FP 164293	TN 277476	Negative	FP 69837	TN 255863	Negative	FP 36273	TN 238219	Negative	FP 45783	TN 241792

Table 7. Image 4 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 490374	FN 38003	Positive	TP 643461	FN 164741	Positive	TP 563416	FN 90054	Positive	TP 682765	FN 225540
Negative	FP 235111	TN 824096	Negative	FP 82024	TN 697358	Negative	FP 162069	TN 636550	Negative	FP 42720	TN 636550

Table 8. Image 5 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 783234	FN 53727	Positive	TP 919601	FN 200614	Positive	TP 959719	FN 258679	Positive	TP 975775	FN 289985
Negative	FP 253995	TN 1084164	Negative	FP 117628	TN 937277	Negative	FP 77510	TN 879212	Negative	FP 61454	TN 847906

Table 9. Image 6 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 3440922	FN 245964	Positive	TP 3710296	FN 626815	Positive	TP 3325024	FN 275105	Positive	TP 3052316	FN 164850
Negative	FP 524125	TN 3429014	Negative	FP 254751	TN 3048163	Negative	FP 640023	TN 3399873	Negative	FP 912731	TN 3510128

Table 10. Image 7 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative	Predicted/Actually Happening	Positive	Negative
Positive	TP 48623	FN 4197	Positive	TP 44537	FN 1973	Positive	TP 45383	FN 2484	Positive	TP 41568	FN 340

Negative	FP 5506	TN 50010	Negative	FP 9592	TN 52234	Negative	FP 8746	TN 51723	Negative	FP 12561	TN 53867

Table 11. Image 8 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negaive	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative
Positive	TP 90553	FN 5035	Positive	TP 115638	FN 11833	Positive	TP 110090	FN 10191	Positive	TP 128076	FN 63628
Negative	FP 43636	TN 182076	Negative	FP 18551	TN 175278	Negative	FP 24099	TN 176920	Negative	FP 6113	TN 123076

Table 12. Image 9 Confision Matrix

Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative	Predicted/ Actually Happening	Positive	Negative
Positive	TP 2288415	FN 65401	Positive	TP 2326918	FN 76170	Positive	TP 2393806	FN 108358	Positive	TP 1646286	FN 30394
Negative	FP 140778	TN 1897707	Negative	FP 102275	TN 1886938	Negative	FP 35387	TN 1854750	Negative	FP 782907	TN 1932714

