

EXTENDED MAMMOGRAM CLASSIFICATION FROM TEXTURAL FEATURES

ADÉL BAJCSI, CAMELIA CHIRA, AND ANCA ANDREICA

ABSTRACT. The efficient analysis of digital mammograms has an important role in the early detection of breast cancer and can lead to a higher percentage of recovery. This paper presents an extended computer-aided diagnosis system for the classification of mammograms into three classes (normal, benign and malignant). The performance of the system is evaluated for two different mammogram databases (MIAS and DDSM) in order to assess its robustness. We discuss the changes required in the system, particularly at the level of the image preprocessing and feature extraction. Computational experiments are performed based on different methods for feature extraction, selection and classification. The results indicate an accuracy of 66.95% for the MIAS dataset and 54.1% for DDSM obtained using genetic algorithm based feature selection and Random Forest classification.

1. INTRODUCTION

Computer-aided detection and diagnosis (CAD) relies on medical image processing, being used nowadays for a big variety of diseases, including breast cancer. In the current research, our aim is to create a CAD system able to detect breast cancer from mammograms. Input images are used from two different mammogram datasets (Mammographic Image Analysis Society – MIAS [15], Digital Database for Screening Mammography – DDSM [7]) in order to verify a stable performance for the stages of preprocessing, feature extraction and classification.

Received by the editors: 20 September 2022.

2010 *Mathematics Subject Classification.* 68T35.

1998 *CR Categories and Descriptors.* 1.2.1 [**Artificial Intelligence**]: Applications and Expert Systems – *Medicine and science*; 1.2.6 [**Artificial Intelligence**]: Learning – *Knowledge acquisition*; 1.4.7 [**Image Processing and Computer Vision**]: Feature Measurement – *Feature representation*;

Key words and phrases. Breast cancer detection, Mammogram classification, GLRLM, Feature selection, Random Forests, MIAS, DDSM.

A multi-class classifier is built to distinguish normal, benign and malignant classes. We start from a recently proposed CAD system [2], which consists of the following five steps:

(1) mammogram preprocessing, (2) segmentation of the image, (3) feature extraction (FE) by calculating characteristics from GLRLM (Gray-Level Run-Length Matrix [6]), (4) feature selection (FS) and (5) classification (CLS).

In reference [2], multiple experiments are reported with competitive results for medical images from the MIAS database [15] using Principal Component Analysis (PCA) for feature selection and Random Forest (RF) classifiers. As the system was created and tested for MIAS images, our main objective in this study is to determine if the system can be extended to obtain a good performance for other medical images. For DDSM images, the system is likely to generate errors because the intensity of the mammograms is different compared to images in MIAS. Also, images from DDSM have a high variance in their intensities and white patches could be found on DDSM entries, connected to the breast.

Considering the abovementioned issues, in the current paper we extend the previously proposed system mainly concerning the preprocessing and feature extraction steps. The preprocessing step is extended by

(1) removing white patches outside the breast, (2) setting the threshold to define the breast dynamically, instead of hard-coding it, (3) creating a more robust method to define the location of the pectoral muscle, (4) selecting the proper component overlapping with the pectoral muscle and (5) filling its concave parts.

Moreover, feature extraction considers characteristics calculated from Gray-Level Co-Occurrence Matrices (GLCM) with different distances and is reimplemented to use the graphical processing unit. For segmentation, feature selection, and classification, the same methods are used as in [2], namely, k-means, PCA and genetic algorithm (GA) based feature selection, and Decision Tree (DT) and RF respectively.

The rest of the article is organized as follows: in section 2 we will discuss existing solutions in the literature to solve the problem. In section 3 the proposed approach is presented in detail. Section 4 describes the data used and the results achieved. In section 5 we draw the conclusions of the conducted research and the future directions are defined.

2. RELATED WORK

With the development of image processing techniques and online (freely) accessible databases, experiments for image analysis are being conducted in

various fields. One of these fields is medicine. With the help of machines, illnesses can be easier detected and confirmed by doctors. Breast cancer is one of these illnesses being studied. The problem of detecting cancer from mammograms is highly important and complex because of the difference in the images created with different machines, the size of the mammograms, and the appropriate selection of the used features. In the following paragraphs, we will present existing methods in the literature to detect breast cancer.

