J CLIN PRACT RES

Original Article

Official Journal of Erciyes University Faculty of Medicine

DOI: 10.14744/cpr.2024.38483
J Clin Pract Res 2024;46(6):548–556

Segmentation Techniques in Mammography: Their Value in Breast Cancer Diagnosis

Hanife Avcı, Jale Karakaya

Department of Biostatistics, Hacettepe University Faculty of Medicine, Ankara, Türkiye

ABSTRACT

Objective: Computer-aided diagnosis (CAD) significantly enhances the accurate and early detection of breast cancer. The segmentation step, which is crucial in CAD methods, employs different algorithms for image segmentation as documented in the literature. Segmentation in detecting breast lesions using CAD may affect the features extracted from the images and, accordingly, the classification results. These segmentation methods have advantages and limitations when compared to each other. No study in the literature has yet explored feature matrices obtained by different segmentation methods using different performance criteria. This study aims to investigate the effects of different segmentation methods used in image processing on mammography images for breast cancer detection.

Materials and Methods: In the preprocessing step, images are enhanced using median filtering, Contrast Limited Adaptive Histogram Equalization (CLAHE), and unsharp masking. Texture features are extracted from the regions of interest (ROI) using texture analysis techniques, coupled with an elastic network technique for feature reduction.

Results: The performance of five different segmentation algorithms was compared using various performance measures such as accuracy, sensitivity, and specificity, alongside different classification methods. The k-means algorithm showed higher performance compared to other segmentation methods. It exhibited high efficacy, achieving an accuracy of 1.00 and 0.989 with Support Vector Machine (SVM) and Random Forest (RF) classifiers, respectively.

Conclusion: Segmentation methods used in image processing were found to have an impact on classification results. These computer-aided systems can be instrumental in patient classification.

Keywords: Biomedical image processing, computer-aided methods, decision making, machine learning, segmentation methods.

A very small part of the study was presented at the 23rd National and 6th International Biostatistics Congress held in Ankara, Türkiye (26–29 October 2022).

Cite this article as:
Avcı H, Karakaya J. Segmentation Techniques in Mammography: Their Value in Breast Cancer Diagnosis. JClinPractRes2024;46(6):548–556.

Address for correspondence:
Jale Karakaya.
Department of Biostatistics, Hacettepe University Faculty of Medicine, Ankara, Türkiye
Phone: +90 312 305 14 67
E-mail: jalekarakaya@gmail.com

Submitted: 25.01.2024
Revised: 11.06.2024
Accepted: 16.09.2024
Available Online: 14.10.2024

Erciyes University Faculty of Medicine Publications - Available online at www.jcpres.com

INTRODUCTION

With the discovery of X-ray and the increasing use of medical imaging systems, medical image processing has become increasingly important in healthcare, particularly over the last 20 years.[1] Imaging plays a crucial role in the diagnosis and evaluation of breast cancer. It is extensively used in tomography, magnetic resonance imaging (MRI), mammography, ultrasound, and positron emission tomography (PET), among others. Medical images are acquired using digital methods

and are primarily used for diagnostic purposes. However, the quality and resolution of images obtained from these methods are often low. These images are corrupted by random errors caused by devices and environmental factors. This unwanted information in digital images is referred to as noise, which complicates the diagnosis of diseases.

Various image processing techniques are employed to remove noise, enhance image quality, and sharpen images. Image processing algorithms consist of several components: preprocessing, image segmentation, feature extraction, feature reduction, and classification. These algorithms are commonly known as computer-aided diagnosis (CAD).[2] These methods assist radiologists in detecting both lesion and non-lesion images by improving the quality of medical images.[3]

Digital mammography has long been recognized as a potent imaging modality for the diagnosis of breast cancer.[4] Calcifications and masses manifest as areas of white density on both lesion and non-lesion mammography images. For instance, glands and connective tissues appear in lighter gray tones than fatty tissues in a mammogram image.[5] Different segmentation techniques are used to distinguish these varied regions. Segmentation methods are generally categorized into two types: region-based and edge-based methods. Image segmentation is a crucial and challenging step in CAD methods, as it involves using algorithms to partition an image into uniform regions.[6] Pixels in homogeneous regions share similar characteristics and form certain areas called regions of interest (ROI) in image processing. These ROIs are analyzed based on the content of the image, including texture, color, shape, and other qualities. These are important measurements that provide various characteristics of different regions of the image. Using the extracted features, each obtained region is classified into lesions and lesion-less images.

Different segmentation methods have been applied in the literature.[2,5–9] Some studies have discussed segmentation methods, and there are articles about these methods, but no study comparing their classification performance has been found. However, segmentation methods directly affect the attributes, which is expected to influence the classification performances.

The motivation of this study is to examine the success of five different segmentation methods in determining the lesion presence in features extracted from mammography images, after enhancing them with various preprocessing methods.

MATERIALS AND METHODS

The following subsections describe the theoretical basis of the implemented features.

KEY MESSAGES

- Image processing algorithms and classification methods can be employed for initial assessments, alleviating the burden on individuals with demanding workloads.
- In patient classification, radiologists can benefit from these computer-aided systems. These systems have demonstrated high performance in assisting radiologists by increasing the visibility of mammography images and enhancing lesion detection, making this study a valuable contribution to the literature.
- The segmentation method that provides the highest classification performance for detecting breast masses is the k-means algorithm.

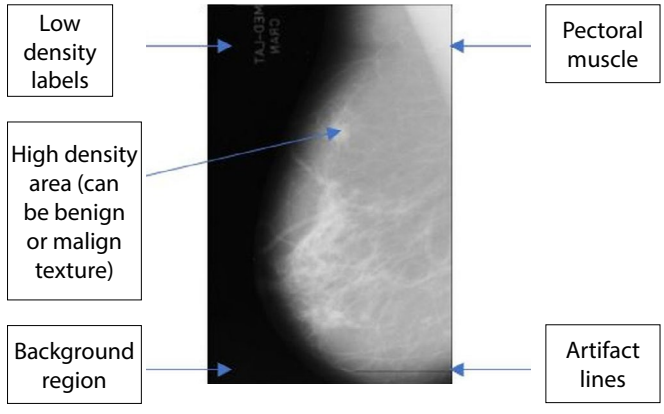

Figure 1. Different components in the left mediolateral oblique (MLO) mammography image featuring a malignant mass.

