

# Improving Breast Cancer Classification Using (Smote) Technique and Pectoral Muscle Removal in Mammographic Images

**Srwa Hasan Abdulla<sup>1</sup>✉, Ali Makki Sagheer<sup>2</sup>, Hadi Veisi<sup>3</sup>**

<sup>1</sup>Basic Science Department, Sulaimani University, Iraq

<sup>2</sup>Computer Science and Information Technology, Al-Qalam University College, Iraq

<sup>3</sup>Computer Engineering, Tehran University, Iran

sirwa.abdulla@univsul.edu.iq✉, prof.ali@alqalam.edu.iq, h.veisi@ut.ac.ir

## Abstract

Computer-aided diagnosis methods are being developed to assist radiologists to improve the interpretation of mammograms for the detection and diagnose of breast cancer, reduce the errors and mistakes made by human beings. In addition, it provides a more accurate and reliable classification of benign and malignant abnormalities. In the mammogram diagnosis, the pectoral muscle appears in Mediolateral oblique views (MLO) of the right and left of the breast. Considering that, the pectoral muscle has the same density as the small suspicious masses in the image and can affect/bias the results of image processing methods. This paper presents a diagnosis method to detect an abnormality in mammograms automatically. Before abnormality identification, image-processing techniques are used to correctly segment the suspicious region-of-interest (ROI). The background of the mammograms has been darkened to distinguish the breast area from any blemishes or writings that will be removed. Then the breast area has been extracted after ignoring the empty regions around the breast in mammogram images. After that, the mammogram image is inverted and the inverted image is then subtracted from the initial image. For pectoral muscle removal, a region growing method using the K-means clustering method is used. Afterward, suspicious ROI is segmented utilizing the K-means with thresholding technique. To detect abnormalities in mammograms, shape-based features, moment invariants, and also fractal dimensions are extracted from the segmented ROI. The Mini-MIAS dataset is used to evaluate the proposed method and is predominately composed of benign samples with only a tiny percent of malignant samples. To accomplish far better classifier efficiency, the SMOTE algorithm is used to present new samples from the minority classes to get a balanced dataset. Random forest classifier utilized to classify the segmented region as benign and malignant. The experimental results obtained an accuracy of 97.1%, the sensitivity of 95.1%, and the achieved specificity is 98.5%.

**Keywords:** Breast Cancer, Mammogram, Pectoral Muscle, K-means Clustering, SMOTE, Random Forest.

Received: 03 November 2021

Accepted: 14 December 2021

Published: 21 December 2021

## 1 Introduction

In regards to cancer, after lung cancer, breast cancer is the second greatest cause of death [1]. It is mostly discovered in females and according to World Health Organization statistics, the first widely occurring cancer worldwide [23], [24].

Mammography has long been regarded as one of the most important technique for identifying breast cancer. It can be utilized to diagnose illness at an early stage when there is still a chance of recovery [15, 17, 13, 35]. The goal of a computer-aided diagnosis (CAD) system is to peruse mammograms, find the suspicious areas of irregularities and evaluate their attributes. The efficiency and integrity of a CAD depend upon accurate segmentation of the lesions as well as proper feature selection [29]. The quality of the images, indeed,

has a significant influence on the performance of any CAD system. As well as the limits of X-Ray technology, mammogram images frequently have low contrast, making it impossible to understand the features of mammograms. As a result, most images must be improved in quality before further analysis [8]. The images impact the segmentation method of mass diagnosis algorithms. Intensity degrees of both mass, as well as pectoral muscle are considerably similar. They have relatively high values of intensity than the remaining regions. To get more accurate results, it is strongly recommended to eliminate the pectoral muscles before implementing any segmentation method.

The number of observations from the malignant class in the Mini-MIAS dataset [32] is significantly less than from the benign class. Machine learning algo-

rithms' performance is usually measured using accuracy. Nonetheless, this is not ideal when the data is unbalanced and/or the expenses of different errors differ considerably [3]. The Synthetic Minority Over-Sampling Technique (SMOTE) is a popular preprocessing approach for dealing with unbalanced datasets, in which the minority class is oversampled by creating synthetic instances in feature space [10].