In [13] a comprehensive review is presented on the methods used from 1998 to 2018. In the review 129 papers were included. As the review highlights, there are three major types of experiments:

- (1) studies where only normal and abnormal classes are considered [16]
- (2) where benign and malignant classes are considered [1, 4, 4, 6, 8, 8] and
- (3) where all three classes (normal, benign and malignant) are considered [10, 12, 14].

The authors concluded that the most used classifiers were Artificial Neural Networks (ANN), Support Vector Machines (SVM) and K-Nearest Neighbors (k-NN). As for evaluation metrics most of the researchers used accuracy, area under curve, sensitivity and/or specificity.

Reference [5] conducted an experiment to compare several feature extraction methods, namely, First Order Statistic (FOS), Gray Level Co-Occurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GLRLM), and Gray Level Difference Matrix (GLDM). The authors concluded that the best result (93.98% accuracy) was achieved using the GLRLM features for building ECOC SVM classifiers.

Another survey [6] focused on various feature extraction and selection methods. Besides the four features mentioned before, the authors analyzed Tamura features, Gabor features, Wavelet transform features, Hu's invariant moments features, and other shape features such as perimeter, area, compactness, aspect ratio, and so on. For feature selection, Tabu search, Genetic algorithm, ReliefF algorithm, and Sequential forward/backward selection are included. The authors [6] showed that for building a classifier, GLRLM features are the most appropriate. Using these features 66.66% and 90.9% respective to AUC and precision were achieved [6].

In [1] is reported 88% for both accuracy and AUC using classification with Neural Networks. As input the authors proposed using a concatenation of different Convolutional Neural Networks (AlexNet, VGG16, GoogLeNet, ResNet18, InceptionResNet). In reference [10] a novel approach is proposed to solve the problem of classifying the mammograms into normal, benign and malignant classes. Two types of features are extracted from the images:

- (1) low-level features calculated using ResNet and transfer learning and
- (2) high-level features calculated using RNN-LSTM.

From the resulting feature set, a CNN selects the relevant features, then the final classification is made by an ensemble learning model (using RF learning and boosting). With this method, the authors achieved 96% accuracy. In [16] ANNs are used for feature extraction (DenseNet and MobileNet) and fed to a fully connected network. With this approach, accuracy of 96.34% was achieved using DenseNet features.

In [4] the application of super-resolution is presented to better distinguish abnormal masses. With this method, the authors boosted the result and achieved 96.7% accuracy. The authors used a combination of FOS and GLRLM features as input to nine different classifiers. Reference [14] proposes the extraction of features from the spatial pyramid and called Pyramid Histogram of Colors. Feeding an ANN with the calculated features, 82.1% AUC was achieved. In [12] a Forest Optimized Algorithm is proposed to select features from characteristics calculated from GLCM and wavelet transform. For classification, the authors used SVM, k-NN and Decision Trees (DTs).

Reference [8] presents a Fuzzy Rule-Based interpolative classifier for making differences between benign and malignant lesions. The input of the classifier consists of 18 features included shape features (shape and morphological characteristics), margin features (sharpness and roughness of the boundary), and density features. With this method, the authors get accuracy of 91.65%.

3. PROPOSED APPROACH

In our research, we extended the approach presented in [2] in order to be able to apply it to other mammogram datasets. The system in [2] is not working accordingly if the breast tissue has too low intensity or if white patches appear on the image. In this section, we present the original CAD system proposed in [2] and then detail the method's extension and the main improvements proposed in the current paper.

3.1. The base CAD system. As already mentioned, the CAD system introduced in [2] has five main steps: preprocessing, segmentation, feature extraction, selection and classification. Essential aspects for a successful computer-aided diagnosis include calculating the features and classification of the observations. The most important is to properly clear the image (remove the information – pixel –, which is not important from the problem's perspective). For instance, to diagnose breast cancer from medical images (such as mammograms), everything outside the breast tissue and the pectoral muscle is irrelevant. The pectoral muscle is not a possible location of a lesion, leading to its disregardment during the image analysis process.