The Mammographic Image Analysis Society Digital Mammogram Database (MIAS)

The open-access mini-MIAS database was used in this study.[10] One of the reasons this dataset is widely used is that it provides an opportunity to compare results in the literature. In this study, 209 normal, 61 benign, and 52 malignant mammography images were used. The images have been evaluated by experienced radiologists, and abnormal lesions have been labeled. This database is available at http://peipa.essex.ac.uk./info/mias.html (Accessed on June 2, 2023). A sample malignant mammography from the collected database is shown in Figure 1. In this study, Fiji-ImageJ,[11] MATLAB version R2017b[12] for image processing methods, and RStudio (version 4.3.1)[13] software were used to examine the performance of classification methods. The "ggplot2" package[14] in RStudio was

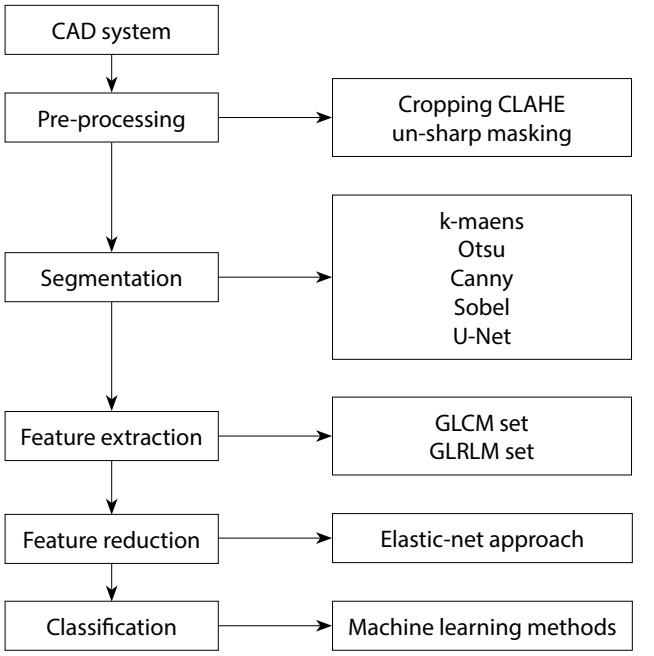

Figure 2. Graphical representation of the image processing workflow, illustrating the classification of mammography images into normal/abnormal and benign/malignant categories.

utilized to create the plots. The "caret" package[15] was used for the Leave One Out Cross Validation (LOOCV) method. Various algorithms used in the CAD method proposed in this study are shown in Figure 2.

Mammography Image Preprocessing

Mammography images are complex and difficult to interpret medical images. The objective of preprocessing is to enhance the image resolution by smoothing the image and eliminating unwanted information such as noise, low contrast, and artifacts that can lead to false positives. In previous studies, different filtering methods were used to eliminate various types of noise, smooth images, and improve or detect low frequencies.[2,7–9,16,17]. In a study we conducted on the comparison of these preprocessing methods, it was observed that combinations of Contrast Limited Adaptive Histogram Equalization (CLAHE), median filter, and unsharp masking algorithms yielded the best results.[18] Based on this study, these preprocessing algorithms (median filter, CLAHE, unsharp masking) were used to compare segmentation methods. In this study, the median filter method was applied to reduce noise and soften the images in mammography images. CLAHE and unsharp masking algorithms were applied to enhance the contrast of suspicious regions or regions of interest in the images and to make the edge regions clearer. These operations were performed on MATLAB R2017b.[12] The change in the mammography image after the preprocessing methods are applied is shown in Figure 3.

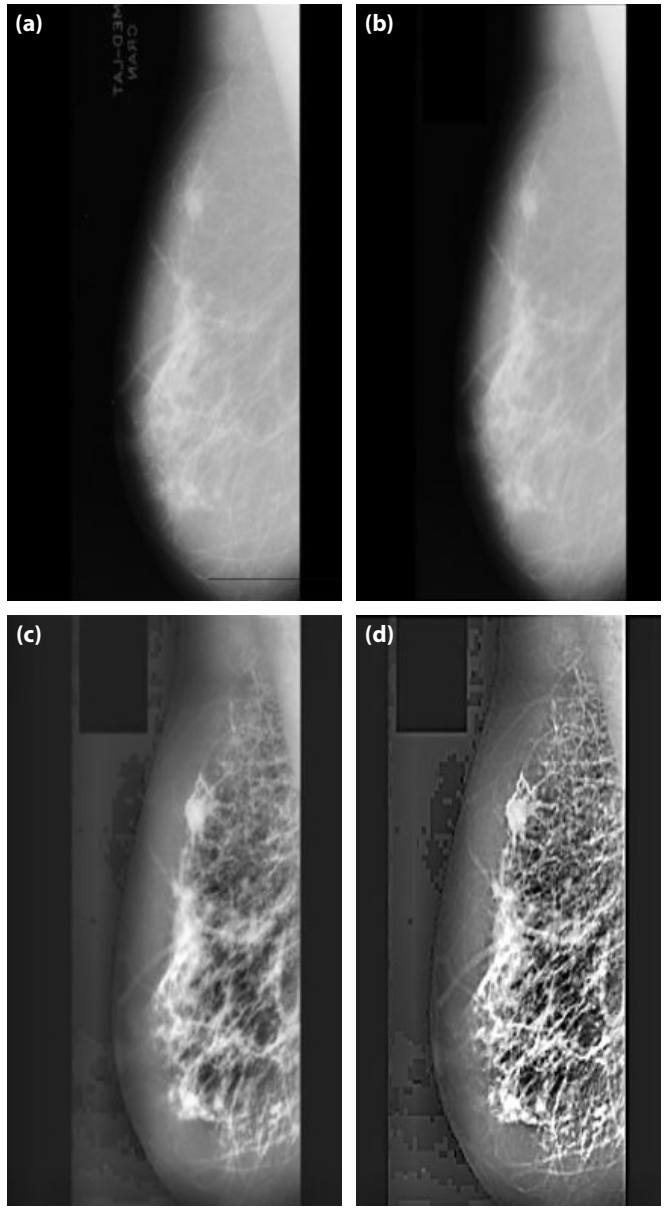

Figure 3. (a) Original left MLO malignant mammography image (mdb023) from the mini-MIAS database (the Mammographic Image Analysis Society Digital Mammogram Database). **(b)** Image processed with a median filter and label information removed. **(c)** Contrast-enhanced image using the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. **(d)** Enhanced edges of suspicious areas using the Unsharp Masking algorithm.