In this paper, a method will be proposed for the removal of pectoral muscle to improve the segmentation process, which directly affects the improvement of classification results, in addition to solving the imbalance problem in the Mini-MIAS dataset.

The structure of this paper is as follows: The associated work for the proposed approach is explained in Section 2. Section 3 provides a theoretical background about the research problem. Section 4 provides a detailed overview of the proposed methodology. The performance analysis that demonstrates the classification efficiency of the suggested system is given in Section 5. The conclusion of this work is obtained in Section 6.

## 2 Related Works

Previous researches reveal that CAD systems can streamline the process of translating mammograms and also offering precise result [23]. The result of a CAD is utilized to assist the radiologist in the discovery of breast cancer [31]. Over the previous two decades, scientists proposed various sorts of classifiers in order to create an efficient and also optimum CAD for mammograms.

Jaleel, et al. in [12] proposed a model that extracted texture features using the Discrete Wavelet Transform (DWT) and the Gray Level Co-occurrence Matrix (GLCM). Then, Artificial Neural Network was used for classifying images into benign or malignant classes. The performance of the proposed model was consulted with Mini-MIAS data. The accuracy accomplished by utilizing this model was 93.7% with GLCM and also 94.6% by utilizing DWT features, respectively.

Makandar and Halalli [20] proposed a system based on extracting the suspicious area coming from the breast. The region growing method was used for image segmentation. The Mini-MIAS database was used to determine the system's effectiveness. The accuracy was stated to be 95.86%.

Xie, et al. in [34] offered a CAD system for the medical analysis of breast cancer that was based upon the Extreme Learning Machine (ELM). The system achieved common reliability of 96.02% by utilizing Mini-MIAS and DDSM datasets.

Kaur, et al. in [16] proposed a new method, utilized on the Mini-MIAS dataset of 322 images, including a pre-processing technique and additionally integrated feature extraction using K-means clustering for Speed-Up Robust Features (SURF) selection. The obtained results for the support vector machine (SVM), linear discriminant analysis (LDA), K-nearest neighbor

(KNN), and Decision Tree were 96.9%, 89.7%, 93.8%, and 88.7%, respectively.

Kamil & Jassam in [14] presents a feature extraction technique for breast mass recognition utilizing the (GLCM) using mammography images, distinguishing between normal and abnormal breast masses. A k-nearest neighbor classifier has the highest sensitivity, specificity, and accuracy which was 92%, 86%, and 86.1%, respectively.

Sharma and Mehra in [30] Two approaches (hand-crafted features and transfer learning) are extensively studied and compared for the purpose of automatic multi-classification relying on magnification on the balancing BreakHis dataset for breast cancer detecting. The Hu moment, color histogram, and Haralick textures are used in the first technique to extract hand-crafted features. The best accuracy is provided by the VGG16 network with linear SVM, which is calculated in two ways: (a) patch-based accuracies 93.97% and (b) patient-based accuracies 93.25%.

Radhi and Kamil in [26] applied two approaches of active contour which are snakes and level sets. A mini-MIAS database was used to test the proposed methods. The proposed models were evaluated in two ways: statistically, the Chan-Vese method yielded the best results (90%, 95%, 98%, 97%, and 97%) for Jaccard, Dice, PF-Score, Precision, and Sensitivity, respectively. Chan-Vese was able to properly determine the location and shape of the tumor based on the characteristics of the segmented region.

As discussed above, various image-processing techniques and segmentation algorithms have been used by researchers to analyze samples and enhance visual accuracy to discover and interpret regions of interest. In this paper, the focus will be on removing pectoral muscle to improve the segmentation process and solve the problem of data imbalance as in the case of the Mini-MIAS dataset to avoid its impact on the accuracy of classification results.

## 3 Background Theory

In this section, the theoretical background of the current research project will be briefly reviewed. An overview of the most important theories and techniques in image processing, segmentation of regions of interest, feature extraction, data balancing and classification will be presented.

### 3.1 Pre-processing

In any image processing technique, preprocessing is regarded as the fundamental step. The ultimate objective of this technique is to improve the image quality and the image characteristics that are necessary for further processing. Mammogram images are difficult to view compared with other medical images, so pre-processing is important [18], [7].