Table 13. Performance metrics of processes

No		Number of Clusters	FN (Pixel, False Negative)	FP (Pixel, False Positive)	TN (Pixel, True Negative)	TP (pixels, True Positive)	TPR (True Positive Rate)	FPR (False Positive Rate)	Similarity	Accuracy	Sensitivity	Precision	Specificity	F score	Duration (sec)
1	Cluster	K=3	53111	66281	569144	595464	0.91	0.10	0.83	0.90	0.91	0.89	0.89	0.90	3.71
		K=4	65522	45062	556733	616683	0.90	0.07	0.84	0.91	0.90	0.93	0.92	0.91	
		K=5	54167	57483	568088	604262	0.91	0.09	0.84	0.91	0.91	0.91	0.90	0.91	
	Otsu	Otsu	176944	3509	445311	658236	0.78	0.01	0.78	0.85	0.78	0.99	0.99	0.87	0.50
2	Cluster	K=3	119194	235191	1009638	588778	0.83	0.18	0.62	0.81	0.83	0.71	0.81	0.76	4.56
		K=4	180354	124180	948478	699789	0.79	0.11	0.69	0.84	0.79	0.84	0.88	0.82	
		K=5	221313	59177	907519	764792	0.77	0.05	0.73	0.85	0.77	0.92	0.93	0.84	
	Otsu	Otsu	206171	86613	922661	737356	0.78	0.08	0.71	0.85	0.78	0.89	0.91	0.83	0.63
3	Cluster	K=3	764792	32279	222240	432458	0.85	0.12	0.80	0.86	0.85	0.93	0.87	0.89	2.82
		K=4	34907	85716	259716	379021	0.91	0.24	0.75	0.85	0.91	0.84	0.78	0.87	
		K=5	38240	79691	256383	385046	0.90	0.23	0.82	0.87	0.90	0.92	0.86	0.90	
	Otsu	Otsu	41250	58012	253373	406725	0.90	0.18	0.76	0.84	0.90	0.90	0.81	0.89	0.38
4	Cluster	K=3	38003	235111	824096	490374	0.92	0.22	0.64	0.82	0.92	0.67	0.77	0.78	4.60
		K=4	164741	82024	697358	643461	0.79	0.10	0.72	0.84	0.79	0.88	0.89	0.83	
		K=5	90054	162069	772045	563416	0.86	0.17	0.69	0.84	0.86	0.77	0.82	0.81	
	Otsu	Otsu	225549	42720	636550	682765	0.75	0.06	0.71	0.83	0.75	0.94	0.93	0.83	0.64
5	Cluster	K=3	53727	253995	1084164	783234	0.93	0.18	0.71	0.85	0.93	0.75	0.81	0.83	4.86
		K=4	209174	109907	928717	927322	0.81	0.10	0.74	0.85	0.81	0.89	0.89	0.85	
		K=5	258679	77510	879212	959719	0.78	0.08	0.74	0.84	0.78	0.92	0.91	0.85	
	Otsu	Otsu	289985	61454	847906	975775	0.77	0.06	0.73	0.83	0.77	0.84	0.83	0.84	0.68
6	Cluster	K=3	245964	524125	3429014	3440922	0.93	0.13	0.81	0.89	0.93	0.86	0.86	0.89	11.32
		K=4	626815	254751	3048163	3710296	0.85	0.07	0.80	0.88	0.85	0.93	0.92	0.89	
		K=5	275105	640023	3399873	3325024	0.92	0.15	0.78	0.88	0.92	0.83	0.84	0.87	
	Otsu	Otsu	164850	912731	3510128	3052316	0.94	0.20	0.73	0.85	0.94	0.76	0.79	0.84	1.80
7	Cluster	K=3	4197	5506	50010	48623	0.92	0.09	0.83	0.91	0.92	0.89	0.90	0.90	1.70
		K=4	1973	9592	52234	44537	0.95	0.15	0.79	0.89	0.95	0.82	0.84	0.88	
		K=5	2484	8746	51723	45383	0.94	0.14	0.80	0.89	0.94	0.83	0.85	0.88	
	Otsu	Otsu	340	12561	53867	41568	0.99	0.18	0.76	0.88	0.99	0.76	0.81	0.81	0.26
8	Cluster	K=3	5035	43636	182076	90553	0.94	0.19	0.65	0.84	0.94	0.67	0.80	0.78	2.04
		K=4	11833	18551	175278	115638	0.90	0.09	0.79	0.90	0.90	0.86	0.90	0.88	
		K=5	10191	24099	176920	110090	0.91	0.11	0.76	0.89	0.91	0.82	0.88	0.86	
	Otsu	Otsu	63628	6113	123483	128076	0.66	0.04	0.64	0.78	0.66	0.75	0.85	0.78	0.30
9	Cluster	K=3	65401	140778	1897707	2288415	0.97	0.06	0.91	0.95	0.97	0.94	0.93	0.95	9.38
		K=4	76170	102275	1886938	2326918	0.96	0.05	0.92	0.95	0.96	0.95	0.94	0.96	
		K=5	108358	35387	1854750	2393806	0.95	0.01	0.94	0.96	0.95	0.98	0.98	0.97	
	Otsu	Otsu	30394	782907	1932714	1646286	0.98	0.01	0.66	0.81	0.98	0.67	0.71	0.80	1.11

Table 14. Mean process performance metrics

	TPR	FPR	Similarity	Accuracy	Sensitivity	Precision	Specificity	F score
Clustering Mean Performance Metrics	**0.89**	**0.14**	**0.77**	**0.87**	**0.89**	**0.86**	**0.87**	**0.87**
Otsu Threshold Mean Performance Metrics	**0.84**	**0.12**	**0.73**	**0.84**	**0.84**	**0.86**	**0.87**	**0.84**