Mammography Segmentation

Texture features must be extracted, and the region of interest must be identified on the mammogram prior to classification. Another important step in identifying ROIs in mammography images and then extracting features from these regions is the segmentation step. Image segmentation entails dividing an image into separate areas or categories, where each area corresponds to different objects or parts of objects.[19] If image segmentation is done correctly, all other steps in image analysis become simpler. Therefore, the quality and reliability of the segmentation determine whether an image analysis will be successful. At the same time, the success of the segmentation step directly affects the classification performance of data mining algorithms. The purpose of segmentation methods is to extract the object of interest from an image with different backgrounds and complex regions; this means segmenting ROIs based on similar pixel values in the image. Segmentation techniques that can be used when dividing images into relevant regions can be divided into five basic categories: region-based (thresholding, region-growing, watershed, split and merge, and clustering), edge-based (Roberts, Sobel, Prewitt, Laplacian, and Canny), atlas-based, model-based, and deep learning techniques.[2,3,6–9,16,17,19–21]

Segmentation algorithms are generally divided into two types: region-based techniques based on the similarity of the intensity values of the image, and edge-based techniques based on the discontinuity feature.[17] In addition to region and edge-based algorithms, deep learning-based U-Net algorithms have been developed since 2015. It has been observed that deep learning models used in image processing have improved their performance in recent years. Therefore, in addition to region and edge-based segmentation algorithms, the deep learning-based U-Net algorithm has also been used.[22,23] In order to see the effect, five segmentation methods [two region-based (k-means clustering and Otsu), two edge-based (Canny and Sobel), and one deep learning-based (U-Net)] were applied to mammography images. After applying the appropriate segmentation method, the resulting ROIs are wrapped as a masking filter on the original image, and the resulting images are called ground reality images.

K-Means Clustering

The k-means clustering algorithm is a popular method for region-based segmentation commonly applied in image processing tasks. For our mammogram image segmentation task, we selected the number of clusters (k) as 3. This choice was based on domain knowledge and preliminary experiments, ensuring a balance between capturing relevant tissue structures and computational efficiency. The selection of k=3 allowed us to segment the images into meaningful regions corresponding to different tissue types and potential abnormalities. Each cluster has a specific center (centroid), and pixels are clustered according to their distances to these centers. For each pixel, the distance to all available centroids is calculated. The pixel is then assigned to the centroid with the shortest distance, thus grouping pixels into specific clusters. In this study, we calculated the distances using the Euclidean distance. Detailed information about the k-means algorithm is included in our previous study.[18]

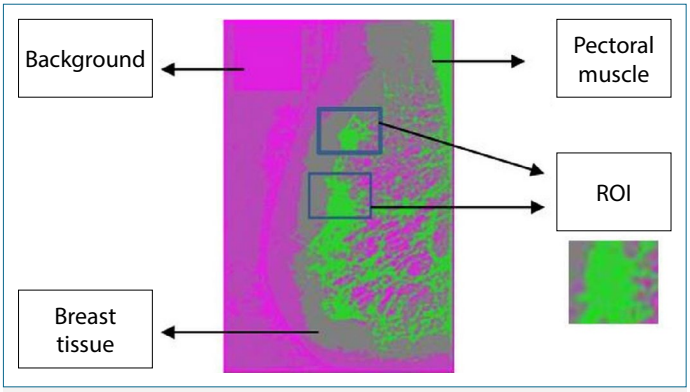

Figure 4. Output images following k-means segmentation.

As shown in Figure 4, after applying the k-means segmentation algorithm, mammography images are divided into three main clusters: pectoral muscle, breast tissue, and background.

Otsu's Method

Otsu's method is a binarization threshold image segmentation algorithm. It divides the image into light-colored input areas and dark gray areas based on their grayscale properties. The greatest difference between the two regions indicates the most optimal threshold. Otsu's method performs automatic thresholding on the image while segmenting the mammography image.

Canny and Sobel Edge Detection Operators

The most common approach used to detect discontinuity in the intensity values of pixels at the gray level is edge detection. An edge is the border between two regions with different gray levels. The edge contains information such as the location, shape, and texture of objects in the image. It is used to highlight or detect sudden changes in the gray level. Edge-based detectors can be used to detect these changes. In this study, the Canny and Sobel edge detectors, two of the edge-based techniques, were applied to mammography images. All edges determined by the Canny edge detection method.[16] are very similar to real edges. With the Canny algorithm, the distance between the point marked as the edge and the actual edge center is minimized.[17] The Sobel

edge operator is a standard edge detection operator that uses the average of adjacent pixels both vertically and horizontally. Due to the weighted average, this operator not only captures edge information but also helps reduce noise.[24]

U-Net

U-Net was developed by Olaf Ronneberger and his team in 2015 due to their work on biomedical images.[25] U-Net is a convolutional neural network with an encoder-decoder architecture, extensively used in medical imaging tasks. The name U-Net derives from its U-shaped architectural structure. U-Net has a symmetric architecture and has been successful in tasks such as image segmentation.

U-Net's architecture consists of two main parts: the encoder and the decoder. The encoder section includes various convolutional layers that extract feature maps from the input image. These layers perform convolution and pooling operations to extract the image features and reduce their dimensions. This process results in a compressed feature map that represents the higher-level features of the image. The decoder part then uses this feature map to perform convolution and up-sampling operations through a series of layers to produce a higher-resolution output. The decoder expands the feature map back to the dimensions of the original input image and generates the final segmentation map.

The combination of these components in the U-Net architecture allows it to excel in tasks like image segmentation, thanks to its strong feature extraction capabilities and the ability to produce high-resolution outputs.

Feature Extraction and Reduction

An image can exhibit various properties such as color, shape, and texture. Texture properties have been utilized in many previous studies to assess the classification performance of mammography images.[2]

After obtaining ROIs from mammography images, feature extraction was performed. In this study, the Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM) methods were used as feature extraction techniques. Consequently, a digital data matrix consisting of pixel values was obtained from the mammography images. The elastic-net approach was employed for dimension reduction of the feature space. In our study, we set alpha to 0.5, meaning we gave equal weight to both L1 and L2 penalties. We determined the best lambda value using cross-validation. Specifically, we employed the cv.glmnet function from the glmnet package,[26] which performs k-fold cross-validation to identify the optimal lambda value that minimizes the cross-validated error. Thirteen features were selected from the 33 features obtained.

Classification

The feature matrices obtained from the mammography images were first classified as normal/abnormal and then as images with benign/malignant lesions. Classification algorithms used included Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), k-Nearest Neighbor (k-NN), Naive Bayes (NB), and Decision Tree (DT) in the study.