### 3.2 ROI Segmentation

Segmentation is a process of partitioning digital mammograms into non-overlapping segments to distinguish masses from the background. Various algorithms, such as local and global thresholding based on image histograms are used to segment the breast masses [24]. In addition, methods based on regions, as an example, in the region growing, a seed point is utilized and the region is grown until it meets homogeneity criteria [22].

### 3.3 Features Extraction

In classification, feature extraction is crucial [4], [27], and [33]. Several types of image features have been examined and evaluated for classification applications. Fractal dimension using pixel range calculation methodology, with three region-based features i.e., eccentricity, solidity, and extent of the segmented region are collected. Then from the segmented suspicious mass region, seven moment invariants are obtained.

### 3.4 Data Balancing

Class-imbalanced data is a data in which the result is binary and the frequency of the observed classes is skewed to one of two possible results, the majority class versus the minority class. The problem is that many ML methods are “biased” towards the majority class in the presence of imbalanced data [11]. SMOTE is a well-known method for generating balanced datasets by oversampling the minority class. By practicing each minority class sample and including synthetic instances along any/all of the k minority class’s nearest neighbors’ line segments, it oversamples the minority class. Neighbors from the k-nearest neighbors are randomly selected depending on the amount of oversampling needed [5].

Synthetic samples are generated by considering the difference between the nearest neighbor and the feature vector (sample). Add this difference to the feature vector under consideration by multiplying it by a random integer between 1 and 0. As a result, a random point along the line segment between two distinct features is selected [3].

The SMOTE algorithm is detailed below:

1. For each sample, find the k-nearest neighbors.
2. Choose samples from a k-nearest neighbor randomly.
3. Find new samples by multiplying the initial samples by the difference and by the gap of (0, 1).
4. Add the additional samples to the minority. After that, a new set is generated.

### 3.5 Classification

Random forest (RF) is an example of supervised machine learning, which has become one of the most

widely used methods by offering a broad range of solutions to classification problems, and is known for its high predictive accuracy and ability to handle high-dimensional data [25].

Assume that  $T_r$  is a training set with properties and  $i$  cases, and that  $T_k$  is a bootstrap training set sampled from  $T_r$  with replacement and containing  $r$  random properties ( $r \leq a$ ) with  $i$  cases.

A Random Tree is a tree with  $r$  random properties drawn at random from a collection of possible trees. Each tree has an equal chance of being sampled, as the word “at random” means. Random Trees can be efficiently generated, and huge collections of Random Trees can be combined to build accurate models in most scenarios.

A Random Forest is a classifier consisting of a set of tree-structured classifiers  $\{hk(C, T_k)\}$ ,  $k = 1, 2, \dots, L$ , where  $T_k$  are independent identically distributed random samples and each tree performs a unit vote for the most common class at input  $C$  [21], [2].

### 3.6 Performance Evaluation

To evaluate machine learning techniques, evaluations metrics such as accuracy, sensitivity, and specificity are required. For the calculation of these metrics, the confusion matrix distinguishes the terms TP (true-positive), TN (true-negative), FP (false-positive), and FN (false negative) from the predicted and ground truth result.

Accuracy, sensitivity, and specificity are calculated utilizing equations (1, 2, and 3), measure the efficiency of the proposed method outlined in the paper. The quality of classification can be determined as [28]:

- Sensitivity: Sensitivity is a test that decides the chances of outcomes that are correctly identified when the cancer is present.

$$\text{Sensitivity} = TP / (TP + FN) \quad (1)$$

- Specificity: is a test that determines the chances of the outcomes that are true negative which are correctly identified.

$$\text{Specificity} = TN / (TN + FP) \quad (2)$$

- Accuracy: is a test that determines the chances that how many samples are correctly identified.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

Receiver Operating Characteristics (ROC) curve: A ROC curve is created by collating the sensitivities and specificities for various values of a continuous test measure. This gives you a list of different test values, as well as the sensitivity and specificity of the test at those levels. For each of the tabulated data, the ROC curve is generated by plotting sensitivity on the y-axis versus 1– specificity on the x-axis. As a result, a diagnostic test with adequate accuracy should have a ROC curve in the top left triangle [9].