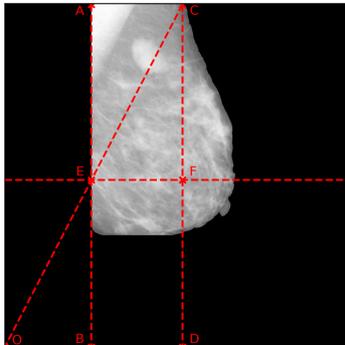


FIGURE 1. Defining the possible location of the pectoral muscle on image mdb015 from MIAS.

In the first step, the tissue is separated from the rest of the image. To get a precise result, noise reduction (morphological opening) and emphasis (histogram equalisation) are applied. To remove the pectoral muscle, the Seeded Region Growing (SRG) algorithm, proposed by Maitra et al. [9], is used by selecting the pixels corresponding to the muscle. The requirement of this approach is that the pectoral muscle must be on the left side of the image. To decrease the image to the possible location of the muscle first, the left- (A) and rightmost bounds (C) of the breast at the top (vertical lines across the left-/rightmost points – lines AB and CD on image fig. 1) are defined, and then the top of the right bounding line is connected to the lower-left corner (O) of the image. As a result, a triangle should be formatted (on the top between the two bounding lines – ACE), containing the pectoral muscle. For this triangle, the SRG was applied using as seeds the pixels from the section between the right angle and the bisector of the hypotenuse. The final step of the preprocessing is to selecting a smaller bounding box surrounding the breast. Next (segmentation step), k-means was performed to segment the images into 12 clusters. During the feature extraction, from both the original and the segmented image GLRLMs are calculated and characteristics are extracted (resulting in 44 features per image – 11 features \times 4 directions). Following with the feature selection step, which aims to reduce the size of the input. For this purpose, two algorithms are implemented: PCA and GA. Later, the result of the feature selection method is passed to the classification. To differentiate normal, benign and malignant classes, DT and RF models are built.

3.2. Extended method.

When introducing new images to the system, presented in the previous subsection, there could be cases when the execution differs from the expected behaviour because it was specified for mammograms from a specific dataset. For instance, referring to images from DDSM, there are cases when the intensity of the breast tissue is low, and the tissue disappears after the thresholding, or the position of the pectoral muscle is misdected. The most vulnerable part is the preprocessing. This subsection details the changes made in the previously presented system.

On the side(s) of some images (from DDSM), a white column/row could appear. Previously, the position of the pectoral muscle was defined based on the longest vertical line found with Hough transform with an average intensity lower than 180 or if no line was found based on the pixel intensities in the upper corners of the mammogram. This method worked for images from MIAS. Even though MIAS mammograms could contain white columns, they are not connected to the breast and therefore they are removed together with the labels. Images from MIAS are squares and the actual mammogram is positioned in the centre. On the other hand, in DDSM, these columns could be mistaken for the pectoral muscle (because no lines will be detected due to the lack of surrounding black columns). Hence, before the preprocessing, every pixel with high intensity (> 240) on the margin of the image will be set to 0.

The first step of the preprocessing is the removal of the labels using thresholding. For images from MIAS the threshold value of 50 was defined experimentally. However, using the same value on the new images led to the disappearance of the breast tissue in some cases (due to its low intensity as on Fig. 2(b)). To overcome this problem, we decided to use a dynamic definition for the threshold values by calculating the median intensity value from the image – I (omitting pixels with 0 value). The result of the new thresholding is shown in Fig. 2(c). Next, we must ensure that the pectoral muscle is on the left side of the image (this is a prerequisite of the used SRG method).

To define the initial orientation of the breast, the intensities in the upper corners are used. We further break this down into two cases. First, we calculate the first and last non-zero positions from the first row. If one of the values is close enough to the side (within 100 pixels), then the location of the muscle is clear. Otherwise, we add zero padding around the image and apply Hough's line transform to determine vertical lines on the image. The longest line found specifies the position of the pectoral muscle. The zero-padding helps to define the border, even if this is exactly on the edge of the image.