The dataset is split into 70% for training and 30% for testing, and the models are developed using the LOOCV method. We assess the model's performance with accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), the area under the ROC curve (AUC), balanced accuracy (BA), and F1 measure.[27]

External Validity on the INbreast Data Set

The INbreast dataset, a collection of full-field digital mammograms, was made publicly accessible by the Hospital de São João in Porto, Portugal, in 2011.[28] The INbreast dataset is a large and comprehensive dataset of digital mammography images, frequently used for breast cancer research. Created in 2011 by researchers at the University of Porto, this dataset contains a total of 115 patients and 410 mammography images. The dataset includes craniocaudal (CC) and mediolateral oblique (MLO) mammography images. Normal, benign, and malignant classes were created using the Breast Imaging Reporting and Data System (BIRADS) classification in the INbreast dataset.

We performed all analyses on the Mini-MIAS dataset on the INbreast dataset to examine the validity and consistency of the results. We present our validation results for both normal/abnormal and benign/malignant classification in the Appendix. All results have been added to the Appendix.

RESULTS

In this study, the mini-MIAS dataset, which includes 322 mammography images from 161 patients, was utilized. Preprocessing techniques were used to eliminate noise, remove tags and unwanted information from the images, and enhance image resolution.

After applying the segmentation methods, the results of the normal/abnormal classification are given in Appendix 1. Region-based segmentation methods for normal/abnormal tissue classification showed higher classification performances than edge-based segmentation methods and the deep learning-based method (U-Net). Classification performances of k-means and Otsu methods, which are among the region-based techniques, were very close to each other.

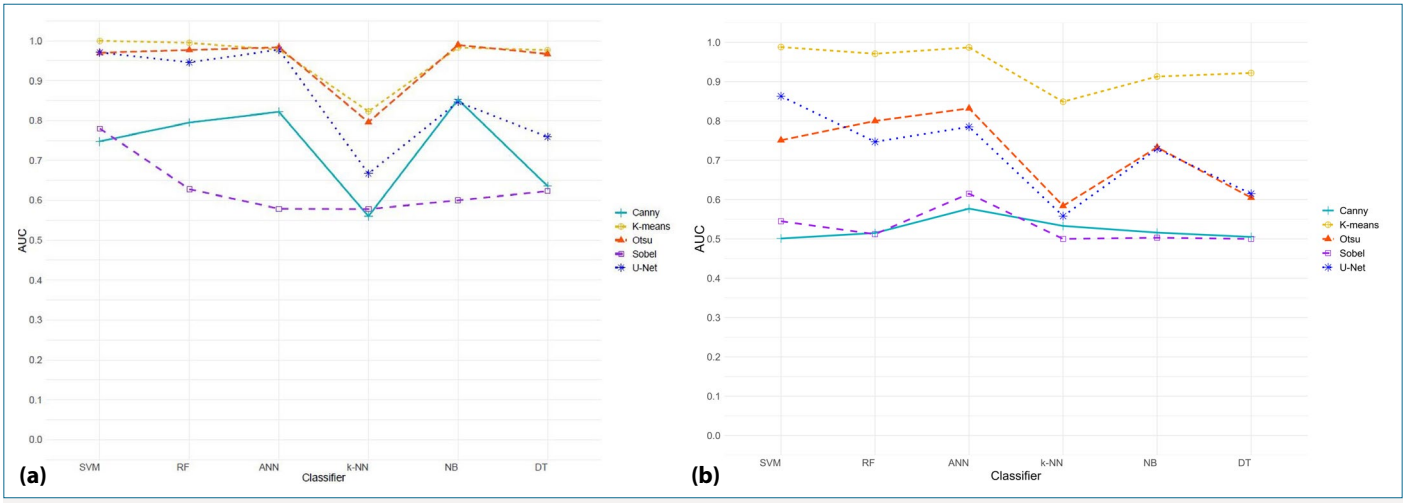

Figure 5. (a) Comparison of Area Under the Curve (AUC) values of classification methods for normal/abnormal segmentation (mini-MIAS). **(b)** Comparison of AUC values of classification methods for benign/malignant segmentation (mini-MIAS).

Table 1. Confusion matrices for normal/abnormal classification for the top performing proposed methods

Segmentation methods			SVM Predicted		RF Predicted		ANN Predicted	
			Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
K-Means	Actual	Normal	63	0	62	1	62	1
		Abnormal	0	34	0	34	1	33
Otsu	Actual	Normal	61	2	62	1	61	2
		Abnormal	1	33	1	33	0	34
Canny	Actual	Normal	52	11	58	5	59	4
		Abnormal	11	23	11	23	10	24
Sobel	Actual	Normal	63	0	55	8	60	3
		Abnormal	8	26	21	13	27	7
U-Net	Actual	Normal	61	2	60	3	62	1
		Abnormal	1	33	2	32	1	33

SVM: Support vector machine; RF: Random forest; ANN: Artificial neural network.

In all classifiers, k-means and Otsu showed higher performance than other edge-based segmentation methods. Regarding the Sobel algorithm, no significant differences were detected among the classification methods. For the normal/abnormal classification, the AUC values obtained from the classification methods of the segmentation methods are presented in Figure 5.

Benign/malignant classification results are given in Appendix 2. Region-based segmentation methods for benign/malignant tissue classification showed higher classification performances than edge-based segmentation methods.

The k-means clustering algorithm, which is a region-based technique, showed the highest performance across all classifiers. Little variation is observed in the classification results when applying features extracted by the Canny-Sobel algorithms. However, differences arise among the classification methods, particularly with respect to k-means, Otsu, and U-Net algorithms. For benign/malignant classification, the

