

Article

Breast Cancer Detection in Mammogram Images Using K-Means++ Clustering Based on Cuckoo Search Optimization

Kittipol Wisaeng 

Technology and Business Information System Unit, Mahasarakham Business School, Mahasarakham University, Mahasarakham 44150, Thailand; kittipol.w@acc.msu.ac.th; Tel.: +66-086-6393870

Abstract: Traditional breast cancer detection algorithms require manual extraction of features from mammogram images and professional medical knowledge. Still, the quality of mammogram images hampers this and extracting high-quality features, which can result in very long processing times. Therefore, this paper proposes a new K-means++ clustering based on Cuckoo Search Optimization (KM++CSO) for breast cancer detection. The pre-processing method is used to improve the proposed KM++CSO method more segmentation efficiently. Furthermore, the interpretability is further enhanced using mathematical morphology and OTSU's threshold. To this end, we tested the effectiveness of the KM++CSO methods on the mammogram image analysis society of the Mini-Mammographic Image Analysis Society (Mini-MIAS), the Digital Database for Screening Mammography (DDSM), and the Breast Cancer Digital Repository (BCDR) dataset through cross-validation. We maximize the accuracy and Jaccard index score, which is a measure that indicates the similarity between detected cancer and their corresponding reference cancer regions. The experimental results showed that the detection method obtained an accuracy of 96.42% (Mini-MIAS), 95.49% (DDSM), and 96.92% (BCDR). On average, the KM++CSO method obtained 96.27% accuracy for three publicly available datasets. In addition, the detection results provided the 91.05% Jaccard index score.

Keywords: breast cancer; mammogram images; K-means++ clustering; cuckoo search optimization



Citation: Wisaeng, K. Breast Cancer Detection in Mammogram Images Using K-Means++ Clustering Based on Cuckoo Search Optimization. *Diagnostics* **2022**, *12*, 3088. <https://doi.org/10.3390/diagnostics12123088>

Received: 15 November 2022

Accepted: 6 December 2022

Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Breast cancer is a challenging and fatal disease that seriously affects women, leading to sudden death in critical cases. Based on the statistics provided by the World Health Organization (WHO) in 2020, there were 2.3 million women diagnosed with breast cancer [1]. As of the end of 2020, WHO reported that 685,000 deaths among women globally were caused by breast cancer. Automated detection of breast cancer from mammogram images is a controversial issue among researchers. Several image processing methods have been proposed. These include a variety of machine learning methods, such as Support Vector Machine (SVM), Convolution Neural Networks (CNN), Naïve Bayesian (NB), and Artificial Neural Networks (ANN). For instance, Karabatak et al. [2] proposed a new algorithm to detect breast cancer using NB on the Wisconsin breast cancer dataset. Additionally, Pak et al. [3] used non-subsampled contour transforms and super-resolution for breast cancer detection. Sabu et al. [4] attempted a comparison of ANN and SVM by selecting the best detection method among 60 cancerous and 60 non-cancerous images. This study focused on breast cancer tumor types by Mini-MIAS, and DDSM. Then, Progressive Support-pixel Correlation Statistical Method (PSCSM) was comparatively used by Yang et al. [5]. Global and local thresholding based on image histogram, Region Growing Method (RGM), Markov Random Field (MRF), stochastic relaxation methods, and fuzzy use membership function were used for breast cancer detection by Punitha et al. [6]. Al-masni et al. [7] suggested deep regional learning based on CNN. Shen et al. [8] proposed the Mixed-Supervision Guided Deep Model (MSGDM) based on U-net model and ResCU-Net. Different machine learning, including K-Nearest Neighbourhood (KNN), Fuzzy Clustering (FC), Neural Network (NN),

Logistic Regression (LG), SVM, and Random Forest (RF), were applied for breast cancer detection. A similarity of three unsupervised learning, including Fuzzy Rough Segmentation (FRS), Rough Segmentation (RS), and Fuzzy Segmentation (FS), was proposed by Punitha et al. [9]. Zebari et al. [10] offered a thresholding method and binary morphological segmentation for breast cancer. Vaka et al. [11] presented end-to-end testing for breast cancer detection using Deep Neural Network with Support Value (DNNSV). Azlan et al. [12] proposed Principal Component Analysis (PCA) and SVM for breast cancer detection. Next, George et al. [13] developed hybrid methods by combining the CNN and SVM on the Breast Cancer Histopathological Database (BreakHis). Alanazi et al. [14] performed a CNN to detect breast cancer disease and took mammogram images from the Kaggle dataset. The details of the existing up-to-date methods for breast cancer classification and detection are illustrated in Table 1.

Table 1. Comparison of machine learning and deep learning methods for segmentation of breast cancer.

Authors	Methods	Dataset	SEN (%)	SPEC (%)	ACC (%)
Karabatak et al. [2]	Naïve Bayesian	Wisconsin breast cancer	99.11	98.25	98.54
Pak et al. [3]	SVM and ANN	Mini-MIAS and DDSM	SVM = 100 ANN = 87.50	SVM = 80 ANN = 93.33	SVM = 92.85 ANN = 91.31
Yang et al. [5]	PSCSM	Mini-MIAS	–	–	90.9
Punita et al. [6]	RGM, MRF segmentation, Stochastic Relaxation methods, and Fuzzy	DDSM	98.10	97.8	98.00
Al-masni et al. [7]	Regional deep learning based on CNN and CLAHE	IN Breast	97.14	92.41	95.64
Shen et al. [8]	MSGDM	IN Breast	97.56	88.89	94.16
Punita et al. [9]	Fuzzy Rough Segmentation (FR), Rough Segmentation (RS), and Fuzzy Segmentation (FS)	Mini-MIAS	FR = 91.23 RS = 80.25 FS = 89.89	FR = 91.23 RS = 79.45 FS = 89.12	FR = 97.45 RS = 82.45 FS = 90.12
Zebari et al. [10]	The thresholding method and binary morphological for segmentation	Mini-MIAS, IN BCDR	98.9	98.4	98.13
Anji et al. [11]	DNNS	MG Cancer	–	–	97.21
Azlan et al. [12]	PCA and SVM	DDSM and Mini-MIAS	93.94	96.61	95.24
George et al. [13]	The hybrid methods by combined the CNN and SVM	BreakHis	97.24	96.18	96.21
Alanazi et al. [14]	CNN architectures	Kaggle 162 H&E	–	–	87.00
Mohammed et al. [15]	SVM	Mini-MIAS	–	–	91.12
Agnes et al. [16]	MA-CNN	Mini-MIAS	96.00	–	–
Yektaei et al. [17]	Multiscale Convolutional Neural Network (MCNN)	Mini-MIAS	95.90	–	97.03
Kaur et al. [18]	Multiclass Support Vector Machine (MSVM)	Mini-MIAS	–	–	96.90
Viswanath et al. [19]	SVM	Raw sample images	–	68	84.84
Jin et al. [20]	Binary classifier with CNNI-BCC	Mini-MIAS	–	–	73.24
Kayode et al. [21]	SVM	Mini-MIAS	94.4	91.3	–
Debelee et al. [22]	SVM and MLP	Mini-MIAS	SVM = 99.48 MLP = 97.40	SVM = 98.16 MLP = 96.26	SVM = 99 MLP = 97
Zhang et al. [23]	Mask R-CNN	DCE-MRI	80	74	75

From a comparative review of machine learning and performance parameters used for evaluating the detection method, each method is suitable for detecting breast cancer on different datasets. Nevertheless, the accuracy of each technique still varies for different datasets. This work aimed to assess the reliability of the diagnosis system for automatically detecting breast cancer in mammogram images. The automated breast cancer detection from mammogram images comprises a series of contributions, which are listed as follows: (1) We resolved the problem of poor quality and contrast enhancement of the original image of the different datasets by using color normalization and noise reduction stages, (2) A comprehensive application that used machine learning such as K-Means (KM), Fuzzy C-Means (FCM), Fuzzy K-Means (FKM), and KM++CSO to select a suitable robust method for the coarse segmentation stage, (3) A fused method for finely detecting breast cancer disease associated with region-based segmentation based on Cuckoo Search Optimization (CSO) is applied, (4) A new method was applied to three open-access datasets (Mini-MIAS, DDSM, and BCDR) with different camera settings, intensities, and additional noise, and (5) Validation of the proposed method by comparing it with Sensitivity (SEN), Specificity (SPEC), Accuracy (ACC), and Jaccard index score.

To achieve all these, we organized the study into four sections. Section 1 presents the background of breast cancer problems and the current literature reviews. It also focuses on segment limitations and how to use the proposed KM++CSO to solve these limitations. Section 2 presents the pre-processing method and the proposed KM++CSO framework phases. Section 3 describes the conducted results and comparisons with the literature. Finally, the conclusion and the analysis of comparison KM++CSO and others are presented in Section 4.

2. Materials and Methods

Since mammogram image is nowadays in digital format, it is possible to create a Computer Aided Diagnosis (CAD) system that automatically segments breast cancer disease. Nevertheless, automatic breast cancer detection using a CAD system requires the education of human screeners or experts for further validation. Therefore, new breast cancer detection is necessary for classifying breast cancer from the image background without unnecessary interference from other regions. Typically, the mammogram image includes noise or artifacts, which could adversely affect breast cancer detection. Sometimes the image is as high-intensity in the center of the image and low-intensity in a region in the posterior margin. This research proposed a fusion method to detect breast cancer from the low-intensity of image and the block diagram for automatic detection of breast cancer, as illustrated in Figure 1.