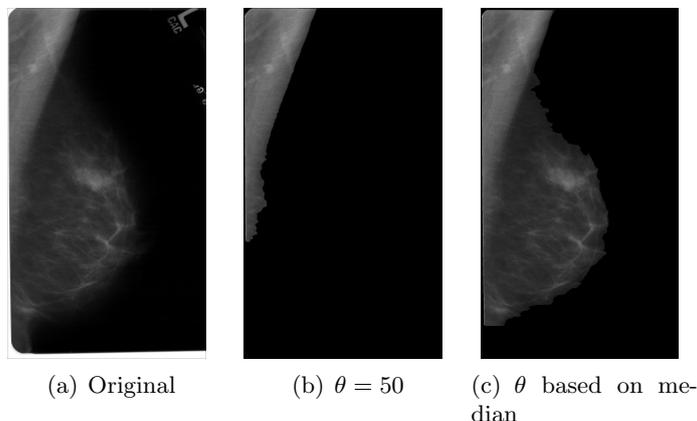


FIGURE 2. Separating foreground and background using thresholding presented on image A-1833-1.LEFT_MLO from DDSM.

The result of SRG, mentioned in the previous section, could contain more (unconnected) components (see Fig. 3(b)). Of these components, we should select the largest and closest to the top left corner of the image (Fig. 3(c)).

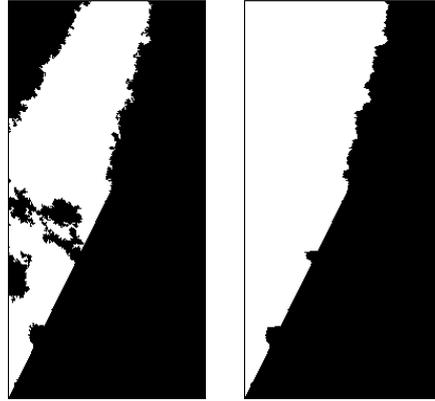
Moreover, we investigate the shape of the different resulting components and found that some of them are concave (Fig. 4(a)). For a more precise removal, we fill these holes in the component (Fig. 4(b)). To accelerate the pipeline, we propose a new feature extraction version that uses a graphical procession unit. In the current research, we had an Nvidia GeForce GTX 960M GPU and achieved $\approx 62\times$ acceleration. Heretofore, the GLRLM was calculated at 265.39s. Even with parallel threads, it takes 14.52s to construct the matrix. However, using GPU, we can get the result matrix in 4.3s (depending on the image's size).

For steps (2), (4) and (5) the same methods are used as in the originally proposed method from [2]. Besides GLRLM features, GLCM features are calculated (for distances 1 - default - and 8) and used for classification. From the constructed (normalized) GLCM, the features are calculated using (3), (4) and (7) to (23), where *level* denotes the number of gray levels on the mammogram. The functions in (1) and (2) help the definition of sum average (18), sum variance (20), sum entropy (19) and difference average (21), difference variance (23), difference entropy (22) respectively.



(a) ROI of pectoral (b) Result of SRG (c) Pectoral mask
(used in the original version)

FIGURE 3. Selecting the mask corresponding to the pectoral muscle from the result of the SRG [9] on image mdb015 from MIAS.



(a) Result of SRG (b) Final mask
(used in the original version)

FIGURE 4. Filling the wholes on the SRG [9] result on image mdb017 from MIAS.

$$(1) \quad p_{x+y}(k) = \sum_{i,j=0}^{levels-1} GLCM_{ij}, \quad (2) \quad p_{x-y}(k) = \sum_{i,j=0}^{levels-1} GLCM_{ij},$$