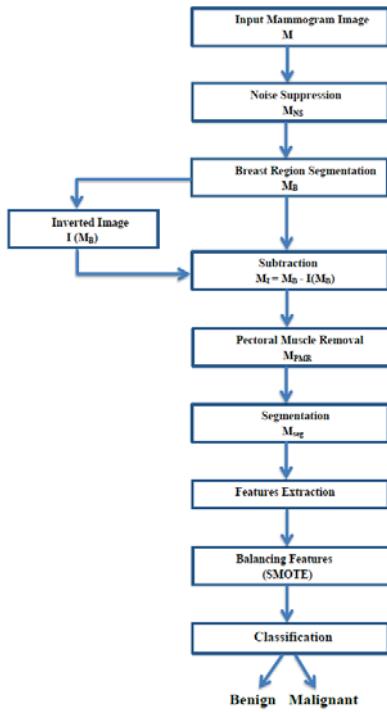


Figure 1: Flowchart of the proposed method.

#### 4 Proposed Method

The proposed methodology involves multiple steps including image processing techniques. The first step is the image acquisition from the Mini-MIAS database of mammogram that is freely available [32]. There are 322 mammograms in this data set: 270 of the sample images are benign, while the remaining 52 are malignant.

The proposed method demonstrated in Fig. 1, which involves noise reduction, breast region differentiation, filtering, pectoral muscle removal, segmentation, feature extraction, balancing the data in the feature space as well as performing the classification. In regards to the segmentation process, the proposed system performs many operations to improve the segmentation process, the most important of which is solving the pectoral muscle problem and ensuring that it does not appear which affects the efficiency of extracting characteristics and thus improve the efficiency of the classification process. Then, the proposed system solves the imbalance problem in the Mini-MAIS dataset to raise the classification efficiency.

The mammogram images include a variety of noises such as tape artifacts, labels, and scanning artifacts. Various noises like these should be removed and not processed anymore. After the noises have been removed, the breast region is segmented. Original mammogram images are binaries by special threshold technique which blacking the background only in order to separate it from the mammogram image content. Then, morphological erosion operation is performed to

remove small regions. After that, the process of determining the breast region (region of interest) is carried out, which is in two steps: The first is to determine the beginning and end of the breast region (left and right), where the maximum width of the breast is adopted, and then determining the beginning and end of the region of interest. The second step includes determining the (top and bottom) of the region of interest, at the top the highest point belonging to the region of the breast is approved. As for the bottom, the end of the breast lump is determined and approved as a lower point for the region of interest. Thus, the area of the breast that is confined between the four points extracted is adopted.

To produce a subtracted image, the obtained breast image was inverted. The original breast image is subtracted from the inverted image.

Depending on the direction of the mammogram, the pectoral muscle appears in the left-top corner or right-top corner of MLO images. In comparison to the main breast tissue, they have higher density values. In order to improve the segmentation process and the classification performance, an automated method for pectoral muscle removal is presented. A mammogram is clustered with a binary split algorithm before applying the steps of the procedure for removing the pectoral muscle that has been proposed. The method is divided into three phases: The alignment of the breast is determined in the first phase; mammograms are either right or left-aligned. In the second phase, a triangle is placed over the mammogram image to isolate the main breast region from the pectoral muscle. To suppress the pectoral muscle, a seeded region growing algorithm is being used in the third step. Only the triangular defined in the second step is used in this procedure. The final segmented unlabeled image is generated by superimposing the remaining breast area on the subtracted image. Segmentation is carried out to divide the images into uniform areas and extract the region-of-interest (ROI). On a sharpened image, K-means clustering with thresholding is used to isolate the segmented suspicious mass area. A feature vector is generated using a collection of shape-based, moment-invariant, and fractal dimension features derived from each segmented suspicious area.