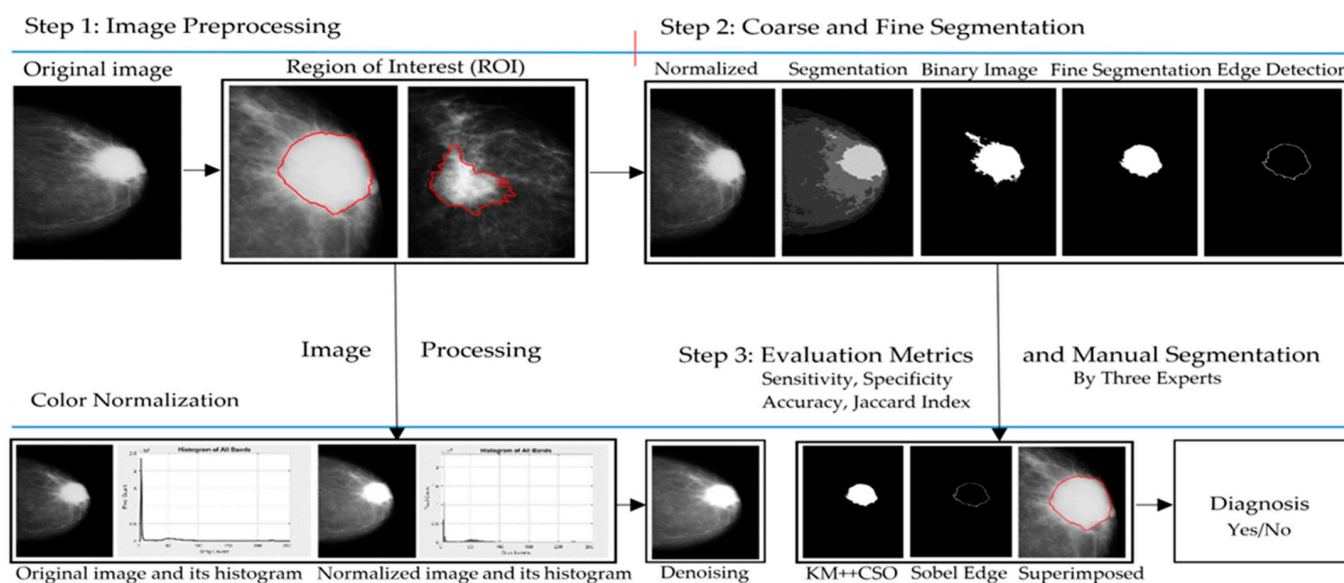


Figure 1. The proposed KM++CSO for segmentation of breast cancer region from mammogram image. The red line represents the area of breast cancer labeled by the specialist.

2.1. Dataset and Image Acquisition

A mammogram image used in this study is drawn from different datasets, some of which are from the Mini-MIAS [5,9,15–19] and DDSM [12], while others are from BCDR [10]. We used the same images to evaluate their detection methods [3–6,9,10,12]. This dataset from DDSM, Mini-MIAS, and BCDR comprises 2620, 323, and 123 images, respectively. Mini-MIAS contains 323 images divided into 207 healthy and 116 abnormal images. DDSM includes 1609 scanned film mammography studies divided into 695 healthy and 914 abnormal images. BCDR contains 18 normal and 105 abnormal images from 123 images. Before discussing the mammogram image, it may be helpful to describe the structure of the image; see Figure 2 [5,9,10,12,15–19]. Figure 2a is an image from Mini-MIAS, Figure 2b is an image from INbreast, and Figure 2c is an image from BCDR, respectively.

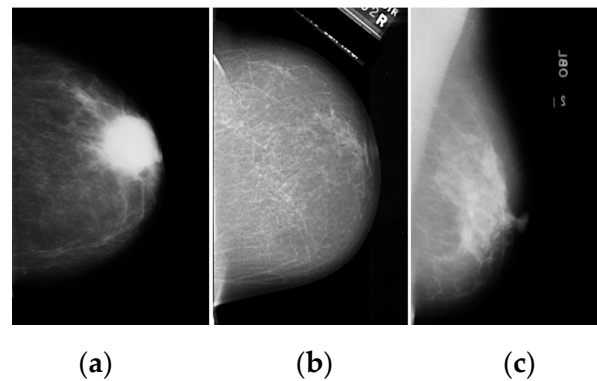


Figure 2. The original image from three open-access datasets: (a) BCDR, (b) DDSM, (c) Mini-MIAS.

2.2. Image Color Normalization

The pre-processing method in many stages is image color normalization, and image analysis is no exception. One of the commonest is the one proposed by Gonzalez [24]. The author adopts a method based on the histogram specification to enhance the intensity of the gray level of the mammogram image. Bejnordi et al. [25], proposed a Whole-Slide Image Color Standardizer (WSMCS) to estimate the background and assumed that a background image usually had distributed color. However, the processing time of WSMCS is very high processing time. Janowczyk et al. [26] proposed stain normalization by using sparse autoencoders for the image to obtain the normalized image. However, the majority involve methods to scan the same image multiple numbers of times. Therefore, it may cause false detection and some of the information lost. In this study, the global histogram specification [27] helps adjust the histogram of the original image or histogram modifications to enhance the contrast and create a new image. For normalizing the values in the range $[0, G]$, the probability density function of the input and output mammogram image that follows a specified shape $f_z(z)$ for $z \in [0, 1]$ is denoted by a subscript (s). If the input is performed to attain processed (output) intensity levels (s), the result would be an image that also has a uniform probability density function using Equation (1).

$$s = G(z) = \int_0^z f_z(w)dw \quad (1)$$

where $f_z(w)$ is the probability density function of the severity levels in the image, G is the grayscale limit (or range), and the transformation function for equalizing the histogram. An adjustable transformation function is generated by histogram specification in the process of the original image with L gray-level values. The implementation of histogram specification and the transformation function $T(r)$ are defined as Equations (2) and (3), respectively.

$$s_k = T(r_k) = \sum_{j=0}^k f_R(r_j) \quad \text{for } k = 0, \frac{1}{L-1}, \frac{1}{L-1}, \dots, 1 \quad (2)$$

$$s_k = G(z_k) = \sum_{i=0}^k f_Z(z_i) \quad \text{for } k = 0, \frac{1}{L-1}, \frac{1}{L-1}, \dots, 1 \quad (3)$$

The shape of the histogram specification is necessary to get the output image for some distinctive implementations. After the implementation of histogram specification, the normalized image is obtained. Here, the images in Figure 3a,b demonstrate the reference image, and its histogram while Figure 3c,d display the obtained image for grayscale after the normalization by histogram and its histogram after histogram specification of grayscale.

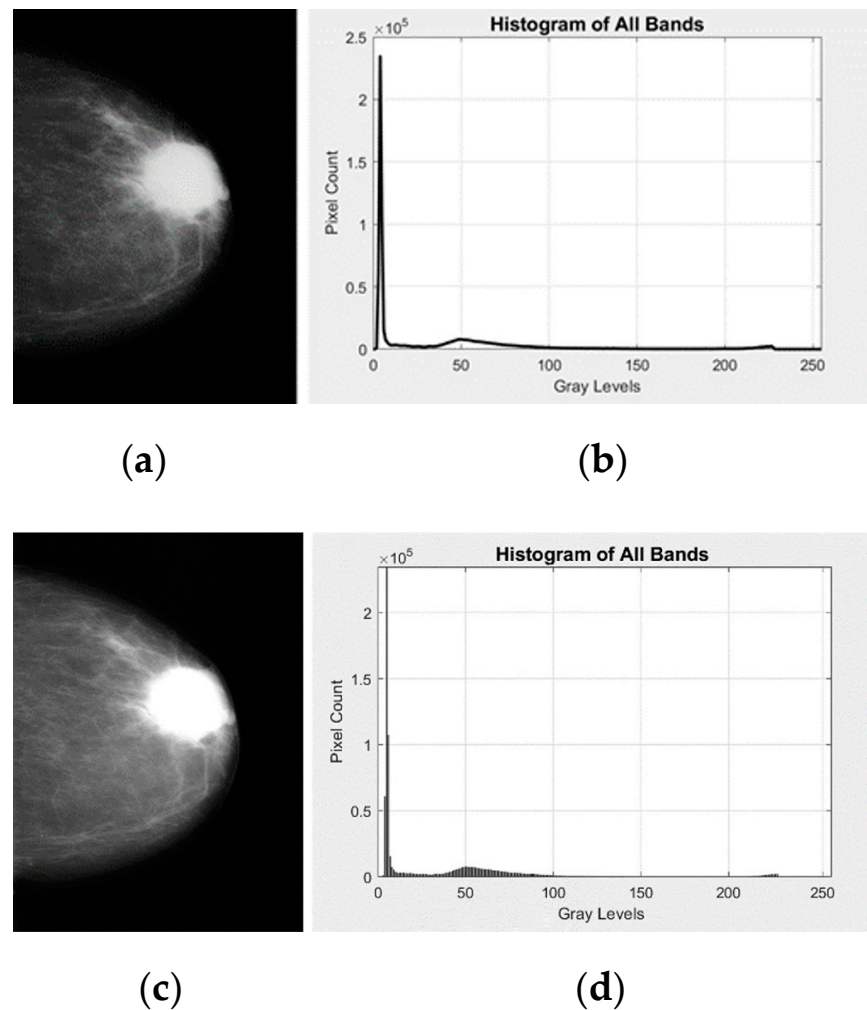


Figure 3. The mammogram image and it is after histogram specification with their respective histograms: (a) looking at the mammogram image, in which the original image appears to contain mostly the same grayscale value, (b) histogram of the original mammogram image, (c) the subtle changes in grayscale images become for the mammogram image to classify in the normalized image, and (d) the peak in the normalized image is approximately 50 grayscale levels, as seen in the normalized histogram. Fortunately, the process improves the segmentation of breast cancer in mammogram images, and it simultaneously enhances the noise in mammogram images.

2.3. Image Denoising

The histogram specification applied in the previous section is helpful for the enhancement of image contrast. However, this step has also increased the intensity of the noise in the image. Therefore, the median filtering [28] was specially designed to remove the noise over-amplification in the image to some extent so that the image looks real. This method runs a window size of a median filter of 3×3 through the whole image. In a window size of the median filter, the targeted pixel is defined as per Equation (4).

$$f(x, y) = \text{median} \left[\sum_{i=-2}^2 \times \sum_{j=-2}^2 \text{window}(x - i, y - j) \right] \quad (4)$$

Median filtering works well when there is one out-layer value in the window. This technique gives a good denoising result when the neighboring pixel values are very close.

2.4. Breast Cancer Detection

In this subsection, we chose a comprehensive application that used machine learning, such as KM, FCM, FKM, and KM++CSO, to select the most suitable algorithm for the coarse segmentation stage. Then, we combined these works with the clustering method coupled with mathematical morphology and OTSU's threshold.