Discussion and Conclusion

In this study, a fully automated computer-aided diagnosis (CAD) algorithm was designed for manually segmented breast cancer images. When the number of clusters was selected as 3, the pixels were seen in white, black, and gray tones. When 4 and 5 were selected, they were divided into black, white, and different shades of gray. It was seen only in the black and white (binary) form in Otsu. Images from the database weremarked by the expert radiologist. Ground truth (reference images) and tumor region images obtained as a result of clustering and the Otsu threshold process were compared. Performance metrics were used to determine segmentation performance. In performance measurement metrics, means were compared to make an overall comparison for 36 images. For the clustering process, TPR was 0.89, FPR was 0.14, the similarity was 0.67, accuracy was 0.87, sensitivity was 0.89, sensitivity was 0.86, specificity was 0.87, F score was 0.87. For Otsu, TPR was 0.84, FPR was 0.12, the similarity was 0.73, accuracy was 0.84, sensitivity was 0.84, sensitivity was 0.86, specificity was 0.87, F score was 0.84. Both methods were found to be successful and close to each other.

FP (False Positive) is pixels that cannot be monitored as a radiologist's tumor section and were actually seen as tumors (in software) and were shown in green. For FN (False Negative); working as a radiologist tumor zone was the part that was not actually seen as a tumor zone (in software). These encounters of his appearance were made pixel by pixel. This study aimed to reduce a radiographic error in examinations of cancerous tissue in patients diagnosed with breast cancer, which can be better determined with software. In the name of dividing into benign(benign) and malignant(malignant) tumor; determining the boundaries of the cancerous lesion was important for patients diagnosed with breast cancer who went to routine check-ups with short periods of time, comparing with previous tumor exams, whether the tumor was benign or malignant, and the course of treatment was before the opening.

With a slight difference, it can be said that the clustering algorithm was more suitable and usable in terms of tumor detection. In addition, it can be given among its other advantages that it showed the pixel color toning in more detail, and the process wascompletedin a shorter time. On the other hand, the Otsu algorithm resulted in a much shorter time compared to the clustering algorithm. The tumor had also performed the results obtained in the determination of scientific data on the accuracy of criteria used in the studies are listed.These criteria were evaluated depending on the parameters TP, TN, FP, FN, the part that the radiologist marks as a tumor pixel by pixel, and the part that the software considers a tumor. In addition, coloring was done to distinguish it. The reason why it was made pixel-based was to minimize the error rate in the study.

Kapoor and Singhal compared K-Means, K-Means++, and Fuzzy C-Means clustering algorithms. Experimental results, similar to our study, showed that in case of an increase in the number of data points, the number of iterations, which greatly affects the cluster performance, was reduced, the duration was shortened, the fluctuations in the cluster center and the time complexity were reduced, in addition, the sum of the distance that changes the performance was minimized (Kapoor and Singhal, 2017). Dallali et al. (2018) found that the Otsu thresholding algorithm was less inaccurate and provided optimum performance with an accuracy of 98.83% in mammography images (Dallali et al., 2018). Similarly, in a study conducted with 36 patient data, the area and volumes obtained using K-means clustering and Otsu thresholding approaches on single or multi-section images were compared by a nuclear medicine specialist. As a result, it was observed that the Otsu thresholding algorithm was more selective (Tianwen et al., 2019). Malali et al. (2020), in their study, reached a 90% accuracy rate in mammography images with the K-means algorithm recommended in breast cancer (Malali et al., 2020). The recommended algorithm increased the sensitivity by 21%. While Aswathy and Jagannath (2020) obtained 91% accuracy based on SVM, it had 93% maximum segmentation accuracy with K-means clustering (Aswathy and Jagannath, 2020). K-means and Otsu thresholding were applied to mammogram images taken from MIAS. The results showed that the proposed methods were easy and high sensitivity of 92.93% was achieved with a high reduction in 1.98 FPPI (Aksebzeci, 2017). In the study of Dubey et al.(2018), the highest and lowest clustering accuracies were 94.7%, 77.1%, and 94.4%, 88.5% for fuzzy and random centroid, respectively. The accuracy obtained with this approach was approximately 92% (Dubey et al., 2018).