In the Mini-MAIS dataset [32], the number of observations coming from the malignant class is substantially less than those belonging to the benign class. The predictive model created utilizing typical machine learning methods may be biased and inaccurate in this situation. This happens because Machine Learning Algorithms are normally designed to increase accuracy by reducing the number of errors. As a consequence, they don't consider the distribution of classes, their proportions, or the balance of classes [6], [19]. Because of that, employ Synthetic Minority Over-sampling Technique (SMOTE) to cover the impact of this on the accuracy of the classifier. A subset of data from the minority class is selected, and then new synthetic comparable

instances are created. After that, the initial dataset is supplemented with these synthetic instances. To train the classification models, a new dataset is used for this purpose. The random forest classifier is made use of to identify the segmented region as benign or malignant.

## 5 Results and Discussion

The suggested algorithm is evaluated using 322 mammogram images from the Mammographic Image Analysis Society (Mini MIAS) dataset [32], which is publicly available. Initially, using a bilinear interpolation approach, all instance images are reduced to 256\*256 pixels. After that tape artifacts and labels are removed.

The intermediate findings of breast area segmentation are seen in Fig. 2. The initial mammogram, darkening background mammogram after performing special threshold operation, and mammogram after performing morphological erosion operation (enhanced image) are shown in Fig. 2(a), (b), and (c), respectively.

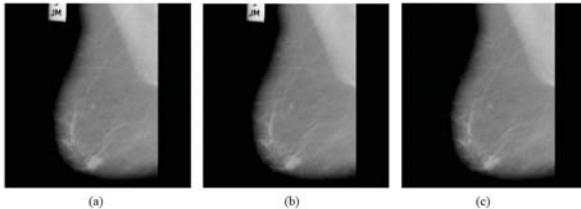


Figure 2: Breast region identification results: a) Original mammogram, b) Darkening background image, c) Morphological erosion image.

Fig. 3 shows the breast region segmentation process. The original revised image, the first step of the breast region segmentation, and the second step of the breast region segmentation, are shown in Fig 3(a), (b), and (c), respectively.

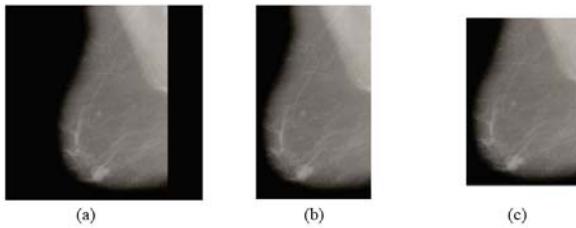


Figure 3: Breast region segmentation process: a) Original revised image, b) (step 1) breast segmentation, c) (step 2) breast segmentation.

Inverted and subtracted images are shown in Fig. 4. In the subtracted images, the suspicious mass area is clearly improved. Fig. 5 illustrates the elimination of the pectoral muscle and the segmentation of suspicious mass regions. Fig. 6 shows sample segmented images after applying all of the proposed model's operations.

The first three features of the seven moments invariant, fractal dimension using pixel range calculation

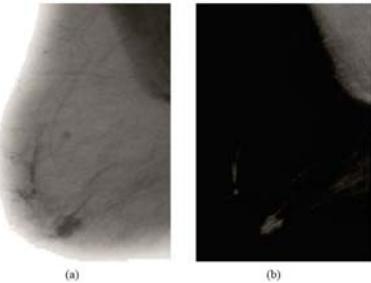


Figure 4: (a) Inverted image (b) Subtracted image.

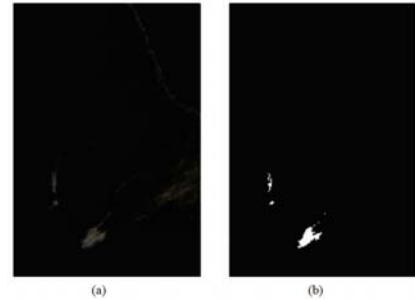


Figure 5: (a) Pectoral muscle removal process (b) Segmented image.

methodology, and three area-based features, eccentricity, solidity, and extent, are derived from the segmented suspicious mass region. The feature extraction process ends after the completion of the extraction of the above features for all the mammograms in the Mini-MIAS dataset.