2.4.1. KM Clustering Method

KM clustering [29] is the simplest of machine learning that solves the well-known clustering problem. Given a dataset $\{x\}_1^n \subset \mathbb{R}^d$ and a set of initial center locations $\{c_j\}_1^k \subset \mathbb{R}^d$, we merely alternate between the following two steps repeatedly (by assigning each point x_i to its nearest center and relocating each center c_j to the centroid of all points assigned to it), until center locations converge. Formally, the KM task is an optimization problem that partitions data into clusters while minimizing the sum of the squared distances between each data point and the cluster center it belongs to. Given a set of data points $\{x\}_1^n \subset \mathbb{R}^d$ and a set of initial center locations $\{c_j\}_1^k \subset \mathbb{R}^d$, we defined the minimization objective function $J(x, c)$ as the sum of squared errors given in Equation (5).

$$J(x, c) = \sum_{i=1}^n \underset{j}{\operatorname{argmin}} \|x_i - c_j\|^2 \quad (5)$$

There are n ways to partition a set of n points into k possibly empty clusters. If we insist on strictly non-empty clusters, then the number of ways to partition the set is given by the Stirling numbers of the second kind k , which we define below in Equation (6).

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n \quad (6)$$

The results segmented by using standard KM clustering to mammogram images are depicted in Figure 4. In this stage, $k = 2, 3, 4, 5$, and 6 run 110 iterations, which are used as the initial values for starting the detection. Figure 4b displaying the image detected using $k = 3$ is a “good” value and gives satisfactory results. Since KM is the simplest and most frequently unsupervised, we can implement it efficiently. Moreover, KM requires a few computations for comparing distance among data points and grouping clusters. However, KM has several limitations. First, the cluster that the method produces may be arbitrary and far from the optimal clustering. Second, the technique requires the user to specify k and the number of clusters and provides an initial guess of cluster locations. Finally, KM takes a long time to run this application.

2.4.2. FCM Clustering

FCM has been widely used and developed for biomedical image segmentation. The FCM, presented by Bezdek et al. [30], describes the fuzzy segmentation for the pixels by calculating the membership values for fuzzy sets. The objective function minimizes the FCM for the fuzzy C -partition U and many cluster centers V defined as per Equation (7).

$$\text{Minimize } J_{\text{FCM}}(U, V) = \sum_{k=1}^N \sum_{i=1}^C (u_{ik})^m d^2(x_k, v_i) \quad (7)$$

where $X = \{x_1, x_2, \dots, x_N\} \subseteq \mathbb{R}^p$, U is the fuzzy C -partition, C is the number of cluster centers (In this work, an experimental variation from two to six is used), m represents the weight exponentially, $d^2(x_k, v_i)$ is a measuring distance between pixel x_k and cluster center of v_i (usually the Euclidean distance), N is the number of the feature vector in an image, u_{ik} is the membership values for fuzzy sets of pixel k in cluster i , v_i is the cluster center for subset i in feature space $(U, V) = \sum_{k=1}^N \sum_{i=1}^C (u_{ik})^m d^2$ and is an overall weight sum of generalized A-error. On solving the objective function $(U$ and $V)$, minimizing an objective

function will be obtained so that the subject is subject to the constraint in Equation (8) is satisfying.

$$\sum_{i=1}^C u_{ik} = 1, \forall k \in \{1, \dots, N\} \quad (8)$$

with these observations, we can decompose of J_{FCM} be (u_{ik}, c_k) . The necessary conditions for (u_{ik}, c_k) are calculated using Equation (9) and Equation (10), respectively.

$$u_{ik} = \frac{1}{\sum_{j=1}^C \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \quad (9)$$

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^m x_k}{\sum_{k=1}^N (u_{ik})^m} \quad (10)$$

The algorithm of FCM is composed as seen in the following steps: (1) Fix the initialize the fuzzy membership matrix c , $2 \leq c \leq N$; Fix m , $1 < m < \infty$; (2) Initialize step of the fuzzy C-partition $U(0)$; (3) Assume distribution to be normal steps as $b = 1, 2, \dots, N$; (4) Calculate the c fuzzy cluster centers v_i as per Equation (10); (5) Update $U^{(k)}, U^{(k+1)}$ by calculate the new calculated u_{ik} as per Equation (9); (6) Compute the change in the membership values using an appropriate norm; if $U^{(k)} - U^{(k+1)} < T$, STOP; otherwise, set $k = k + 1$ and return to steps 4 and (7) A termination criterion is needed to stop the iterative process when the objective function is less than (T) and otherwise return to step 2.

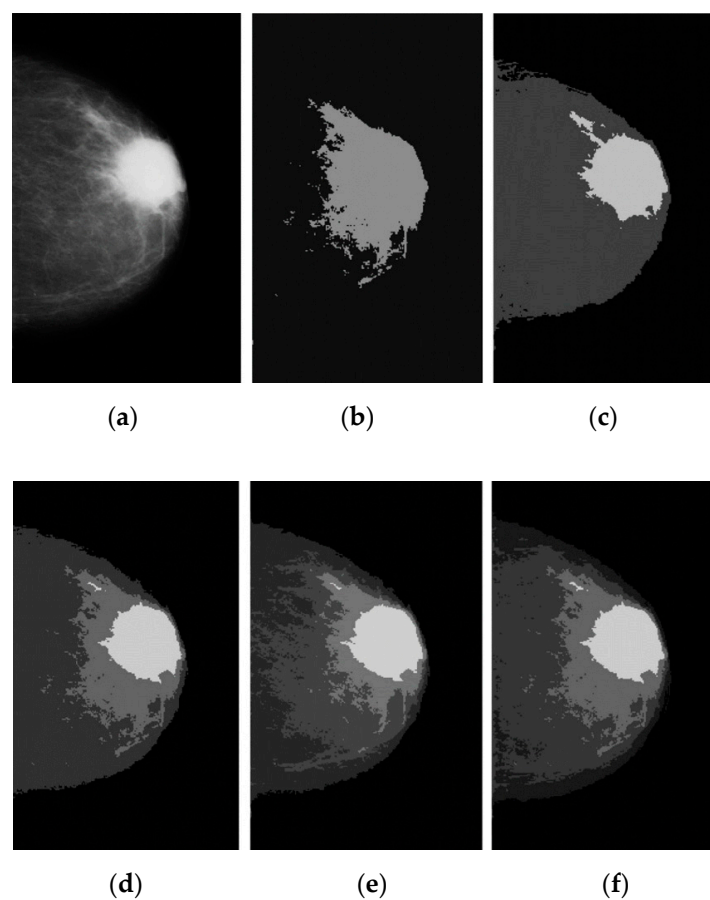


Figure 4. An original mammogram image is shown in a KM segmented version: (a) pre-processed image, (b) the first image segmented with $k = 2$, (c) good quality segmented with $k = 3$, (d) another example of segmenting breast cancer with $k = 4$, (e) segmenting cancer with $k = 5$, and (f) region segmented as candidate cancer area with $k = 6$ after 110 iterations.

The images segmented as test images for applying the FCM method with the various clusters are illustrated in Figure 5b–e. Figure 5b displays the three clusters designated using $k = 3$ (57 iterations) and the segmented image. Figure 5c shows the four clusters formed by selecting $k = 4$ (62 iterations). Figure 5d displays the five clusters formed by selecting $k = 5$ (100 iterations). Figure 5e depicts the six clusters formed by selecting $k = 6$. Figure 5f shows the seven clusters formed by selecting $k = 7$. This study concludes empirically that $k = 4$ is a “good” value. This value forms further advantages of simplifying the update equations and can therefore speed up computer implementations of the method.

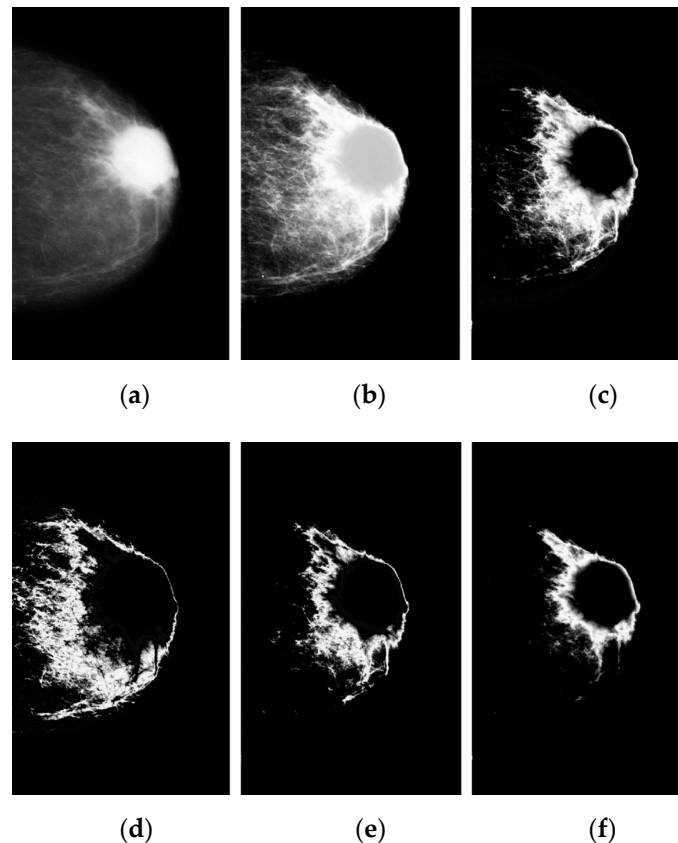


Figure 5. The segmentation of breast cancer using the FCM method: (a) preprocessed image on the left, (b–f) results of breast cancer regions segmented with $k = 2, 3, 4, 5$, and 6 after 100 iterations.