where $i + j = k$ where $|i - j| = k$

$$\begin{aligned}
 (3) \quad \mu_x &= \sum_{i,j=0}^{levels-1} iGLCM_{ij} & (4) \quad \sigma_x^2 &= \sum_{i,j=0}^{levels-1} (i - \mu_x)^2 GLCM_{ij} \\
 (5) \quad \mu_y &= \sum_{i,j=0}^{levels-1} jGLCM_{ij} & (6) \quad \sigma_y^2 &= \sum_{i,j=0}^{levels-1} (j - \mu_y)^2 GLCM_{ij} \\
 (7) \quad ent &= - \sum_{i,j=0}^{levels-1} GLCM_{ij} \ln GLCM_{ij} & (8) \quad cont &= \sum_{i,j=0}^{levels-1} (i - j)^2 GLCM_{ij} \\
 (9) \quad ASM &= \sum_{i,j=0}^{levels-1} GLCM_{ij}^2 & (10) \quad energy &= \sqrt{ASM} \\
 (11) \quad diss &= \sum_{i,j=0}^{levels-1} |i - j| GLCM_{ij} & (12) \quad corr &= \sum_{i,j=0}^{levels-1} \frac{(i - \mu_x)(j - \mu_y)}{\sqrt{\sigma_x^2 \sigma_y^2}} GLCM_{ij} \\
 (13) \quad IDM &= \sum_{i,j=0}^{levels-1} \frac{1}{1 + (i - j)^2} GLCM_{ij} & (14) \quad sim &= \sum_{i,j=0}^{levels-1} \frac{1}{1 + |i - j|} GLCM_{ij} \\
 (15) \quad DM &= \sum_{i,j=0}^{levels-1} \sqrt{\frac{|i - j|}{2}} GLCM_{ij} & (16) \quad CP &= \sum_{i=0}^{levels-1} (i + j - \mu_x - \mu_y)^4 GLCM_{ij} \\
 (17) \quad CS &= \sum_{i=0}^{levels-1} (i + j - \mu_x - \mu_y)^3 GLCM_{ij} & (18) \quad SA &= \sum_{i=2}^{2levels} ip_{x+y}(i) \\
 (19) \quad SE &= - \sum_{i=2}^{2levels} p_{x+y}(i) \ln p_{x+y}(i) & (20) \quad SV &= \sum_{i=2}^{2levels} (i - SA)^2 p_{x+y}(i) \\
 (21) \quad DA &= \sum_{i=0}^{levels-1} ip_{x-y}(i) & (22) \quad DE &= - \sum_{i=0}^{levels-1} p_{x-y}(i) \ln p_{x-y}(i)
 \end{aligned}$$

$$(23) \quad DV = \sum_{i=0}^{levels-1} (i - DA)^2 p_{x-y}(i)$$

4. EXPERIMENTS AND RESULTS

In this section, we present the details of our experiments, starting with the used datasets, the conditions of the splitting (into train and test) the datasets and the results of our research.

4.1. Datasets. MIAS [15] is a database of MLO (mediolateral oblique view) mammograms published in 1994. It contains 322 (161 pairs) images. Of these, 207 are from normal tissues, 64 are from tissues containing benign lesions and the remaining 51 contain malignant lesions.

DDSM [7] is another database that was first published in 1998. Over the years it was further developed. The database consists of 7808 mammograms from approximately 1950 patients. Compared to MIAS, DDSM contains both MLO and CC (craniocaudal view) view images. From the total images, we will use only the MLO ones, so 3904 images. The distribution of the observations is as follows: 1204 normal, 1342 benign and 1358 malignant.

4.2. Experiment setup. Each dataset is split into train and test sets containing 75% and 25% of the data. When creating the split, we maintain the same distribution in the result sets as in the original data (stratified sampling). Also, to reduce the possible bias of the classification, we place the observations from one person into the same set. Considering these, from MIAS, there will be 242 mammograms in the train set and 80 in the test set. As for the DDSM 2928 mammograms will be in the train set and the remaining 976 in the test set.

To define the optimal parameters, K-fold cross-validation was applied. The best parameter is defined based on the highest mean accuracy across the models built during cross-validation. In these experiments 5-folds are used.

As shown above, the datasets are imbalanced. Therefore, stratification is used to preserve the original distribution of the classes in each fold. With the use of stratified cross-validation, the average of the result metrics will be a close approximation to the result on the original sets.

For evaluation purpose accuracy (A), precision (P), recall (R) and f1-score (F_1) measures are used. As also shown in [13], the first three metrics are frequently used in studies related to breast cancer detection or diagnosis.

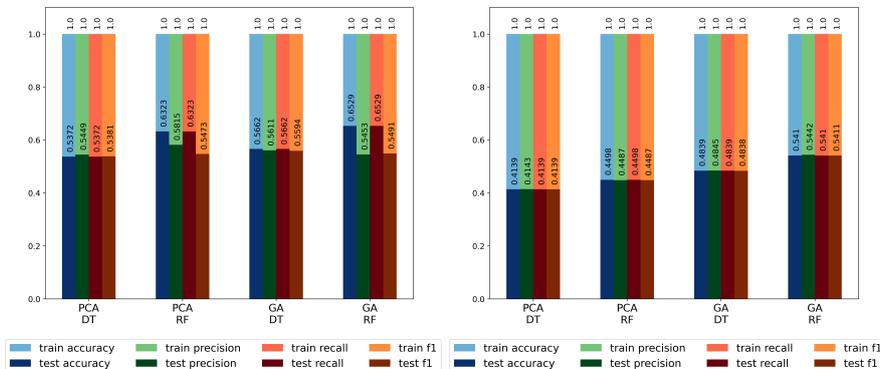


FIGURE 5. Result metrics using GLRLM features calculated from MIAS images.