Bradley and Fayyad (1998) used the K-means algorithm to improve the starting points and achieved an acceptable low run time (Bradley and Fayyad, 1998). Similarly, Karen et al. (2021) used the K-means algorithm to improve groups and used colony optimization to improve cluster quality (Karen et al., 2021). Ghosh and Dubey (2013) presented the comparison between KM and FCM based on sample number and K. Experimental results showed that the K-means clustering algorithm was much better than FCM because it took more time to perform fuzzymeasurement calculations (Ghosh and Dubey, 2013).

Time complexity affected the outcome. Thus, there was no doubt that FCM produced results as good as those produced by KM results, but the time complexity was still relatively high. Banerjee et al. (2015) compared various variants of KM, bisecting KM, FCM, and genetic KM. Genetic KM performed best for both internal and external indices (Banerjee et al., 2015). On the other hand, Kaygisiz and Cakir (2020) achieved successful results with Otsu thresholding (Kaygisiz and Cakir, 2020).

In this study, it was aimed to create prototyping with a high success rate. More reliable results could be obtained with a richer data set. However, building a model was our main goal. In addition, we achieved high performance in a short time with our simple algorithm without the need for very complex processes. Our findings may prove that the algorithm can be used by doctors to diagnose breast cancer. This tool is more useful for areas far from urban or rural areas where medical professionals or oncologists may not be available. Thanks to advances in image acquisition and appropriate tools, the diagnosis can be confirmed using this system, serving the automated diagnosis of breast cancer.

However, in the study, images with a low accuracy rate were obtained as well as images with high accuracy. Radiologists may be mistaken when marking sites, or marked values may be only approximate. Therefore, ground truth accuracy is also a controversial issue. Also, after following the diagnostic system steps, the first stage is image development; however, breast images often contain artifacts such as uneven lighting, adipose tissue, milk ducts, and rich vascular structure. In conclusion, robust methods are needed to remove artifacts and detect lesion borders in breast images. In future work, we also plan to test our methods on advanced neural networks and machine learning so that we can shed some light on some of these "ground truth" issues. In this study, it was aimed to determine tumor boundaries more accurately with pixel-based segmentation, to reduce the need for human beings, and to develop medical devices used in field imaging in the field of health with computer-aided software using image processing methods in a shorter time with fewer data sets.

In future work, it is planned to diagnose breast cancer using deep learning methods. Deep learning has come to the forefront as the rising trend of recent years in the diagnosis of diseases from medical images. There are important studies in the literature on the diagnosis of various diseases with deep learning. To name a few examples, breast cancer diagnosis (Shen et al., 2019), brain tumor diagnosis (Irmak, 2021), malaria disease detection (Irmak, 2021), COVID-19 disease detection (Irmak, 2020) are some important applications of deep learning in the diagnosis of medical diseases. It will be interesting to use deep learning methods in the diagnosis of breast cancer.

Acknowledgments

This project and article were supported by Osmaniye Korkut Ata University and Institute of Natural and Applied Sciences, OKUBAP-2020-PT3-011 Bap unit.

Conflict of Interest

There is no conflict of interest between the authors.

References

Akay M. Automatic mass segmentation in mammographic images. Universitat de Girona, Spain, 2006.

Aksebzeci B. Computer-aided classification of breast cancer histopathological image., 2017.

Al-Azhar. Assessment of the diagnostic accuracy of contrast-enhanced digital mammography in the differentiation between benign and malignant breast mass lesions. International Medical Journal, 2021; 2(1): 90-96.

Aswathy MA., Jagannath M. Performance analysis of segmentation algorithms for the detection of breast cancer. Procedia Computer Science 2020; 167: 666-676.

Banerjee S., Choudhary A., Pal S. Empirical evaluation of k-means, k-means bisection, fuzzy c-means and genetic k-means clustering algorithms. IEEE international WIE conference on Electrical and Computer Engineering 2015; 168-172.

Birdwell RL., Ikeda DM., O'Shaughnessy KF., Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 2001; 219(1): 192–202.

Bottou L., Lin CJ. Support vector machine solvers. Large Scale Kernel 2007; 3(1): 301–320.

Bradley PS., Fayyad UM. Refining ınitial points for k -means clustering. 15th International Conference on Machine Learning, San Francisco-ABD, 1998; 91-99.