2.4.3. FKM Clustering

FKM clustering [31] is one method that provides a better result than KM for overlapping datasets. This method is utilized for the dataset in which the data points are between the centers. The technique works on the objective to minimize by as Equation (11).

$$F(X, C) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \|x_i - c_j\|^2 \quad (11)$$

where m is the fuzzifier and any real number greater than 1, u_{ij} the degree of membership in which data points belong to each cluster x_i to the cluster center c_j with the limitation that $u_{ij} \geq 0$, $\sum_{j=1}^k u_{ij} = 1 \forall i$ and $\|x_i - c_j\|^2$ is the distance between the data points to calculate x_i and c_j , is an assign of the distance of the n data point to the closest cluster by its sum of squared distance from the cluster’s centroid. KM is a popular method that has been used in many segmentation domains. It is a good candidate for an extension to work with fuzzy feature vectors. The iterative approach of optimizing the objective function $F(X, C)$ by updating the degree of participation of the data point x_i to the cluster centroids

c_j and the cluster center c_j results within the clustering of the data is calculated based on Equations (12) and (13), respectively.

$$u_{ij} = \sum_{j=1}^k \|x_i - c_j\|^{\frac{-2}{m-1}} / \sum_{j=1}^k \left(\|x_i - c_j\|^{\frac{-2}{m-1}} \right) \quad (12)$$

$$c_j = \sum_{i=1}^n u_{ij}^m \cdot x_i / \sum_{i=1}^n u_{ij}^m \quad (13)$$

As the value of m increases, the algorithm becomes fuzzy. The proposed FKM is tested on the image by setting the number of clusters to $K = 2$. The resultant images are displayed in Figure 6b. We compared these methods to those FKM clustering methods' capability to provide "good" classification performance as the number of clusters increases. Figure 6c–f show the breast cancer classification performance as set in the number of clusters $K = 3, 4, 5$, and 6 , respectively. Figure 6d shows "good" breast cancer classification performance with a number of clusters $k = 4$.

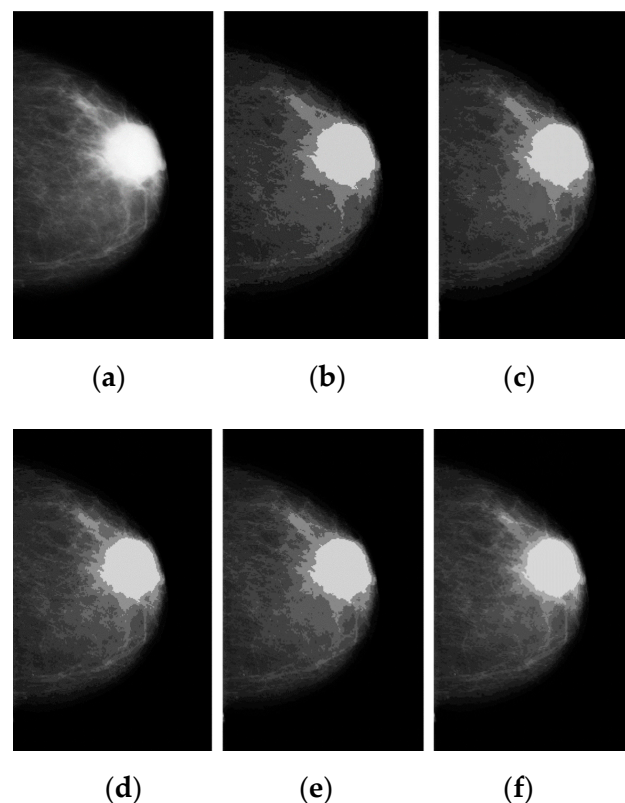


Figure 6. The segmentation of breast cancer using the FKM method: (a) preprocessed image on the left, (b–f) results of breast cancer segmentation on the gray level image with $k = 2, 3, 4, 5$, and 6 after 100 iterations.

2.4.4. Optimization Design with KM++CSO

KM++ is an algorithm to initialize the number of clusters k , which is given as an input to the KM method [32]. Since selecting the appropriate value of K prior is difficult, KM++ provides a method for determining the optimum value of K before proceeding to cluster the data. Impressively, this method offers theoretical guarantees about the output clustering quality. KM++ initialization method chooses the first center c_1 selected uniformly at random from the dataset X , but each successive center c_j is chosen with probability by Equation (14).

$$\frac{D^2(c_j)}{\sum_{x \in X} D^2(x_i)} \quad (14)$$

Here $D(c_j)$ denotes the minimum distance from a data pixel x to the closest point between c_j and all previously chosen centers. Then, the essential component of KM is presented as follows: (1) A starting with center c_1 , a point is chosen uniformly at random from the dataset X ; (2) Take another center c_{i1} , $c_{i1} = x \in X$ with probability $D^2(c_j)/\sum_{x \in X} D^2(x_i)$ where $D^2(c_j)$ is the shortest distance between the data point x to the closest center which is chosen; (3) Repeat step 1 to 2 until the values of k center are chosen, and (4) Perform the KM using the k center as the initial center. In step 2, “ D^2 weighting” is the applied weighting. As a result of the segmentation error of KM, this seeding algorithm gives out considerable improvements. The clustering performed by KM++ does not provide any guarantees that it theoretically performs the best clustering. So, we can use an improved method with CSO [33] to find a better solution for initial centroids. The Cuckoo’s reproductive behavior inspired it as a particular species by laying its eggs in the nests of other host birds or species. We can summarize more details for KM++CSO in the pseudocode (Algorithm 1).

Algorithm 1: KM++CSO.

- (1) Select an appropriate value for K which is the number of clusters and obtain a number of data points.
 - (2) Randomly initialize of K cluster center (c_1, \dots, c_k)
 - (3) Keep repeating steps 4 and 5 until the algorithm convergence to a certain value or the centroids do not change
 - (4) In each generation, all data points (x_i) are now assigned the point to the cluster with the closest centroids (c_1, \dots, c_k)
 - (5) Assign each data point to the closest cluster centroids $j = 1, 2, \dots, k$
 - (6) End
 - (7) A solution of the objective function (CSO); $f(x)$, $x = (x_1, x_2, \dots, x_k)$
 - (8) Populate the initial population of n host nests x_i , which is generated at random
 - (9) While $(t < \text{Max Generation})$ or (Stopping criterion) do
 - (10) For each loop Cuckoo i do
 - (11) Get a Cuckoo randomly (named i) and replaced it by performing intensity values
 - (12) Evaluate its quality/fitness F_i
 - (13) Choose randomly a nest from n (named j)
 - (14) If $(F_i > F_j)$ minimization them
 - (15) Replace j with i (generate a new solution)
 - (16) End if
 - (17) The number of host nests is fixed with a probability (P_a) of the eggs with low fitness being abandoned and rebuilt other nests somewhere
 - (18) Keep the best nests with a higher fitness value
 - (19) Rank the solution and find the best current one
 - (20) Pass the current best solutions to the next generation
 - (21) End while
-

We can summarize the KM++CSO: (1) Domain space: Initial centroids by CSO generated, and the used KM++ start with generated centroids. The result of clustering after classification is measured by a Peak–Signal–to–Noise Ratio (PSNR) between the input image and output image; (2) Improved optimization: the closet cluster centroid using CSO and KM++ finds the optimal initial image centroids by CSO, so the automatic detection is done using a clustering result and (3) An acquisition function: although the initial parameter selections take extra time, KM++CSO converges very fast for the coarse segmentation stage and lowers the computation time. KM++ determines a set of k clusters with parameters set to 2, 3, 4, 5, and 6 (100 iterations). Then the best solutions to optimize the results obtained by the Cuckoo search algorithm using the fraction $P_a = 0.2$ is used to compare the capability in segmentation performance as the number of clusters increases. The resultant images are displayed in Figure 7b–f, respectively. Figure 7d shows “good” segmentation performance in the number of clusters (k) is set to 4.

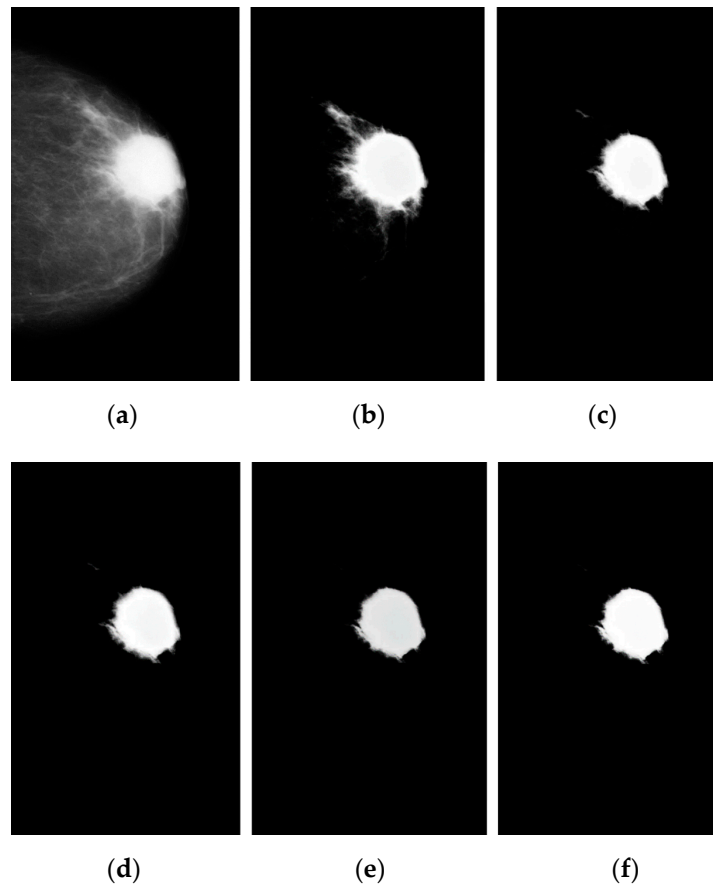


Figure 7. The segmentation of breast cancer using KM++CSO, (a) preprocessed image on the left, (b–f) the results of breast cancer segmentation with $k = 2, 3, 4, 5$, and 6 after 110 iterations.