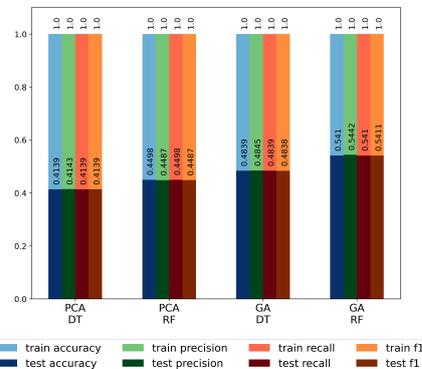


FIGURE 6. Result metrics using GLRLM features calculated from DDSM images.

4.3. Results. The results of the classifications are presented in Tables 1 and 2 and figs. (5) to (10). In columns A , P , R and F_1 of the tables, the mean values are reported over the folds with their corresponding standard deviation. In Table 1 and Table 2 only the test results are presented. During the training of the models, we achieved 100% (as show on Figs. (5) to (10)) with each parameter mentioned in the setup.

From Table 1, we can see that the results on MIAS are slightly different from the ones reported in [2]. This can be explained with the changes mentioned in Section 3.2. However, these modifications are necessary in order to use the proposed method on images from DDSM.

The best results for MIAS are obtained using $GLCM_8$ (for step (3)), GA feature selection (for step (4)) and RF classifier (for step (5)). With k-fold cross-validation we reached 66.95% average test accuracy. For the extended method, the average test accuracy using GLRLM, PCA and RF was 63.23%. The made in the preprocessing can cause the difference between the results of the original [2] and extended system. Now each image has the same orientation (previously the flip and the pectoral removal did not work on some images) and by applying CLAHE (contrast limited adaptive histogram equalization) filtering to the images. The different split, explained above, can also affect the result.

Table 2 presents the results obtained on DDSM images using the extended system proposed in this paper. The best classification result on DDSM is 54.1% for average test accuracy, when classifying the observations into three classes using features calculated from GLRLM, GA feature selection and RF classification.

FE	FS	Dir. ^a	CLS	A	P	R	F_1
GLRLM	PCA	↖	DT	0.5372 ±0.0478	0.5449 ±0.0255	0.5372 ±0.0478	0.5381 ±0.0346
		↘	RF	0.6323 ±0.0184	0.5815 ±0.0852	0.6323 ±0.0184	0.5473 ±0.0310
	GA	↑	DT	0.5662 ±0.0469	0.5611 ±0.0478	0.5662 ±0.0469	0.5594 ±0.0433
		↖	RF	0.6529 ±0.0239	0.5453 ±0.1135	0.6529 ±0.0239	0.5491 ±0.0319
GLCM	PCA	↖	DT	0.5373 ±0.0641	0.5112 ±0.0818	0.5373 ±0.0641	0.5099 ±0.0543
		↑	RF	0.5869 ±0.0497	0.5023 ±0.0721	0.5869 ±0.0497	0.5288 ±0.0545
	GA	↖	DT	0.5949 ±0.0543	0.5880 ±0.0660	0.5949 ±0.0543	0.5785 ±0.0492
		↑	RF	0.6491 ±0.0541	0.5573 ±0.0718	0.6491 ±0.0541	0.5681 ±0.0555
GLCM ₈	PCA	↖	DT	0.5327 ±0.0415	0.5326 ±0.0170	0.5327 ±0.0415	0.5301 ±0.0271
		↘	RF	0.5910 ±0.0415	0.4733 ±0.0608	0.5910 ±0.0415	0.5143 ±0.0462
	GA	↖	DT	0.5538 ±0.0569	0.5581 ±0.0564	0.5538 ±0.0569	0.5512 ±0.0553
		↖	RF	0.6695 ±0.0214	0.6318 ±0.0640	0.6695 ±0.0214	0.5938 ±0.0171

^a the column marks the direction of the used features.