As we mentioned earlier, the number of mammograms coming from the malignant class is much lower than those from the benign class. To fix this weakness, SMOTE algorithm was used to produce a balanced dataset. A random forest algorithm is used to classify the mammograms. Sensitivity, specificity, accuracy, and the ROC curve are all used to assess the classification's efficiency as shown in Fig. 7. The percentage of actual positives that are correctly identified as malignant is measured by sensitivity. The proportion of actual negatives correctly identified as benign is measured by specificity.

Table 1 shows the performance dimension of the proposed algorithm. The results are promising, with a 97.1% overall accuracy. It performed well in detecting both benign and malignant, with sensitivity and specificity values of 95.1% and 98.5%, respectively. Table 2 compares the proposed algorithm's efficiency with the various other existing approaches in this domain.

Table 1: The performance measures of the proposed algorithm.

Features	Sensitivity	Specificity	Accuracy
Moment invariant, fractal dimension and Region-based features	95.1	98.5	97.1

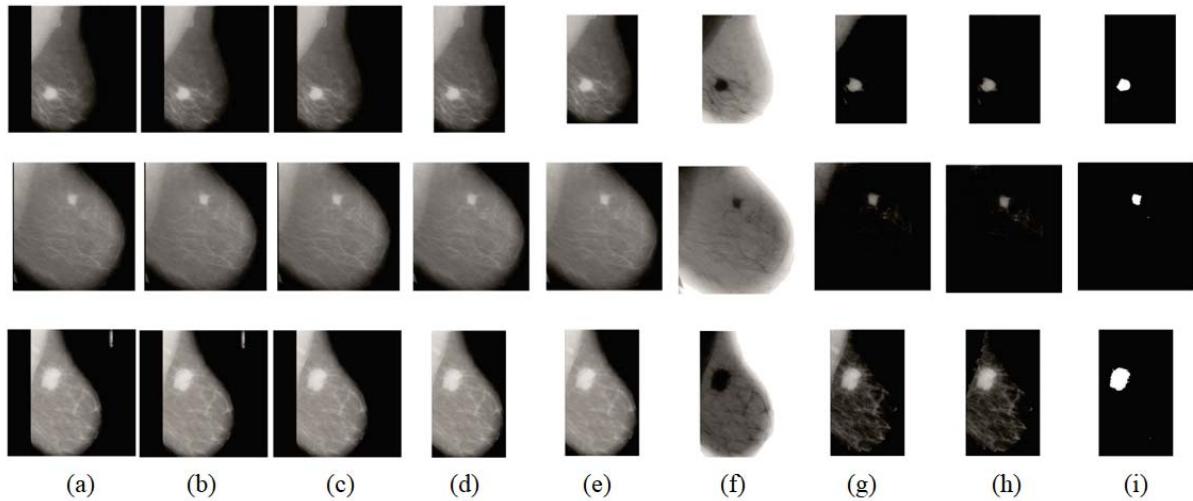


Figure 6: Segmentation of suspicious regions outputs: a) Original image , b) Binarized image , c) Morphological erosion, d) First step of the breast region segmentation, e) Second step of the breast region segmentation, f) Inverted image, g) Subtracted image, h) Pectoral muscle removal process, i) Segmented image.

Table 2: Comparison with existing methods.

Methods	Dataset	Features	Classifier	Accuracy
Jaleel et al. [12]	Mini MIAS	Discrete Wavelet Transform and GLCM	ANN	93.7% with GLCM and 94.6% with DWT
Kashyap et al. [15]	Mini MIAS	Moment Invariant, fractal dimension	SVM	96.92%
Xie et al. [34]	Mini MIAS+DDSM	Gray level features and textural features	ELM and SVM	96.02%
Kaur et al. [16]	Mini-MIAS	SURF	SVM, KNN, LDA and DT	96.9%, 93.8%, 89.7% and 88.7%
Kamil et al. [14]	Mini-MIAS	Gray Level Co-occurrence Matrix	KNN	86.1%
Proposed model	Mini-MIAS	Moment Invariant, fractal dimension and Region-based	RF	97.1%