2.4.5. Binary Image Detection

In this stage, binary image segmentation using the Gaussian OTSU's threshold [34] is used. Given an image: $P(i)$ is the normalized frequency of i , and is calculated according to Equation (15).

$$P(i) = \frac{\text{number}\{(r, c) | \text{image}(r, c) = i\}}{(R, C)} \quad (15)$$

Here r represents an index for rows, and c is a column of the image. R is the number of rows, and C is the columns of the image respectively. The value calculates the “good” of the threshold at value T^* by minimizing the within-class variance (weighted) and the between-class variance, which is the sum of the weighted variances of the two classes. If this minimum is not unique, obtain T^* by averaging the T value corresponding to the various minimized segments. This threshold is determined by minimizing within-class intensity variance at any threshold, where T is calculated by Equation (16).

$$\sigma_w^2 = w_{mb}(T) * \sigma_{mb}^2(T) + w_{bc}(T) * \sigma_{bc}^2(T) \quad (16)$$

Here $w_{mb}(T)$ and $w_{bc}(T)$ are the probability of the number of mammogram background and breast cancer pixels in each class for this threshold T and variance of the two classes T_0 with a set of intensity levels from 0 to T , $\sigma_{mb}^2(T)$ and $\sigma_{bc}^2(T)$ represent the variance of color values and variance of class T_1 with a set of intensity level from $T + 1$ to L , or $T_1 = \{T, T+, \dots, L-, L\}$ and σ_w^2 is the weighted sum of group variances. Based entirely on computations performed on the histogram (1–D) of the image, the normalized histogram

of pixels that will be detected as mammogram background (w_{mb}) and breast cancer (w_{bc}) pixels, is given by Equation (17).

$$w_{mb}(T) = \sum_{i=1}^T P(i), w_{bc}(T) = \sum_{i=T+1}^L P(i) \quad (17)$$

The probability of pixel intensity (P) is computed for each pixel value in two separated clusters (namely background and breast cancer pixels) using the cluster probability functions expressed as per Equation (18).

$$\mu_{mb} = \sum_{i=1}^T P(i)/W_{mb}, \mu_{bc} = \sum_{i=T+1}^L P(i)/W_{bc} \quad (18)$$

where L is the number of pixels in a grayscale image containing information between 1 and L $\{1, 2, 3, \dots, L\}$ and $\sum_{i=T+1}^L P(i)$ is the sum of $P(i)$, where values between $i = T + 1$ and L $\{i = T + 1, T + 2, T + 3, \dots, L\}$. The class variance of the background and breast cancer pixels is given by Equations (19) and (20).

$$\sigma_{mb}^2(T) = \frac{1}{w_{mb}(T)} \sum_{i=1}^T (i - \mu_{mb})^2 * P(i), \quad (19)$$

$$\sigma_{bc}^2(T) = \frac{1}{w_{bc}(T)} \sum_{i=T+1}^L (i - \mu_{bc})^2 * P(i) \quad (20)$$

The between-class variance is defined as per Equation (21).

$$\begin{aligned} \sigma_{mb}^2 &= \sigma^2 - \sigma_w^2 \\ \sigma_{mb}^2 &= w_{mb} * \sigma_{mb}^2 + w_{bc} * \sigma_{bc}^2 + w_{mb} * (\mu_{mb} - \mu)^2 + w_{bc} * (\mu_{bc} - \mu)^2 \\ \sigma_{mb}^2 &= w_{mb} * w_{bc} (\mu_{mb} - \mu_{bc})^2 \end{aligned} \quad (21)$$

The procedure for determining the appropriate threshold value (T) is related to the maximum variance between groups. According to Equation (17), if $(W_{mb} = \sum_{i=1}^T P(i) = 0.138)$ and sets the probability of pixels in the background image $W_{mb} = \{0.138, 0.291, 0.268\}$ while if $(W_{bc} = \sum_{i=T+1}^L P(i) = 0.638)$ and sets the probability of pixels in the background image $W_{bc} = \{0.638, 0.734, 0.725\}$. According to Equation (18), if $\mu_{mb} = \{0.138, 0.121, 0.110\}$ and $\mu_{bc} = \{0.710, 0.821, 2.823\}$, the average background and breast cancer pixels can be calculated by $\mu_{mb} = \sum_{i=1}^T P(i)/W_{mb}$ and $\mu_{bc} = \sum_{i=T+1}^L P(i)/W_{bc}$, respectively. Therefore, the maximum variance of each group was calculated using Equation (21), which is $\sigma_{mb}^2 = \{0.028, 0.212, 1.427\}$. If the maximum variance between groups occurs in three iterations, then the maximum variance is $0.028 + 0.212 + 1.427 = 0.556$. Hence, in this binary detection process, the optimal threshold setting is 0.556, with 1 for the breast cancer pixel and 0 for the background pixel. The highlighted breast cancer regions were white, while the highlighted mammogram background was black (see Figure 8).

2.4.6. Improved Binary Image Using Mathematical Dilation Operators

Besides the binary detection applied in the previous section, morphological dilation operators are essential for more accurate detection [35]. The dilation is a procedure that enables the pixels to grow along these lines and fill the little holes, i.e., similar to increasing the thickness of the areas and associating the separate points. This operation is mainly performed post-segmenting of the characteristics of regions in mammogram images and adopts the logical processes by using the structure element involving the growing mammogram pixels in a binary image. The dilation of A represented by B is calculated in Equation (22).

$$A \oplus B = (z | (\hat{B})_z \cap A \neq \emptyset) \quad (22)$$

where B represents the structuring element, is a reflection of set B and \emptyset is the empty set. Furthermore, the detailed explanation of the method used in this paper is as follows; (1) Input a binary image by thresholding (Figure 9a) and then the breast cancer region in the image that is filled using mathematical dilation operators; (2) Assign mathematical dilation operators with structuring elements is set to 7 using a 3×3 mask. Dilation operation has effects of smoothing the contour of breast cancer, eliminating thin protrusions, and breaking narrow isthmuses (see Figure 9b,c) Save the resulting image after applying step 2, and the pixels of the breast cancer are preserved, and all small pixels removed after dilation by using thresholding methods [36] again.

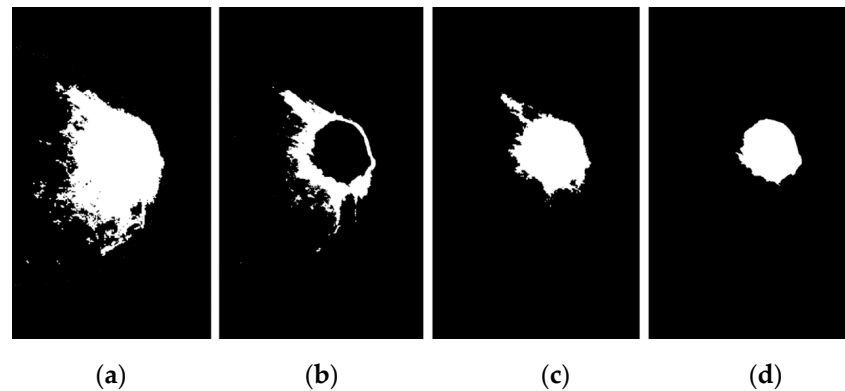


Figure 8. Breast cancer image and results of four segmentation methods, then, obtained binary image based on OTSU's threshold with $T = 0.556$ gives the best result, (a) KM method, (b) FCM method, (c) FKM method, (d) KM++CSO method.

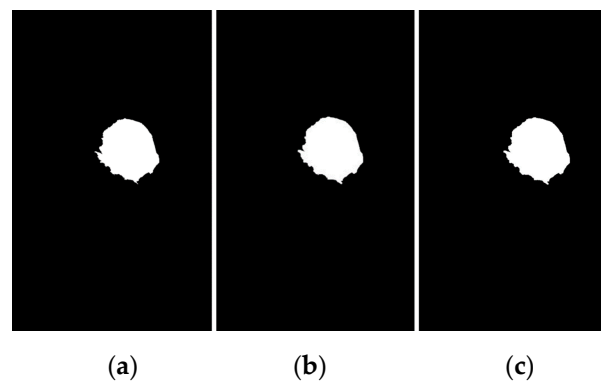


Figure 9. Segmenting and removing small pixels by morphological modeling based on a dilation operator, (a) breast cancer segmented using KM++CSO and OTSU's threshold method, (b) image reconstructed with structure element of 7 to disconnect component pixels of the binary image, (c) regional maxima image preserved and all small objects removed using global thresholding.

This is an iterative algorithm for automatic estimation of threshold T given as follows: (1) An initial threshold from the image intensity histogram values T and partitions the image into two classes, $G1$ and $G2$, using the threshold, (2) Calculate the mean intensity values for $G1$, $G2 \rightarrow \text{mean}(G1)$, and $\text{mean}(G2)$, (3) Select a new threshold value using $T = 0.55 \times (\text{mean}(G1) + \text{mean}(G2))$, (4) Repeat steps 1–3 until the mean intensity values $G1$, and $G2$ in successive iterations do not change ΔT . In this step, start means gray level and $\Delta T = 0$. The algorithm results in $T = 0.67$ after five iterations, so $T = 0.67$ is used. The breast cancer region is shown in Figure 9c.

2.4.7. Edge Detection

As the segmented image in Section 2.4.6, we perform some experiments to compare with various edge detectors such as the Sobel edge operator [37], Robert edge operator [38], Prewitt edge operator [39], and Canny edge operator [39], for the segmentation of the boundaries of breast cancer on Figure 9c. Figure 10 shows the result in this case. According to Figure 10a, the Sobel edge detector will be more accurate in detecting the boundaries of breast cancer and more flexible to unseen image datasets.

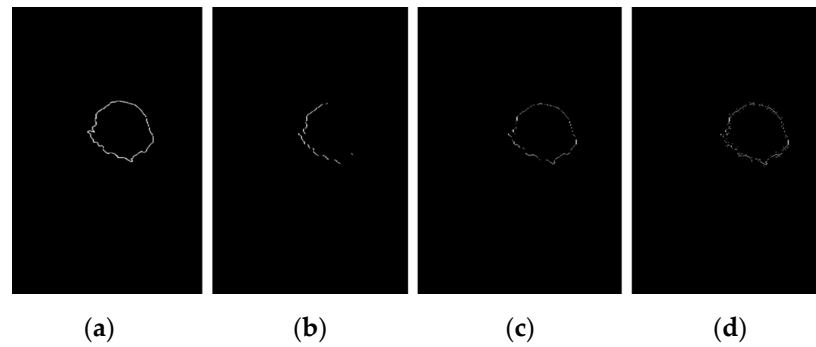


Figure 10. The mammogram image is segmented with various edge operator, (a) Sobel edge operator, (b) Robert edge operator, (c) Prewitt edge operator, (d) Canny edge operator. Canny edge segmentation results are different from the other methods. Prewitt and Robert's edge operators produce almost the same edge segmentation. However, the Sobel edge segmentation result is superior by far to the other methods.