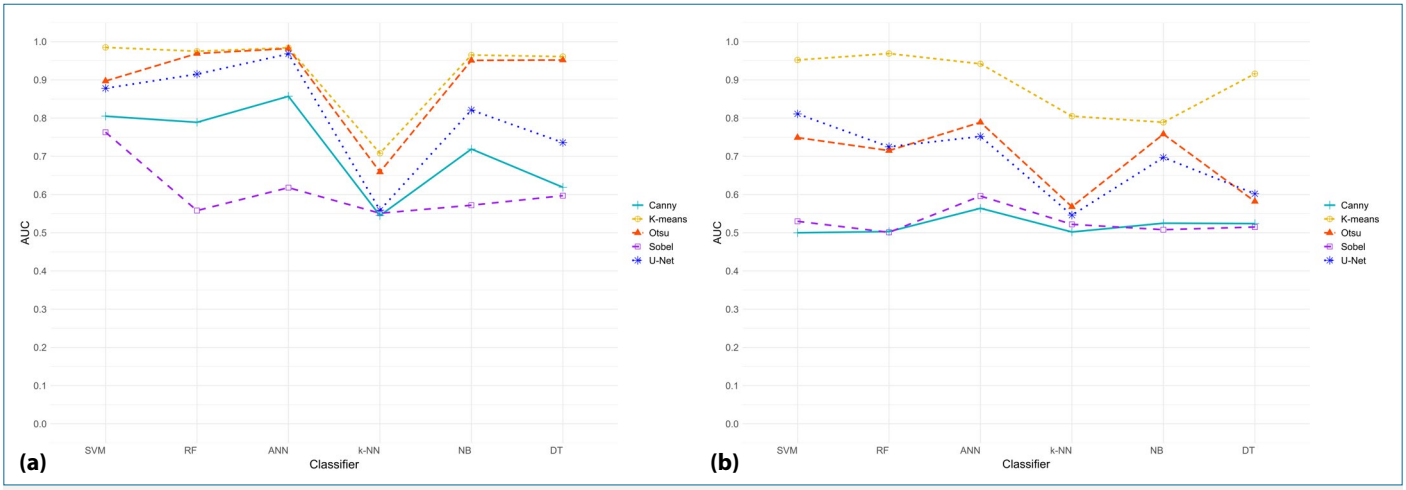

Figure 6. (a) Comparison of AUC values of classification methods for benign/malignant segmentation (INbreast dataset). **(b)** Comparison of AUC values of classification methods for benign/malignant segmentation (INbreast dataset).

AUC values obtained from the classification methods of the segmentation methods are presented in Figure 5.

Based on the results, SVM, RF, ANN, NB, and DT classifiers outperform k-NN in distinguishing between normal/abnormal and benign/malignant cases. Confusion matrices for normal/abnormal classification for the proposed methods showing the highest performance are given in Table 1.

We presented the results for the INbreast data set as Appendix 3, Appendix 4, and Figure 6 in the Appendix. In all classification methods, similar results were obtained, although there were minor differences in terms of performance measures obtained with the mini-MIAS and INbreast datasets for five different segmentation methods.

DISCUSSION

One of the best ways to diagnose breast cancer is through viewing mammogram images. Researchers and radiologists can get support from computer-assisted methods to detect the presence of a mass in images. These methods can both speed up the diagnosis process and increase its accuracy. Research in the literature highlights the critical role of the segmentation step in identifying suspicious areas. Among the various segmentation algorithms, k-means, Otsu, Canny, Sobel, thresholding, Laplacian of Gaussian (LoG), and Watershed algorithms are widely used.[2,7–9,20,21] However, no study compares the segmentation methods used in identifying suspicious lesions (mass or tumor) based on region and edge. Therefore, this study was designed to assess the effects of using different segmentation algorithms on the performance of classification methods. Although there are studies on segmentation methods, there is no study on classification performances of machine learning algorithms. Similarly, comparisons of the U-Net algorithm, which has gained popularity with the rise of deep learning in recent years, with the edge and region-based segmentation methods are not available in the literature.[29]

Limitation and Generalization of the Findings

The imbalance rates according to the number of images in the datasets for normal-abnormal and benign-malignant classification are 0.541 and 0.852, respectively. Thus, classification results were obtained on moderately balanced datasets. There is no imbalance problem that would decrease performance in the datasets we are currently working on. When the imbalance is below 30%, a serious imbalance is mentioned.[30,31] If the imbalance is very high (5–10%), some performance measures tend to be very small. Imbalance in datasets leads to conflicting results. The class imbalance problem is an important issue in machine learning. The classification performances of the machine learning techniques used in the study may differ in imbalanced datasets. New algorithms are proposed for the class imbalance problem. In this case, classification methods can be re-examined with different methods. Common features of these algorithms include synthetic data generation, undersampling, and oversampling.

Another limitation of the study is sample size. In studies conducted with machine learning algorithms, it is generally not possible to achieve high prediction accuracy without sufficient data samples. A suitable sample magnitude plays a crucial role in attaining precise and dependable outcomes.[32] According to research conducted by Cui and Gong in 2018, the predictive accuracy of machine learning algorithms demonstrated an upward trend with an increase in the sample size, irrespective of the specific algorithm utilized.[33]

CONCLUSION

In conclusion, in this paper, we compared the performance of the k-means clustering algorithm, Otsu method, Canny, Sobel, and U-Net segmentation methods for classifying mammography images. As a result of this study, the k-means clustering algorithm was identified as a segmentation method that provides higher classification results than other methods. The results of the deep learning-based U-Net algorithm did not demonstrate performance as high as region-based algorithms. Using machine learning algorithms, research conducted with inadequate data samples is unlikely to achieve a substantial level of predictive accuracy. Choosing an appropriate sample size is a key factor to increase the accuracy and reliability of analyses.

To sum up, the k-means algorithm demonstrates excellent performance across all classification methods, encompassing both normal/abnormal and benign/malignant categorizations. Additionally, for the segmentation of different medical images, it may be suggested that researchers choose the k-means clustering algorithm for its discriminative ability. In patient classification, radiologists can benefit from these computer-aided systems. Computer-Aided Systems have demonstrated high performance in assisting radiologists by increasing the visibility of mammography images and enhancing lesion detection.

Considering that the results may vary for different parameter values of the models, future studies can be designed because Artificial Intelligence (AI), machine learning, and data mining algorithms are used as classifiers in image analysis to help detect the presence of lesions. Image processing algorithms and classification methods can be employed for initial assessments, alleviating the burden on individuals with demanding workloads. Future research endeavors will center around the development of web-based tools, aiming to streamline the adoption of these methods by experts.

Acknowledgements: The author would like to thank Recep Demirci for the image processing interface called "MedPic" used in the segmentation stage of the images in this study.

Ethics Committee Approval: No ethics committee approval is required.

Author Contributions: Concept – JK; Design – JK, HA; Supervision – JK; Data Collection and/or Processing – JK, HA; Analysis and/or Interpretation – JK, HA; Literature Search – JK, HA; Writing – HA; Critical Reviews – JK.

Conflict of Interest: The authors have no conflict of interest to declare.

Informed Consent: Written informed consent was obtained from patients who participated in this study.

Use of AI for Writing Assistance: Not declared.