## 6 Conclusion

The proposed approach includes an algorithm for enhancing, segmenting, and classifying abnormalities identified in mammograms of the Mini-MAIS dataset. The identification of suspicious and pectoral muscle regions in mammograms is improved by subtracting the preprocessed and inverted preprocessed images. To extract pectoral muscle from mammogram images, a region growing with the K-means clustering method is presented. In most of the images, the triangular area specified over an image covers the pectoral muscle. The approach produces a very good result to removes pectoral muscles as shown in the results section. For most classifiers, it is effective in reducing the class-imbalance problem. Because of that, SMOTE has been used for producing synthetic samples that can help to mitigate the issue of class imbalance. The experiments demonstrate that the proposed method that includes (noise reductions, breast region differentiation, filtering, pectoral muscle removal, segmentation, feature extraction, balancing features) obtained an accuracy of 97.1%, the sensitivity achieved is 95.1%, and the speci-

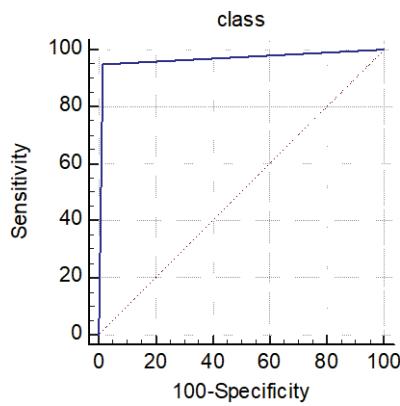


Figure 7: ROC curve of the classification result.

ficity achieved is 98.5%, over other popular methods in recent literature.

## References

- [1] How common is breast cancer?: Breast cancer statistics. <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer/>. Accessed: 2021-09-10.
- [2] BIAU, G., AND SCORNET, E. A random forest guided tour. *Test* 25, 2 (2016), 197–227.
- [3] CHAWLA, N. V., BOWYER, K. W., HALL, L. O., AND KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.
- [4] CHEN, Q., PETRIU, E., AND YANG, X. A comparative study of fourier descriptors and hu's seven moment invariants for image recognition. In *Canadian conference on electrical and computer engineering 2004 (IEEE Cat. No. 04CH37513)* (2004), vol. 1, IEEE, pp. 103–106.
- [5] FERNÁNDEZ, A., GARCIA, S., HERRERA, F., AND CHAWLA, N. V. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research* 61 (2018), 863–905.
- [6] GALAR, M., FERNANDEZ, A., BARRENECHEA, E., BUSTINCE, H., AND 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)* 42, 4 (2011), 463–484.
- [7] GOWRI, D. S., AND AMUDHA, T. A review on mammogram image enhancement techniques for breast cancer detection. In *2014 International Conference on Intelligent Computing Applications* (2014), IEEE, pp. 47–51.
- [8] HAZARIKA, M., AND MAHANTA, L. B. A novel region growing based method to remove pectoral muscle from mlo mammogram images. In *Advances in Electronics, Communication and Computing*. Springer, 2018, pp. 307–316.
- [9] HOO, Z. H., CANDLISH, J., AND TEARE, D. What is an roc curve?, 2017.
- [10] HUSSEIN, A. S., LI, T., YOHANNESE, C. W., AND BASHIR, K. A-smote: A new preprocessing approach for highly imbalanced datasets by improving smote. *International Journal of Computational Intelligence Systems* 12, 2 (2019), 1412–1422.
- [11] ISHWARAN, H., AND O'BRIEN, R. Commentary: the problem of class imbalance in biomedical data. *The Journal of thoracic and cardiovascular surgery* 161, 6 (2021), 1940.
- [12] JALEEL, J. A., SALIM, S., AND ARCHANA, S. Textural features based computer aided diagnostic system for mammogram mass classification. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (2014), IEEE, pp. 806–811.
- [13] JESNECK, J. L., LO, J. Y., AND BAKER, J. A. Breast mass lesions: computer-aided diagnosis models with mammographic and sonographic descriptors. *Radiology* 244, 2 (2007), 390–398.
- [14] KAMIL, M. Y., AND JASSAM, A.-L. A. Analysis of tissue abnormality in mammography images using gray level co-occurrence matrix method. In *Journal of Physics: Conference Series* (2020), vol. 1530, IOP Publishing, p. 012101.
- [15] KASHYAP, K. L., BAJPAI, M. K., AND KHANNA, P. Breast cancer detection in digital mammograms. In *2015 IEEE international conference on imaging systems and techniques (IST)* (2015), IEEE, pp. 1–6.
- [16] KAUR, P., SINGH, G., AND KAUR, P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class svm using deep learning classification. *Informatics in Medicine Unlocked* 16 (2019), 100151.
- [17] KELLY, K. M., DEAN, J., COMULADA, W. S., AND LEE, S.-J. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European radiology* 20, 3 (2010), 734–742.
- [18] KHOSRAVY, M., GUPTA, N., MARINA, N., SETHI, I. K., AND ASHARIF, M. R. Morphological filters: An inspiration from natural geometrical erosion and dilation. In *Nature-inspired computing and optimization*. Springer, 2017, pp. 349–379.
- [19] KRAWCZYK, B. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence* 5, 4 (2016), 221–232.
- [20] MAKANDAR, A., AND HALALLI, B. Combined segmentation technique for suspicious mass detection in mammography. In *2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)* (2015), IEEE, pp. 1–5.
- [21] OSHIRO, T. M., PEREZ, P. S., AND BARANAUSKAS, J. A. How many trees in a random forest? In *International workshop on machine learning and data mining in pattern recognition* (2012), Springer, pp. 154–168.
- [22] PATIL, R. S., AND BIRADAR, N. Improved region growing segmentation for breast cancer detection: progression of optimized fuzzy classifier. *International Journal of Intelligent Computing and Cybernetics* (2020).
- [23] PRATIWI, M., HAREFA, J., NANDA, S., ET AL. Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network. *Procedia Computer Science* 59 (2015), 83–91.