3. Results

The outcomes of the proposed method by using KM, FCM, FKM, and KM++CSO show breast cancer detection through mammogram images from three different datasets. The proposed method was evaluated by SEN, SPEC, and ACC [40] and the Jaccard index score [41]. SEN, SPEC, and ACC were used for standard quality measurement of the segmented output image. SEN is the percentage of breast cancer pixels segmented as breast cancer pixels, and SPEC is the percentage of background detected as background pixels. ACC represents the percentage of the overall pixels that are segmented correctly. These indexes are defined by Equations (23)–(25).

$$\text{Sensitivity (SEN)} = \frac{TP}{TP + FN} * 100 \quad (23)$$

$$\text{Specificity (SPEC)} = \frac{TN}{FP + TN} * 100 \quad (24)$$

$$\text{Accuracy (ACC)} = \frac{TN + TP}{(TP + TN) + (FP + FN)} * 100 \quad (25)$$

where TP (True Positive) is the number of breast cancer pixels that are detected as abnormal, FN (False Negative) is the number of breast cancer pixels that are detected as normal, TN (True Negative) is the number of backgrounds that are detected as normal, and FP (False Positive) is the number of backgrounds that are detected as abnormal. Another metric to evaluate the breast cancer detection between the proposed method and manually detected with three experts was used for comparative analysis, which is widely used in the literal. Another performance aspect of the Jaccard index score [41] is defined as per Equation (26).

$$\text{Jaccard Index} = \frac{|ROI_{KM++CSA} \cap ROI_{GT}|}{|ROI_{KM++CSA} \cup ROI_{GT}|} \quad (26)$$

where $ROI_{KM++CSA}$ was the breast cancer region by using the proposed detection method, ROI_{GT} was the breast cancer region of the ground truth image and $|ROI_{KM++CSA}|$ is its cardinality.

3.1. Breast Cancer Detection Results

The proposed methods are implemented on MATLAB version 2020a (The MathWorks Inc.) and run on a 4.00 GHz Intel(R) Core (TM) i7–6700K CPU, 8GB (RAM) under the Microsoft Windows 10, 32-bit operating system. In Figure 11, we list images with four different detection methods and a number of clusters for every clustering method. Figure 11(a1) is the original image with signs of breast cancer. Figure 11(a2) shows the results of applying KM to detect breast cancer. Figure 11(a3) shows the effect of using OTSU's threshold on the detected image of Figure 11(a2). Figure 11(a4) is the result of running a mathematical morphology on the image (Figure 11(a3)) and applying global thresholding. The detection results revealed that the detected images generated by morphological dilation are the best way to fill the gaps in the segmented stage. Figure 11(a5) shows the result of applying the Sobel edge operator. We tested the proposed method with some images in which SEN, SPEC, ACC, and Jaccard index score indices evaluated the method's effectiveness. Selecting a range of values for a number of clusters k would lead to a not-so-good detection result. If the value for the number of clusters is decreased, the colors of the neighborhood may be confused. Moreover, KM requests that the user indicate the number of clusters k before the detection commences. The number of clusters the user indicates must compare with the number of gray colors. It is not essential to have prior knowledge of the number of gray colors contained by the image since there is an arrangement made for re-inputting the number of clusters as soon as KM gets to the end of the clusters, which indicates that it stops. In this experiment, the proposed method results with the test image can be seen in Figure 11(b1–b5). Figure 11(b1) is the original image of breast cancer. Figure 11(b2) is the result of applying FCM with four clusters performed using $k = 4$. Figure 11(b3) shows the binary detection results for Figure 11(b2) using the OTSU's threshold. The result of applying mathematical morphology based on dilation operators and running global thresholding on the dilated image is shown in Figure 11(b4). Figure 11(b5) displays the result of applying the Sobel edge operator. FCM iterates are based on the number of clusters and come over on the considered image. Unlike KM, the FCM will return the number of clusters after the detection. Next, FKM was tested on a mammogram image, the gray level image, and their results are shown in Figure 11(c1–c5). The image shown in Figure 11(c2) displays the four clusters performed by choosing $k = 4$ and the detected image. Figure 11(c3) shows the result of using the OTSU's threshold method. Figure 11(c4) shows the results of applying mathematical morphology based on dilation operators and running global thresholding on the image. Figure 11(c5) displays the result of using the Sobel edge operator. FKM could be an algorithm produced from FCM and KM but carries more of an FKM property than KM. Finally, KM++CSO is tested on mammogram images, and its performance is compared to that of other machine learning like KM, FCM, and FKM. The first test image Figure 11(d1), is displayed to compare the experimental results of breast cancer. Figure 11(d2) shows the four clusters performed using $k = 4$ and the detected image. Figure 11(d3) is applied to detect breast cancer regions using OTSU's method with threshold values of 0 to 255 and set to 128. The breast cancer is dilated from its background using morphological-based dilation operators, and the results of applying global thresholding are shown in Figure 11(d4). The Sobel edge is again used to detect edges in the regions. By observing Figure 11(d5), we can say that the Sobel edge detects the edges. KM++CSO works on gray-level images like FCM and creates the same number of iterations as in KM and FKM. Time is taken to detect based on the tested image using KM++CSO to be faster than other clustering methods, whereas, in a few cases, KM also shows up to be faster than the original FCM and FKM. Whereas KM++CSO and KM compete in time, FCM has been modified to create the same number of iterations as FKM with a faster operation time. That is, FKM is faster than FCM. The conflict in time between FCM and FKM and KM and KM++CSO is expected to account for

the properties of the image under consideration and the efficiency of the machine on which the algorithms are tested.

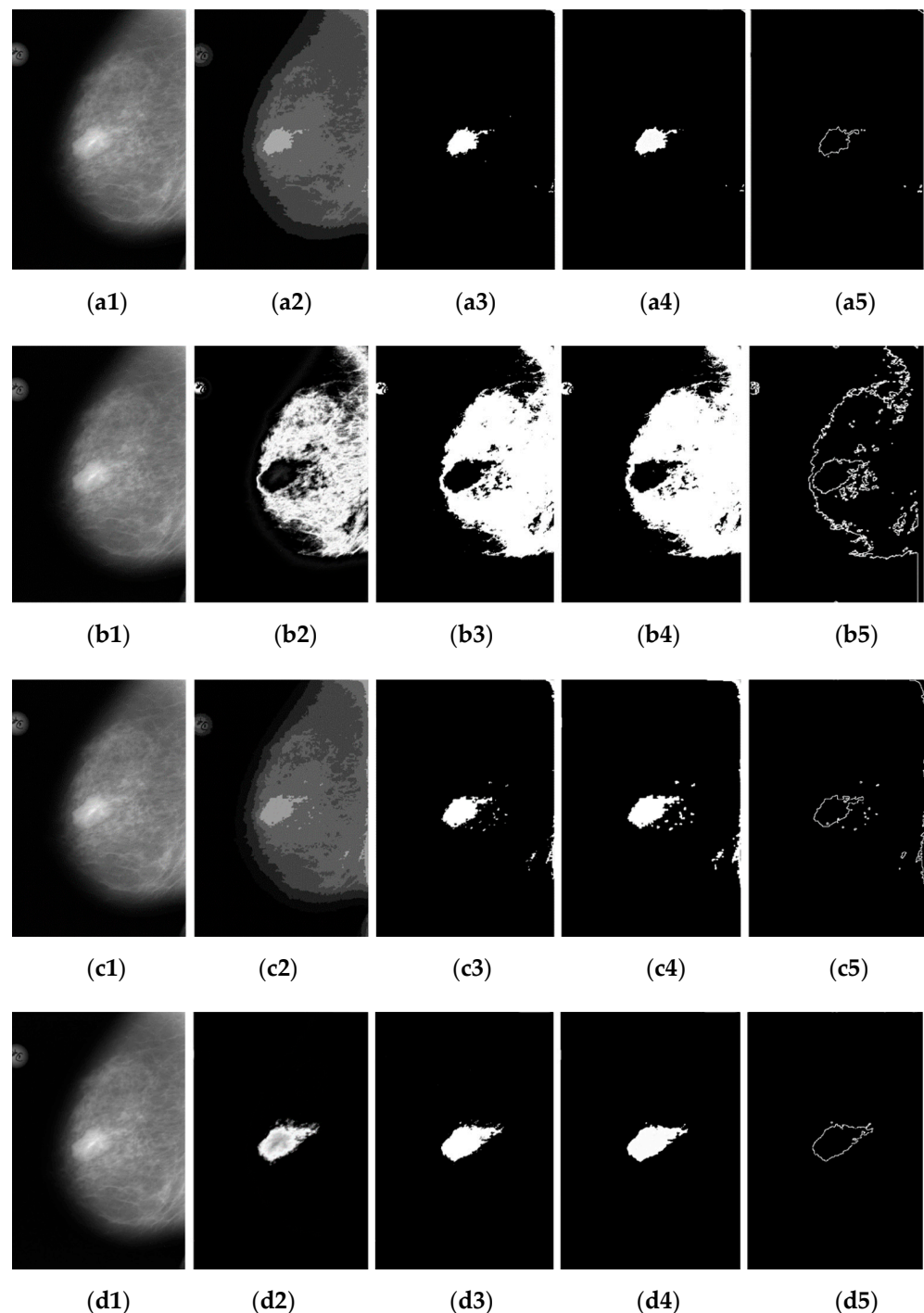


Figure 11. Breast cancer image and result of four detection method with OTSU's threshold, mathematical modeling and Sobel edge operator, (a1–a5) KM, (b1–b5) FCM, (c1–c5) FKM, (d1–d5) KM++CSO.

In terms of average accuracy, the number of iterations is considered. The more iterations, the more accuracy. The iteration that FCM and FKM can perform generally depends on the number of gray images contained by an image, which limits its iterative ability, unlike that of KM++CSO, which segments based on the number of clusters assigned in an image. As a result, FCM and FKM are less accurate than the KM and KM++CSO. The performance of the proposed methods on three open-access datasets is shown in Table 2.