Canadian Cancer Statistics Advisory Committee. Canadian Cancer Statistics, 2018.

Cancer Facts and Figures. American Cancer Society. Atlanta, 2020; 1-76.

Ciecholewski M. Microcalcification segmentation from mammograms: a morphological approach. Journal Digit Imaging 2017; 30(2): 172–184.

Çiklaçandir FGY., Ertaylan A., Bınzat U., Kut A. Lesion detection from the ultrasound images using k-means algorithm. Medical Technologies Congress (TIPTEKNO) 2019; 1-4.

Dallali A., el Khediri S., Amel SA., Kachouri A. Breast tumors segmentation using otsu method and K-means. ATSIP 2018; 1-6.

Das S., Abraham A., Konar A. Automatic clustering using an ımproved differential evolution algorithm. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2008; 38: 218-237.

Dubey KA., Gupta U., Jain S. Comparative study of k-means and fuzzy c-means algorithms on the breast cancer data. International Journal on Advanced Science Engineering Information Technology, 2018; 8(1): 18-29.

Etehadtavakol M., Ng EYK. Survey of numerical bioheat transfer modelling for accurate skin surface measurements. Thermal Science and Engineering Progress, 2020; 20(1): 100681..

Et-taleby A., Boussetta M., Benslimane M. Faults detection for photovoltaic field based on k-means, elbow, and average silhouette techniques through the segmentation of a thermal image. International Journal of Photoenergy, 2020; 6617597.

Ghosh S., Dubey KS. Comparative analysis of k-means and fuzzy c means algorithms. (IJACSA). International Journal of Advanced Computer Science and Applications 2013; 4(4): 35-39.

Gunderman RB. Essential radiology. Clinical presentation, pathophysiology, Germany. 2006.

https://www.cancerimagingarchive.net/ Erişim Tarihi: 20.04.2021

https://www.mathworks.org/.. Erişim Tarihi: 22.04.2021

https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/ Erişim Tarihi: 22.04.2021

Isaac DN., Alexis R., Patrik B., Roberto LG., Karin B., Daniel T., Dilip DG., Jeffrey SR., Sunitha

BT., Katja PD. Multidimensional diffusion magnetic resonance ımaging for characterization of tissue microstructure in breast cancer patients: A prospective pilot study. MDPI Cancers 2021; 13(7): 1606.

Irmak E. Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework, Iranian Journal of Science and Technology-Transactions of Electrical Engineering 2021;45: 1015–1036.

Irmak E. A novel implementation of deep-learning approach on malaria parasite detection from thin blood cell images Electrica 2021; 21(2): 216-224.

Irmak E. Implementation of convolutional neural network approach for COVID-19 disease detection. Physiol Genomics 2020; 52(12): 590-601.

Kapoor A., Singhal A. A comparative study of k-means, k-means++ and fuzzy c-means clustering algorithms. 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT) 2017; 1-6.

Karen SJ., Emily FC., Mary SS. Molecular subtypes of breast cancer: A review for breast radiologists. Journal of Breast Imaging 2021; 3(1): 12–24.

Katz E., Barness Y. Comparison of SNR and Peak-SNR (PSNR) as performance measures and signals for peak-limited two-dimensional (2D) pixelated optical wireless communication, in: Conference on Signals, Systems & Computers. IEEE 2015; 1880–1884.

Kaur A. Comparative analysis of segmentation algorithms for brain tumor detection in MR images. Medicine 2017.

Kaur MN., Klassen AF., Xie F., Bordeleau L., Zhong T., Cano SJ., Tsangaris E., Breitkopf T., Kuspinar A., Pusic AL. An international mixed methods study to develop a new preference-based measure for women with breast cancer: the BREAST-Q Utility module. BMC Women's Health 2021; 21(8): 1-17.

Kaur P., Singh G., Kaur P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. Informatics in Medicine Unlocked, 2019; 16: 100151.

Kaygısız H., Çakır A. FPGA kullanılarak görüntülerin gerçek zamanlı olarak OTSU metodu ile bölütlenmesi. Avrupa Bilim ve Teknoloji Dergisi 2020; 18: 911-917.

Khan SS., Ahmad A. Cluster centre initialization algorithm for k-means cluster. In Pattern Recognition Letters 2004; 1293–1302.