- [24] PUNITHA, S., AMUTHAN, A., AND JOSEPH, K. S. Benign and malignant breast cancer segmentation using optimized region growing technique. *Future Computing and Informatics Journal* 3, 2 (2018), 348–358.
- [25] QUIST, J., TAYLOR, L., STAAF, J., AND GRIGORIADIS, A. Random forest modelling of high-dimensional mixed-type data for breast cancer classification. *Cancers* 13, 5 (2021), 991.
- [26] RADHI, E. A., AND KAMIL, M. Y. Breast tumor detection via active contour technique. *International Journal of Intelligent Engineering and Systems* 14, 4 (2021), 561–570.
- [27] RANGANATH, A., SENAPATI, M. R., AND SAHU, P. K. Estimating the fractal dimension of images using pixel range calculation technique. *The Visual Computer* 37, 3 (2021), 635–650.
- [28] ROEMER, V., AND WALDEN, R. Sensitivity, specificity, receiver-operating characteristic (roc) curves and likelihood ratios for electronic foetal heart rate monitoring using new evaluation techniques. *Zeitschrift für Geburtshilfe und Neonatologie* 214, 03 (2010), 108–118.
- [29] SHARMA, S., AND KHANNA, P. Roi segmentation using local binary image. In *2013 IEEE International Conference on Control System, Computing and Engineering* (2013), IEEE, pp. 136–141.
- [30] SHARMA, S., AND MEHRA, R. Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. *Journal of digital imaging* 33, 3 (2020), 632–654.
- [31] SHIRAIHI, J., LI, Q., APPELBAUM, D., AND DOI, K. Computer-aided diagnosis and artificial intelligence in clinical imaging. In *Seminars in nuclear medicine* (2011), vol. 41, Elsevier, pp. 449–462.
- [32] SUCKLING J, P. The mammographic image analysis society digital mammogram database. *Digital Mammo* (1994), 375–386.
- [33] TOU, J. T. Computer-based particle shape analysis for classification, recognition, utilization. In *Advanced Particulate Morphology*. CRC Press, 2018, pp. 165–170.
- [34] XIE, W., LI, Y., AND MA, Y. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* 173 (2016), 930–941.
- [35] ZHI, H., OU, B., LUO, B.-M., FENG, X., WEN, Y.-L., AND YANG, H.-Y. Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. *Journal of ultrasound in medicine* 26, 6 (2007), 807–815.