Financial Disclosure: The authors declared that this study has received no financial support.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Scholl I, Aach T, Deserno TM, Kuhlen T. Challenges of medical image processing, Computer Science-Research and Development 2011; 26(1): 5-13.
2. Mehdy MM, Ng PY, Shair EF, Saleh NIM, Gomes C. Artificial neural networks in image processing for early detection of breast cancer. Comput Math Methods Med 2017; 2017: 2610628.
3. Besl PJ, Jain RC. Segmentation through variable-order surface fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence 1988; 10(2): 167-92.
4. Andreea GL, Pegza R, Lascu L, Bondari S, Stoica Z, Bondari, A. The role of imaging techniques in diagnosis of breast cancer. Current Health Sciences J 2011; 37(2): 55-61.
5. Woods RW, Sisney GS, Salkowski LR, Shinki K, Lin Y, Burnside ES. The mammographic density of a mass is a significant predictor of breast cancer. Radiology 2011; 258(2): 417-25.
6. Jagya P, Dubey RG. Comparison of two segmentation methods for mammographic image. Int J Computer Applications 2015; 126(1): 31-43.
7. Dubey RB, Hanmandlu M, Gupta SK. A comparison of two methods for the segmentation of masses in the digital mammograms. Comput Med Imaging Graph 2010; 34(3): 185-91.
8. Jaffery ZA, Zaheeruddin, Singh L. Performance analysis of image segmentation methods for the detection of masses in mammograms. Int J Applications 2013; 82(2): 44-50.
9. Kaur J, Agrawal S, Vig R. A comparative analysis of thresholding and edge detection segmentation techniques. Int J Computer Applications 2012; 39(15): 29-34.
10. Suckling J. The mammographic image analysis society digital mammogram database, in exerpta medica. Int Congress Series 1069. York, England;1994. pp.375-8.
11. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012; 9(7): 676-82.
12. MATLAB and Image Processing Toolbox Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States. Available from: URL: https://www.mathworks.com/products/matlab.html. Accessed Sep 16, 2024.
13. RStudio Team (2021). RStudio: Integrated Development for R. RStudio Inc., Boston, MA. Available from: URL: http://www.rstudio.com. Accessed Sep 16, 2024.

14. Wickham H. ggplot2. Elegant Graphics for Data Analysis. New York, NY: Springer-Verlag New York; 2009.
15. Kuhn M. Building predictive models in R using the caret package. J Statistical Software 2008; 28(5): 1-26.
16. Canny J. A computational approach for edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1986; 8(6): 679-98.
17. Gonzales RC, Woods RE. Digital image processing. 3th edition. USA: Pearson Education; 2008.
18. Avcı H, Karakaya J. A novel medical image enhancement algorithm for breast cancer detection on mammography images using machine learning. Diagnostics (Basel) 2023; 13(3): 348.
19. Dey S. Hands-On Image Processing with Python. Packt Publishing Ltd; 2018.
20. Kumar A, Tiwari A. A comparative study of otsu thresholding and K-means algorithm of image segmentation. Int J Engineering and Technical Res 2019; 9(5): 12-4.
21. Liu D, Yu J. Otsu Method and K-means. 2009 Ninth International Conference on Hybrid Intelligent Systems, 2009; pp.344-9.
22. Dhungel N, Carneiro G, Bradley AP. Deep learning and structured prediction for the segmentation of mass in mammograms. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention. Springer; 2015. pp.605-12.
23. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging 2019; 32(4): 582-96.
24. Bouguet JY. Pyramid implementation of the Lucas Kanade feature tracker: description of the algorithm, Intel Corporation, Microsoft Research Labs Technical Report: 2000. Available from: URL: http://robots.stanford.edu/cs223b04/algo_tracking.pdf. Accessed Sep 16, 2024.
25. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention 2019; 9351: 234-41.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33(1): 1-22.
27. Karakaya J. Evaluation of binary diagnostic tests accuracy for medical researches. Turkish J Biochemistry 2021; 46(2): 103-13.
28. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast: toward a full-field digital mammographic database. Acad Radiol 2012; 19(2): 236-48.
29. Cho P, Yoon HJ. Evaluation of U-net-based image segmentation model to digital mammography. Conference: SPIE Medical Imaging-Virtual, Tennessee, United States of America; 2021.
30. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2012; 42(4): 463-84.
31. Chawla N, Japkowicz N, Kotcz A. Editorial: special issue on learning from imbalanced data sets. SIGKDD Explorations 2004; 6(1): 1-6.
32. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics 2023; 24(1): 48.
33. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. Neuroimage 2018; 178: 622-37.

Appendix 1. Performance of classification methods according to k-means, Otsu, Canny, Sobel, and U-Net segmentation algorithms for normal (n=209 images)/abnormal (n=113 images) classification

Data mining methods	Segmentation methods		Performance measures							
			Acc.	Sens.	Spec.	PPV	NPV	AUC	BA	F1
SVM	Region-based segmentation	K-means	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Otsu	0.969	0.971	0.970	0.943	0.984	0.970	0.970	0.952
	Edge-based segmentation	Canny	0.773	0.667	0.828	0.667	0.828	0.748	0.747	0.667
		Sobel	0.791	0.765	1.000	1.000	0.887	0.780	0.888	0.873
	Deep learning segmentation	U-Net	0.968	0.971	0.970	1.000	0.984	0.971	0.971	0.982
RF	Region-based segmentation	K-means	0.989	1.000	0.984	0.971	1.000	0.995	0.992	0.985
		Otsu	0.979	0.971	0.984	0.971	0.984	0.977	0.977	0.971
	Edge-based segmentation	Canny	0.835	0.667	0.922	0.821	0.843	0.795	0.794	0.733
		Sobel	0.701	0.382	0.873	0.619	0.724	0.628	0.628	0.461
	Deep learning segmentation	U-Net	0.952	0.941	0.952	0.914	0.968	0.946	0.947	0.927
ANN	Region-based segmentation	K-means	0.979	0.984	0.971	0.984	0.971	0.978	0.977	0.984
		Otsu	0.979	0.968	1.000	1.000	0.944	0.984	0.984	0.984
	Edge-based segmentation	Canny	0.856	0.706	0.937	0.857	0.855	0.822	0.822	0.774
		Sobel	0.690	0.206	0.952	0.700	0.690	0.579	0.579	0.306
	Deep learning segmentation	U-Net	0.979	0.971	0.984	0.971	0.984	0.978	0.978	0.971
k-NN	Region-based segmentation	K-means	0.742	0.697	0.766	0.605	0.830	0.823	0.731	0.648
		Otsu	0.773	0.461	0.887	0.600	0.818	0.796	0.674	0.522
	Edge-based segmentation	Canny	0.567	0.108	0.850	0.308	0.607	0.560	0.479	0.159
		Sobel	0.670	0.250	0.877	0.500	0.704	0.578	0.563	0.333
	Deep learning segmentation	U-Net	0.671	0.648	0.682	0.676	0.644	0.667	0.665	0.662
NB	Region-based segmentation	K-means	0.979	1.000	0.968	0.944	1.000	0.983	0.984	0.971
		Otsu	0.959	0.961	0.957	0.893	0.985	0.990	0.959	0.926
	Edge-based segmentation	Canny	0.763	0.656	0.815	0.636	0.828	0.853	0.736	0.646
		Sobel	0.577	0.531	0.600	0.395	0.722	0.600	0.566	0.453
	Deep learning segmentation	U-Net	0.871	0.875	0.786	0.783	0.871	0.847	0.831	0.826
DT	Region-based segmentation	K-means	0.969	1.000	0.957	0.900	1.000	0.977	0.978	0.947
		Otsu	0.979	0.961	0.986	0.961	0.986	0.967	0.974	0.961
	Edge-based segmentation	Canny	0.784	0.757	0.800	0.700	0.842	0.636	0.778	0.727
		Sobel	0.659	0.500	0.738	0.485	0.750	0.623	0.619	0.492
	Deep learning segmentation	U-Net	0.827	0.862	0.629	0.625	0.857	0.759	0.745	0.725