TABLE 1. Test result metrics for MIAS

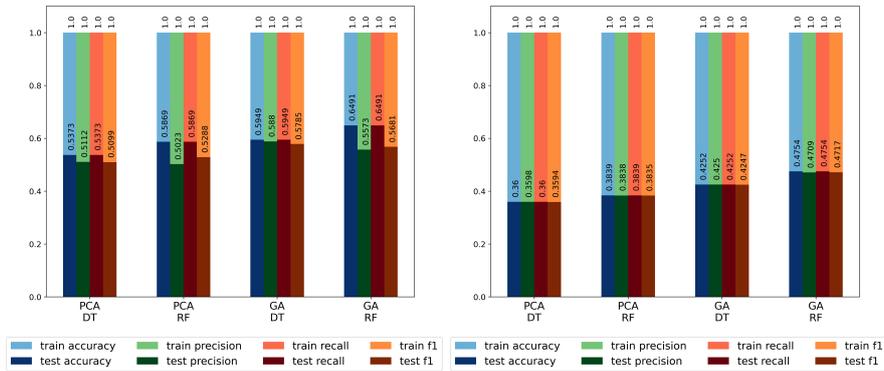


FIGURE 7. Result metrics using GLCM features calculated from MIAS images.

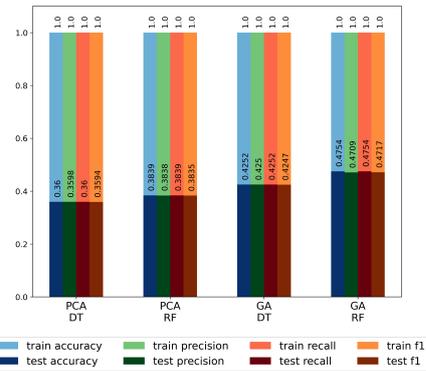


FIGURE 8. Result metrics using GLCM features calculated from DDSM images.

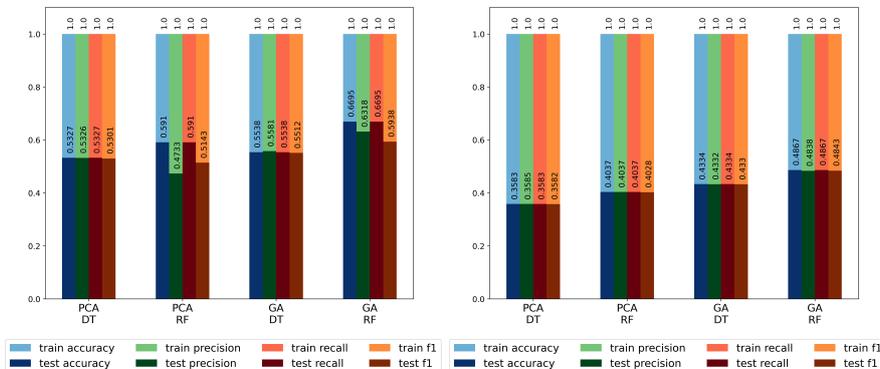


FIGURE 9. Result metrics using GLCM₈ features calculated from MIAS images.

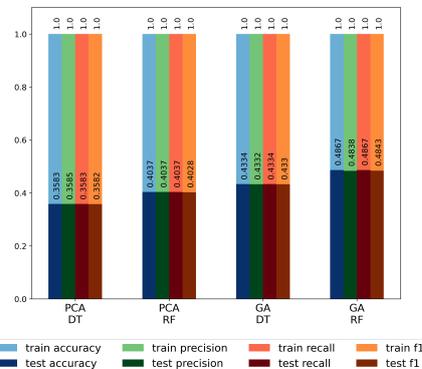


FIGURE 10. Result metrics using GLCM₈ features calculated from DDSM images.

From Table 1 (and Figures 5, 7 and 9) we can also see that the results using GA and RF are very close for all three feature extraction methods. However, in Table 2 (and Figures 6, 8 and 10) the results from GLCM and GLCM₈ are worse, hence, concluding the advantage of using GLRLM (run-length information) over GLCM (pixel correlation). When comparing the results of the two classifiers (DT/RF) we can see that a single DT obtains a lower accuracy value than RF. By using more models (DTs in the forest) we can increase the performance of the classification.