Lin H., Ji Z. Breast cancer prediction based on K-Means and SOM Hybrid Algorithm. In Journal of Physics: Conference Series. IOP Publishing 2020;1624(4): 1-7.

Malali HE., Assir A., Harmouchi M., Rattal M., Lyazidi A., Mouhsen A. Adaptive local gray-level transformation based on variable s-curve for contrast enhancement of mammogram images. Embedded Systems and Artificial Intelligence 2020; 671-679.

Mentari BA., Rasyid Y., Fitri A., Khairul M. Histogram statistics and GLCM features of breast thermograms for early cancer detection, 15th International Conference on Electrical Engineering/Electronics. Computer, Telecommunications and Information Technology (ECTI-NCON2018) 2018; 120-124.

Mini G., Thomas T. A neural network method for mammogram analysis based on statistical features, Convergent Technologies for the AsiaPacific Region Conference, In TENCON 2003.

Mittal H.,Saraswat M. An optimum multi-level image thresholding segmentation using non-local means 2D histogram and exponential Kbest gravitational search algorithm, Engineering Applications of Artificial Intelligence 2018; 71: 226–235.

Ng EYK., Kee EC. Advanced integrated technique in breast cancer thermography. Journal of Medical Engineering & Technology 2008; 32:103-114.

Podgornova YA., Sadykov SS. Comparative analysis of segmentation algorithms for the allocation of microcalcifications on mammograms. Information Technology and Nanotechnology 2019; 2391: 121-127.

Rampuna A., Morrowa PJ., Scotneya BW., Winder J. Fully automated breast boundary and pectoral muscle segmentationin mammograms. Elsevier 2017; 1-14.

Sadeghi B., Karimi M., Mazaheri S. Automatic suspicions lesions segmentation based on variable-size windows in mammography images. Health and Technology 2021; 11(1): 99-110.

Sankar D., Thomas T. A new fast fractal modeling approach for the detection of microcalcifications in mammograms, Journal of Digital Imaging 2010; 23(5): 538-546.

Schönenberger C., Hejduk P., Ciritsis A., Marcon M., Rossi C., Boss A. Classification of mammographic breast microcalcifications using a deep convolutional neural network. Investigative Radiology 2021; 56(4): 224-231.

Shen L., Margolies LR., Rothstein JH., Fluder E., McBride R., Weiva S. Deep learning to improve breast cancer detection on screening mammography, Scientific Reports 2019; 9:12495.

Shokrgozar N., Sobhani FM. Customer segmentation of bank based on iscovering of their transactional relation by using data mining algorithms. Modern Applied Science 2016; 10(10): 283-286.

Subramanyeshwar R., Kamala S., Senthil JR., Sudha SM., Rashmi S., Veeraiah CKT. Accuracy of digital mammography, ultrasound and MRI in predicting the pathological complete response and residual tumor size of breast cancer after completion of neoadjuvant chemotherapy. Indian Journal of Cancer 2021; 58(1).

Sun L., Legood R., Sadique Z. Dos-Santos-Silva I., Yang L. Cost–effectiveness of risk-based breast cancer screening programme China. Bull World Health Organ 2018; 96: 568-577.

Tang T., Chen S., Zhao M., Huang W., Luo J. Very large-scale data classification based on K-means clustering and multi-kernel SVM. Soft Computing 2019; 23(11): 3793-3801.

Tianwen X., Qiufeng Z., Caixia F., Qianming B., Xiaoyan Z., Lihua L., Robert G., Li L., Yajia G., Weijun P. In MRI, complete tumor histogram analysis and otsu threshold method to distinguish breast cancer from others 2019.

Toronto ON: Canadian Cancer Society 2020. Available at: cancer.ca/Canadian-Cancer-Statistics-2018-EN.

Yang MS., Sinaga KP. A feature-reduction multi-view K-means clustering algorithm. IEEE 2019; 7: 114472-114486.

Zhu Y., Tan Y., Hua Y., Zhang G., Zhang J. Automatic segmentation of ground-glass opacities in lung CT ımages by using markov random fieldbased algorithms, Journal of Digital Imaging 2012; 25: 409-422.