Acc: Accuracy; Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under the curve; BA: Balanced accuracy; SVM: Support vector machine; RF: Random forest; ANN: Artificial neural network; k-NN: k-Nearest neighbors; NB: Naïve bayes; DT: Decision tree.

Appendix 2. Performance of classification methods according to k-means, Otsu, Canny, Sobel, and U-Net segmentation algorithms for benign (n=61 image)/malignant (n=52 image) classification

Data mining methods	Segmentation methods		Performance measures							
			Acc.	Sens.	Spec.	PPV	NPV	AUC	BA	F1
SVM	Region-based segmentation	K-means	0.941	0.947	0.937	0.937	0.944	0.988	0.941	0.937
		Otsu	0.765	0.947	0.533	0.720	0.889	0.751	0.740	0.818
	Edge-based segmentation	Canny	0.412	0.357	0.450	0.312	0.500	0.501	0.404	0.333
		Sobel	0.529	0.412	0.647	0.539	0.524	0.545	0.529	0.467
	Deep learning segmentation	U-Net	0.864	0.774	0.883	0.769	0.879	0.863	0.829	0.771
RF	Region-based segmentation	K-means	0.970	1.000	0.947	0.937	1.000	0.971	0.973	0.967
		Otsu	0.735	0.684	0.800	0.812	0.667	0.800	0.742	0.743
	Edge-based segmentation	Canny	0.471	0.400	0.526	0.400	0.526	0.515	0.463	0.400
		Sobel	0.500	0.471	0.529	0.500	0.500	0.512	0.500	0.485
	Deep learning segmentation	U-Net	0.763	0.727	0.780	0.778	0.719	0.747	0.754	0.752
ANN	Region-based segmentation	K-means	0.925	0.929	0.920	0.929	0.930	0.987	0.925	0.929
		Otsu	0.794	0.895	0.667	0.773	0.833	0.832	0.781	0.829
	Edge-based segmentation	Canny	0.554	0.558	0.615	0.553	0.450	0.577	0.586	0.553
		Sobel	0.618	0.529	0.706	0.643	0.600	0.615	0.617	0.581
	Deep learning segmentation	U-Net	0.759	0.792	0.773	0.766	0.790	0.785	0.783	0.779
k-NN	Region-based segmentation	K-means	0.823	1.000	0.667	0.727	1.000	0.849	0.833	0.842
		Otsu	0.588	0.684	0.467	0.619	0.538	0.584	0.575	0.650
	Edge-based segmentation	Canny	0.530	0.545	0.522	0.353	0.706	0.533	0.534	0.428
		Sobel	0.559	0.471	0.647	0.571	0.550	0.500	0.559	0.516
	Deep learning segmentation	U-Net	0.545	0.543	0.563	0.554	0.533	0.558	0.553	0.548
NB	Region-based segmentation	K-means	0.706	1.000	0.545	0.545	1.000	0.913	0.773	0.706
		Otsu	0.676	0.737	0.600	0.700	0.643	0.733	0.668	0.718
	Edge-based segmentation	Canny	0.617	0.467	0.737	0.583	0.636	0.516	0.602	0.519
		Sobel	0.588	0.412	0.765	0.636	0.565	0.503	0.588	0.500
	Deep learning segmentation	U-Net	0.728	0.724	0.732	0.719	0.709	0.729	0.728	0.721
DT	Region-based segmentation	K-means	0.941	1.000	0.888	0.888	1.000	0.922	0.944	0.941
		Otsu	0.588	0.526	0.667	0.667	0.526	0.605	0.596	0.588
	Edge-based segmentation	Canny	0.588	0.714	0.500	0.500	0.714	0.505	0.607	0.588
		Sobel	0.559	0.471	0.647	0.571	0.550	0.500	0.559	0.516
	Deep learning segmentation	U-Net	0.613	0.628	0.575	0.567	0.619	0.615	0.601	0.596

Acc: Accuracy; Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under the curve; BA: Balanced accuracy; SVM: Support vector machine; RF: Random forest; ANN: Artificial neural network; k-NN: k-Nearest neighbors; NB: Naïve bayes; DT: Decision tree.

Appendix 3. Performance of classification methods according to k-means, Otsu, Canny, Sobel, and U-Net segmentation algorithms for normal (n=67 image)/abnormal (n=124 image) classification (INbreast dataset)