They have been tested in four unsupervised machine learnings: KM, FCM, FKM, and MK++CSO. Typically, the training dataset is generally for the samples required; at present, the methods are built with a series of image processing methods, which might still need a training set for a constant learning rate and an optimization for all parameters. From the data analysis, the FCM used on 123 images of BCDR successfully obtained the SEN, SPEC, and ACC rates of 61.29%, 61.19%, and 61.20%, respectively. However, using KM++CSO, the SEN, SPEC, and ACC rate increased to 96.90%, 96.16%, and 96.42%, respectively. Therefore, the accuracy has been increased to 35.22%. For FCM used on 123 breast cancer images of Mini-MIAS, the algorithm successfully obtained the SEN, SPEC, and ACC rates of 60.67%, 60.97%, and 60.78%. Then, KM++CSO achieved SEN, SPEC, and ACC rates of 96.78%, 97.10%, and 96.92%, respectively. Therefore, the rate of accuracy has been increased to 36.14%. Also, 2620 images have been used to evaluate and compare the performance of the fusion methods. Similarly, for DDSM, we obtain SEN, SPEC, and ACC. FCM received the SEN, SPEC, and ACC rates of 67.22%, 69.12%, and 61.10%. Then, KM++CSO achieved SEN, SPEC, and ACC rates of 95.68%, 95.10%, and 95.49%, respectively. Therefore, the accuracy has been increased to 26.39%. For this reason, the detection method has achieved perfect detection results. Overall, the proposed method demonstrated that the detection results are good quality and suitable for the breast cancer detection region from the mammogram image. However, comparing the proposed methods directly with other methods is difficult because of the difference in the number of images, differences in the dataset, and evaluation performance of detection.

Table 2. Comparison between the proposed segmentation method in terms of SEN, SPEC, ACC, and Jaccard index score.

Methods	Datasets	Percentage			
		SEN (%)	SPEC (%)	ACC (%)	Jaccard Index (%)
KM	DDSM	88.38	89.10	88.89	80.45 (2108/2620)
FCM		67.22	69.12	69.10	70.95 (1859/2620)
FKM		90.15	90.18	90.16	83.55 (2189/2620)
KM++CSO		95.68	95.10	95.49	87.37 (2289/2620)
KM	Mini-MIAS	90.10	89.78	89.92	85.13 (275/323)
FCM		60.67	60.97	60.78	64.08 (207/323)
FKM		91.12	91.20	91.14	86.99 (281/323)
KM++CSO		96.78	97.10	96.92	89.47 (289/323)
KM	BCDR	89.96	89.89	89.92	78.86 (97/123)
FCM		61.29	61.19	61.20	65.85 (81/123)
FKM		91.40	91.56	91.48	88.62 (109/123)
KM++CSO		96.90	96.16	96.42	91.05 (112/123)

3.2. Manual Detection Versus Measurement

In this section, the proposed detection method has been evaluated of the breast cancer region by three experts versus measurements automatically detected by the proposed methods. The correlation between the area of breast cancer detected and the measures created by three experts is calculated as per Equation (26). For BCDR, this paper used 123 images based on KM++CSO, and we obtained 112 images successfully, while 11 images were misdiagnosed. Therefore, the proposed algorithm achieved a Jaccard index score of 91.05%. For Mini-MIAS, the proposed method accurately detected 289 images out of 323. Therefore, it is shown that 289 images with a Jaccard index score of 89.47% are detected correctly, while 34 images with a score of 10.53% are detected incorrectly. Moreover, the method accurately detected 2289 images out of 2620 images (87.37% Jaccard index score) in DDSM, while 331 images with a score of 12.63% were detected incorrectly. Figure 12 illustrates the segmentation results for mammogram images randomly selected from BCDR, Mini-MIAS, and DDSM databases. The first segmentation of Figure 12a is demonstrated as the best

breast cancer segmentation, which obtained a 91.05% Jaccard Index score on the BCDR database. The segmentation results of Figure 12b correctly depict the breast cancer region, which received an 89.47% Jaccard Index score on the Mini-MIAS. The segmentation results of Figure 12c correctly display the breast cancer region, which obtained an 87.37% Jaccard Index score on the DDSM. Based on the experimental results of our proposed method, its suitability for breast cancer segmentation has been successfully verified. Moreover, the proposed method outperformed breast cancer segmentation compared with BCDR, Mini-MIAS, and DDSM datasets. However, the proposed method obtained lower results by SEN, SPEC, ACC, and Jaccard Index scores in the DDSM for breast cancer segmentation.

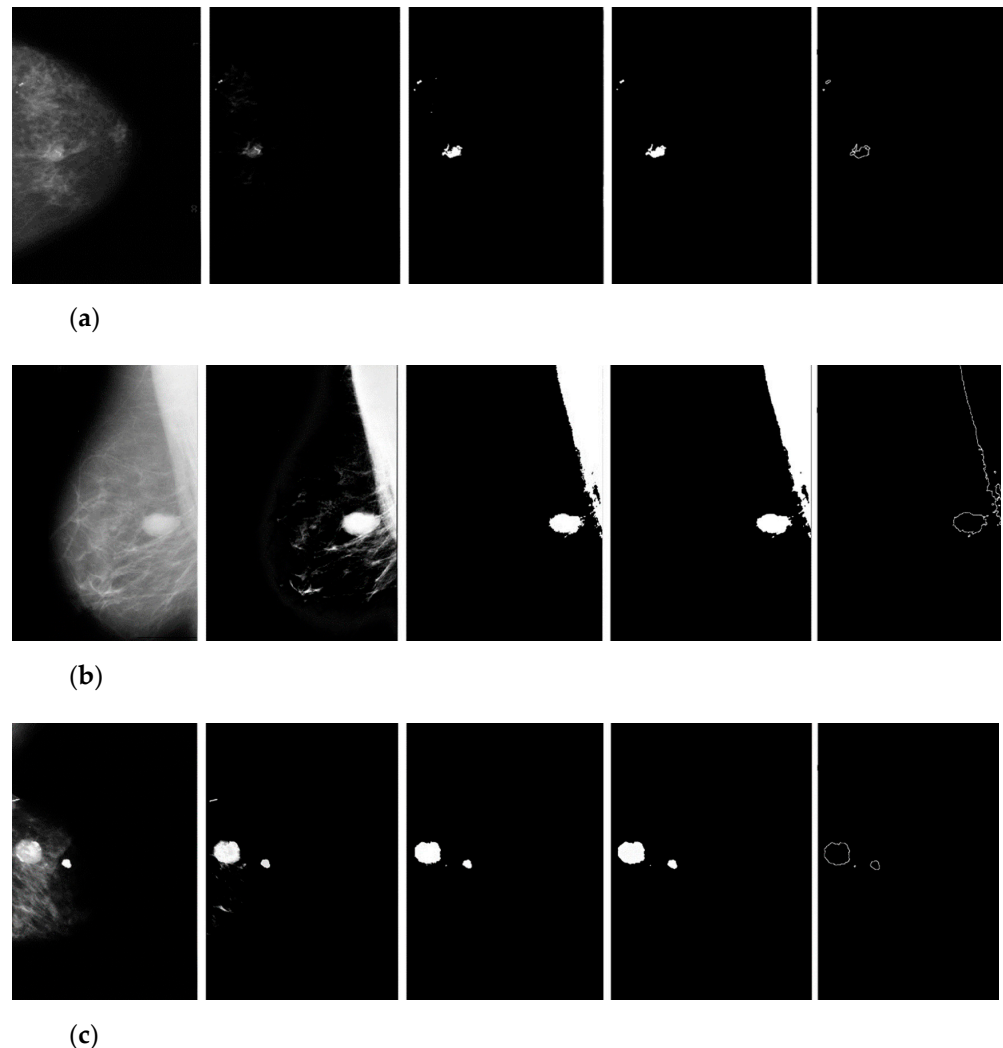


Figure 12. Examples of results of breast cancer segmentation from different databases: (a) BCDR, (b) Mini-MIAS, (c) DDSM.

4. Discussion

The diagnosis system of artificial intelligence in medicine has increased in recent years, which may be seen in the number of published articles [2–14]. The KM++ based on the CSO method for mammogram evaluation used within this work presented high sensitivity, specificity, and accuracy for the segmentation of breast cancer. The following contributions to the research community were provided in this article:

(1) We provided an overview of the breast cancer segmentation methods on three publicly available datasets, optimizing the parameters of KM++CSO for each dataset. With KM++CSO, the best SEN, SPEC, ACC, and Jaccard index scores in detecting breast cancer

using mammogram images among the four methods are 96.90%, 96.16%, 96.42%, and 91.05%, respectively. A description of the experimental results is provided in Table 2.

(2) When it comes to the detection of healthy and breast cancer-infected mammogram images, the time taken to segment based on the tested image with a KM++CSO appears to be faster than the KM, FCM and FKM whereas, in a few cases, KM also shows up to be faster than FCM and FKM. The average execution time required for KM, FCM, and FKM to get the optimal solution with three clusters is 7, 12, and 9 seconds per image, respectively, using the computer specifications described earlier. The average time for KM++CSO for the same cluster with the same clusters and iterations is 4 seconds per image. The proposition of the new method and its performance are shown in Section 3.1.

(3) It is also shown that the pre-processing methods have improved the accuracy and enhanced the coarse detection performance of the KM++CSO, including OTSU's thresholding and mathematical morphology. This paper indicates that by cutting-edge performance in KM++ and applying CSO, it is possible to achieve a higher level of performance, improving breast cancer detection accuracy.

(4) A new method for accurately detecting breast cancer was proposed based on KM++CSO in mammogram images. The technique allows for more accurate detection of breast cancer by tracking the lesions on the mammogram surface. It can also aid the experts in the manual classification and detection of breast cancer. The proposed methods can produce good detection with no effort in terms of coarse to fine detection of breast cancer in a mammogram image.