Data mining methods	Segmentation methods		Performance measures							
			Acc.	Sens.	Spec.	PPV	NPV	AUC	BA	F1
SVM	Region-based segmentation	K-means	0.986	0.975	0.991	0.965	0.959	0.985	0.983	0.970
		Otsu	0.883	0.879	0.901	0.898	0.882	0.897	0.890	0.888
	Edge-based segmentation	Canny	0.743	0.658	0.816	0.643	0.784	0.805	0.737	0.650
		Sobel	0.707	0.649	0.796	0.703	0.784	0.763	0.723	0.675
	Deep learning segmentation	U-Net	0.925	0.866	0.918	0.902	0.912	0.878	0.892	0.884
RF	Region-based segmentation	K-means	0.973	0.965	0.998	0.991	0.958	0.975	0.982	0.978
		Otsu	0.963	0.959	0.972	0.965	0.979	0.969	0.966	0.962
	Edge-based segmentation	Canny	0.802	0.629	0.819	0.795	0.806	0.789	0.724	0.702
		Sobel	0.698	0.493	0.819	0.503	0.734	0.558	0.656	0.498
	Deep learning segmentation	U-Net	0.929	0.905	0.948	0.952	0.937	0.915	0.927	0.928
ANN	Region-based segmentation	K-means	0.981	0.974	0.986	0.978	0.989	0.984	0.980	0.976
		Otsu	0.976	0.952	0.984	0.964	0.982	0.982	0.968	0.958
	Edge-based segmentation	Canny	0.828	0.713	0.918	0.803	0.826	0.857	0.816	0.755
		Sobel	0.701	0.314	0.892	0.729	0.693	0.618	0.603	0.464
	Deep learning segmentation	U-Net	0.968	0.957	0.983	0.963	0.949	0.968	0.970	0.960
k-NN	Region-based segmentation	K-means	0.706	0.695	0.714	0.725	0.739	0.708	0.705	0.710
		Otsu	0.726	0.504	0.746	0.531	0.738	0.659	0.625	0.517
	Edge-based segmentation	Canny	0.507	0.206	0.769	0.298	0.594	0.545	0.488	0.244
		Sobel	0.618	0.307	0.863	0.519	0.683	0.551	0.585	0.386
	Deep learning segmentation	U-Net	0.652	0.539	0.672	0.648	0.627	0.559	0.601	0.588
NB	Region-based segmentation	K-means	0.958	0.995	0.949	0.989	0.968	0.965	0.972	0.992
		Otsu	0.945	0.963	0.959	0.971	0.960	0.951	0.961	0.967
	Edge-based segmentation	Canny	0.668	0.639	0.721	0.632	0.727	0.719	0.680	0.635
		Sobel	0.543	0.517	0.592	0.403	0.628	0.572	0.555	0.453
	Deep learning segmentation	U-Net	0.852	0.861	0.759	0.802	0.843	0.821	0.810	0.830
DT	Region-based segmentation	K-means	0.949	0.989	0.967	0.974	0.982	0.961	0.978	0.981
		Otsu	0.937	0.976	0.948	0.958	0.978	0.952	0.962	0.967
	Edge-based segmentation	Canny	0.762	0.743	0.787	0.696	0.825	0.619	0.765	0.719
		Sobel	0.643	0.505	0.665	0.496	0.684	0.597	0.585	0.500
	Deep learning segmentation	U-Net	0.802	0.847	0.726	0.719	0.817	0.736	0.787	0.778

Acc: Accuracy; Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under the curve; BA: Balanced accuracy; SVM: Support vector machine; RF: Random forest; ANN: Artificial neural network; k-NN: k-Nearest neighbors; NB: Naïve bayes; DT: Decision tree.

Appendix 4. Performance of classification methods according to k-means, Otsu, Canny, Sobel, and U-Net segmentation algorithms for benign (n=67 image)/malignant (n=57 image) classification (INbreast dataset)

Data mining methods	Segmentation methods		Performance measures							
			Acc.	Sens.	Spec.	PPV	NPV	AUC	BA	F1
SVM	Region-based segmentation	K-means	0.926	0.932	0.919	0.935	0.928	0.952	0.926	0.933
		Otsu	0.751	0.893	0.501	0.703	0.874	0.749	0.697	0.787
	Edge-based segmentation	Canny	0.401	0.447	0.329	0.501	0.495	0.500	0.373	0.472
		Sobel	0.518	0.405	0.635	0.527	0.515	0.530	0.520	0.458
	Deep learning segmentation	U-Net	0.837	0.762	0.875	0.759	0.868	0.811	0.819	0.760
RF	Region-based segmentation	K-means	0.958	0.982	0.938	0.895	0.987	0.969	0.960	0.936
		Otsu	0.756	0.674	0.718	0.795	0.703	0.715	0.696	0.730
	Edge-based segmentation	Canny	0.432	0.398	0.519	0.402	0.515	0.503	0.459	0.400
		Sobel	0.492	0.501	0.514	0.498	0.501	0.501	0.507	0.499
	Deep learning segmentation	U-Net	0.739	0.719	0.753	0.764	0.698	0.725	0.736	0.741
ANN	Region-based segmentation	K-means	0.895	0.914	0.893	0.920	0.917	0.942	0.904	0.917
		Otsu	0.768	0.882	0.648	0.698	0.802	0.789	0.765	0.779
	Edge-based segmentation	Canny	0.538	0.516	0.604	0.549	0.502	0.564	0.560	0.532
		Sobel	0.613	0.505	0.694	0.627	0.597	0.596	0.599	0.559
	Deep learning segmentation	U-Net	0.736	0.782	0.697	0.757	0.783	0.752	0.739	0.769
k-NN	Region-based segmentation	K-means	0.785	0.827	0.703	0.714	0.818	0.805	0.765	0.766
		Otsu	0.563	0.578	0.438	0.596	0.527	0.568	0.508	0.587
	Edge-based segmentation	Canny	0.522	0.518	0.483	0.439	0.508	0.502	0.500	0.500
		Sobel	0.531	0.523	0.547	0.603	0.543	0.522	0.535	0.560
	Deep learning segmentation	U-Net	0.539	0.527	0.541	0.515	0.537	0.546	0.534	0.521
NB	Region-based segmentation	K-means	0.803	0.897	0.639	0.785	0.826	0.789	0.768	0.837
		Otsu	0.765	0.728	0.794	0.637	0.758	0.758	0.761	0.679
	Edge-based segmentation	Canny	0.596	0.505	0.598	0.564	0.603	0.525	0.552	0.533
		Sobel	0.572	0.498	0.604	0.563	0.574	0.508	0.551	0.529
	Deep learning segmentation	U-Net	0.694	0.702	0.721	0.695	0.704	0.697	0.712	0.698
DT	Region-based segmentation	K-means	0.919	0.938	0.825	0.839	0.927	0.916	0.882	0.886
		Otsu	0.567	0.539	0.603	0.628	0.615	0.582	0.571	0.580
	Edge-based segmentation	Canny	0.549	0.684	0.505	0.507	0.677	0.524	0.595	0.582
		Sobel	0.541	0.503	0.625	0.538	0.617	0.515	0.564	0.520
	Deep learning segmentation	U-Net	0.595	0.617	0.563	0.548	0.607	0.602	0.590	0.580

Acc: Accuracy; Sens: Sensitivity; Spec: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; AUC: Area under the curve; BA: Balanced accuracy; SVM: Support vector machine; RF: Random forest; ANN: Artificial neural network; k-NN: k-Nearest neighbors; NB: Naïve bayes; DT: Decision tree.