The results from the mammogram image analysis revealed the possibility of the successful detection of breast cancer. The results from the image segmentation confirmed differences in the position, shape, and cancer features between the healthy image and people with breast cancer. However, there are limitations to this work. The proposed method is evaluated on small datasets, although it provides data for appropriate segmentation analysis. The second limitation is setting the multiple algorithms to compare with the ground truth image as the basis of three expert reports. This process requires more complexity in time. However, for future work, the execution time in the detection process is a region of interest. Sometimes, the methodology, which does not require fixing the number of clusters before detecting, may be used for coarse detection.

5. Conclusions

Within this work's limitations, we can conclude that the tested KM++CSO can be helpful for an initial evaluation of breast cancer screening for medical diagnostics. Moreover, the result of the proposed method outperformed breast cancer segmentation when compared with the existing approach on BCDR, Mini-MIAS, and DDSM datasets.

Funding: This research was financially supported by Mahasarakham University, Thailand (Grant No: 6516006-2565).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research project was financially supported by Mahasarakham University, Thailand. The authors' thankfulness also goes to the following specialist experts: Ekkarat Pothiruk, KhonKaen Hospital, Thailand, WiranutSuttisa, Kantharawichai Hospital, Thailand, Sakrit Moksiri, Borabue Hospital, Thailand, for their kind provisions of the breast cancer segmentation for this study.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Siegel, L.R.; Kimberly, M.D.; Hannah, F.E.; Jemal, A. Cancer statistics, CA: A cancer. *J. Clin.* **2021**, *71*, 7–33.
2. Karabatak, M. A new classifier for breast cancer detection based on Naïve Bayesian. *Measurement* **2015**, *72*, 32–36. [[CrossRef](#)]

3. Pak, F.; Kanan, H.R.; Alikhassi, A. Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution. *Comput. Methods Programs Biomed.* **2015**, *122*, 89–107. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Sabu, R.S. Analysis of Mammographic Images for Early Detection of Breast Cancer Using Machine Learning Technique. Master's Thesis, Department of Biotechnology and Medical Engineering, National Institute of Technology, Orissa, India, 2015.
5. Yang, S.C. A robust approach for subject segmentation of medical Images: Illustration with mammograms and breast magnetic resonance images. *Comput. Electr. Eng.* **2017**, *62*, 151–165. [\[CrossRef\]](#)
6. Punita, S.; Amuthan, A.; Joseph, K.S. Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Comput. Inform. J.* **2018**, *3*, 348–358. [\[CrossRef\]](#)
7. Al-masni, M.A.; Park, J.M.; Gi, G.; Kim, T.Y.; Rivera, P. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput. Methods Programs Biomed.* **2018**, *157*, 85–94. [\[CrossRef\]](#)
8. Shen, T.; Gou, C.; Wang, J.; Wang, F.Y. Simultaneous segmentation and classification of mass region from mammograms using a mixed-supervision guided deep model. *IEEE Signal Process. Lett.* **2019**, *27*, 196–200. [\[CrossRef\]](#)
9. Punita, M.; Perumal, K. Hybrid segmentation and feature extraction approach to detect tumors based on fuzzy rough-in mammogram images. *Procedia Comput. Sci.* **2019**, *165*, 478–484. [\[CrossRef\]](#)
10. Zebari, D.A.; Zeebaree, D.Q.; Abdulazeez, A.M.; Haron, H.; Hamed, H.N.A. Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. *IEEE Access* **2020**, *8*, 203097–203116. [\[CrossRef\]](#)
11. Vaka, A.R.; Soni, B.; Reddy, K.S. Breast cancer detection by leveraging machine learning. *ICT Express* **2020**, *6*, 320–324. [\[CrossRef\]](#)
12. Azlan, N.A.N.; Lu, C.K.; Elamvazuthi, I.; Tang, T.B. Automatic detection of masses from mammographic images via artificial intelligence techniques. *IEEE Sens. J.* **2020**, *21*, 13094–13102. [\[CrossRef\]](#)
13. George, K.; Faziludeen, S.; Sankaran, P.; Joseph, K.P. Breast cancer detection from biopsy images using nucleus guided transfer learning and belief-based fusion. *Comput. Biol. Med.* **2020**, *124*, 103954. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Alanazi, S.A.; Kamruzzaman, M.M.; Sarker, M.N.I.; Alruwaili, M.; Alhwaiti, Y. Boosting breast cancer detection using convolutional neural network. *J. Healthc. Eng.* **2021**, *2021*, 5528622. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Mohammed, S.J.; Abbas, T. Comparison of a classifier performance testing methods: Support vector machine classifier on mammogram images classification. *J. Kufa Math. Comput.* **2019**, *6*, 8–12. [\[CrossRef\]](#)
16. Agnes, S.A.; Anitha, J.; Pandian, S.I.A.; Peter, J.D. Classification of mammogram images using multiscale all convolutional neural network (MA-CNN). *J. Med. Syst.* **2020**, *44*, 30. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Yektaei, H.; Manthouri, M.; Farivar, F. Diagnosis of breast cancer using multi scale convolutional neural network. *Biomed. Eng. Appl. Basis Commun.* **2019**, *31*, 1950034. [\[CrossRef\]](#)
18. Kaur, P.; Singh, G.; Kaur, P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Inform. Med. Unlocked* **2019**, *16*, 100151. [\[CrossRef\]](#)
19. Viswanath, V.H.; Guachi, L.G.; Thirumuruganandham, S.P. Breast cancer detection using image processing techniques and classification algorithms. *EasyChair* **2019**, *2101*, 1–11.
20. Jin, Y.; Zheng, Y. Medical image processing with deep learning: Mammogram classification and automatic lesion detection. *Comput. Sci. Med.* **2019**, 1–19. Available online: <https://nics.utk.edu/wp-content/uploads/sites/275/2021/08/19-Jin-final-report.pdf> (accessed on 14 November 2022).
21. Kayode, A.A.; Akande, N.O.; Adegun, A.A.; Adebiyi, M.O. An automated mammogram classification system using modified support vector machine. *Med. Devices (Auckl.)* **2019**, *12*, 275–284. [\[CrossRef\]](#)
22. Debele, T.G.; Gebreselasie, A.; Schwenker, F.; Amirian, M.; Ashenafi, D.Y. Classification of mammograms using texture and CNN based extracted features. *J. Biomater. Biomed. Eng.* **2019**, *42*, 79–97. [\[CrossRef\]](#)
23. Zhang, Y.; Chan, S.; Park, V.Y.; Chang, K.T.; Mehta, S. Automatic detection and segmentation of breast cancer on MRI using mask R-CNN trained on non-fat-sat images and tested on fat-sat images. *Acad. Radiol.* **2020**, *29*, S135–S144. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Gonzalez, R.C. *Digital Image Processing*, 4th ed.; Pearson India: Bangalore, India, 2018.
25. Bejnordi, B.E.; Litjens, G.; Timofeeva, N. Stain specific standardization of whole-slide histopathological images. *IEEE Trans. Med. Imaging* **2016**, *35*, 404–415. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Janowczyk, A.; Basavanahally, A.; Madabhushi, A. Stain normalization using sparse auto encoders (StaNoSA): Application to digital pathology. *Comput. Med. Imaging Graph.* **2017**, *57*, 50–61. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Thomas, G.; Tapia, D.F.; Pistorius, S. Histogram specification: A fast and flexible method to process digital images. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 1565–1578. [\[CrossRef\]](#)
28. Gao, Z. An adaptive median filtering of salt and pepper noise based on local pixel distribution. In Proceedings of the 2018 International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018), Chengdu, China, 30–31 October 2018; Atlantis Press: Paris, France, 2018; Volume 161, pp. 473–483.
29. Zheng, X.; Lei, Q.; Yao, R.; Gong, Y.; Yin, Q. Image segmentation based on adaptive k-means algorithm. *EURASIP J. Image Video Process.* **2018**, *2018*, 68. [\[CrossRef\]](#)
30. Bezdek, J.C.; Ehrlich, R. FCM-the fuzzy c-means clustering-algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [\[CrossRef\]](#)
31. Li, M.J.; Ng, M.K.; Cheung, Y.; Huang, J.Z. Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 1519–1534. [\[CrossRef\]](#)

32. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics Philadelphia, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
33. Yang, X.S.; Deb, S. Cuckoo Search via Lévy flights. In Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009; pp. 210–214.
34. Bangare, S.L.; Dubal, A.; Bangare, P.S.; Patil, S. Reviewing Otsu’s method for image thresholding. *Int. J. Appl. Eng. Res.* **2015**, *10*, 21777–21783. [[CrossRef](#)]
35. Chudasama, D.; Patel, T.A.; Joshi, S.; Prajapati, G. Image segmentation using morphological operations. *Int. J. Comput. Appl.* **2015**, *117*, 16–19. [[CrossRef](#)]
36. Jang, J.; Lee, S.; Hwang, H.J.; Baek, K. Global thresholding algorithm based on boundary selection. In Proceedings of the 2013 13th International Conference on Conference: Control, Automation and Systems (ICCAS), Gwangju, Korea, 20–23 October 2013; pp. 1–3.
37. Han, L.; Tian, Y.; Qi, Q. Research on edge detection algorithm based on improved Sobel operator. In Proceedings of the 2019 International Conference on Computer Science Communication and Network Security (CSCNS2019), Sanya, China, 22–23 December 2019; Volume 309, pp. 2–6.
38. Shrivakshan, G.T.; Chandramouli, C. A comparison of various edge detection techniques used in image processing. *Int. J. Comput. Sci.* **2012**, *9*, 269–276.
39. Ferhat, F.A.; Ait, L.; Kerdjadj, O.; Messaoudi, K.; Abdelwahhab, B. Implementation of sobel, prewitt, roberts edge detection on FPGA. In Proceedings of the 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing, Las Vegas, NV, USA, 22–25 July 2013; pp. 1–4.
40. Wong, H.B.; Lim, G.H. Measures of diagnostic accuracy: Sensitivity, specificity, PPV and NPV. *Proc. Singap. Healthc.* **2011**, *20*, 316–318. [[CrossRef](#)]
41. Fletcher, S.; Islam, M.Z. Comparing sets of patterns with the Jaccard index. *Australas. J. Inf. Syst.* **2018**, *22*, 1–17. [[CrossRef](#)]