4.4. Discussion. In the previous subsection, the results of the proposed extended method have been compared to the system proposed in [2]. We now focus on a comparative analysis between the obtained results for MIAS and DDSM and relevant results from the literature.

In [3], the authors report 60.7% test accuracy using an approach based on GLCM features, correlation feature selection and RF classification (on MIAS). In our research, we achieved higher accuracy which can be explained by the different feature selection (GA) used in the proposed approach.

In [11] presents a comparison of different combinations of feature selection and classification (evaluated on MIAS). The best result reported was 70.53% using Local Binary Pattern and Deep Neural Network. This is comparable with our result. Results achieved with homogeneity, energy, HOG features and DNN, NB, NN, SVM classification are also reported in [11] with accuracies between 42.6% and 59.6%. These methods are outperformed by the proposed extended method.

FE	FS	Dir. ^b	CLS	A	P	R	F_1
GLRLM	PCA	↖	DT	0.4139 ±0.0255	0.4143 ±0.0252	0.4139 ±0.0255	0.4139 ±0.0254
		↗	RF	0.4498 ±0.0234	0.4487 ±0.0247	0.4498 ±0.0234	0.4487 ±0.0243
	GA	↘	DT	0.4839 ±0.0191	0.4845 ±0.0185	0.4839 ±0.0191	0.4838 ±0.0187
		↙	RF	0.5410 ±0.0248	0.5442 ±0.0225	0.5410 ±0.0248	0.5411 ±0.0239
GLCM	PCA	↘	DT	0.3600 ±0.0106	0.3598 ±0.0113	0.3600 ±0.0106	0.3594 ±0.0112
		↗	RF	0.3839 ±0.0175	0.3838 ±0.0182	0.3839 ±0.0175	0.3835 ±0.0180
	GA	→	DT	0.4252 ±0.0098	0.4250 ±0.0097	0.4252 ±0.0098	0.4247 ±0.0098
		↖	RF	0.4754 ±0.0124	0.4709 ±0.0123	0.4754 ±0.0124	0.4717 ±0.0127
GLCM ₈	PCA	↖	DT	0.3583 ±0.0034	0.3585 ±0.0035	0.3585 x±0.0034	0.3582 ±0.0034
		↗	RF	0.4037 ±0.0231	0.4037 ±0.0233	0.4037 ±0.0231	0.4028 ±0.0234
	GA	→	DT	0.4334 ±0.0260	0.4332 ±0.0266	0.4334 ±0.0260	0.4330 ±0.0263
		↙	RF	0.4867 ±0.0229	0.4838 ±0.0225	0.4867 ±0.0229	0.4843 ±0.0224

^b the column marks the direction of the used features.

TABLE 2. Test result metrics for DDSM

In [10], the authors presented the results of CNNs and reported 95% accuracy for classifying images in MIAS into three classes and 96% on the DDSM. From this, we can see that with this basic classification further improvements are necessary to outperform ANNs.

5. CONCLUSIONS AND FUTURE WORK

The scope of the current paper is to apply the CAD system presented in [2] to images from DDSM. However, some changes are needed to adapt the previously presented method to the new images. To determine the best parameters 5-fold cross-validation is applied. Based on the presented results we can see an overfitting of the classification (100% train metrics while around 60% test metrics). The extended system achieved 54.1% accuracy on DDSM using GLRLM, GA and RF. On MIAS, the best result is 66.95% from GLCM₈, applying GA

and RF. They are comparative with related results from the literature. However, the results are lower than expected, therefore the system needs further investigation.

As highlighted in [13] building a classifier using images from more datasets can lead to a more robust solution. Hence, in future work, we will investigate the result of the proposed solution on a combined input using images from MIAS and DDSM at the same time. Also, according to the same review, Random Forests and Decision Trees are least frequently used than other classifiers, such as Support Vector Machines and Artificial Neural Networks. Thus, we will dive deeper into these classifiers in order to explore their potential for breast cancer detection. Likewise, we will experiment building two separate classifiers that distinguish: (1) normal and abnormal and (2) benign and malignant tissues. In future work it would be worth experimenting some typical failure cases and interpretation of behaviour.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian Ministry of Education and Research, CCCDI - UEFISCDI, project number PN-III-P2-2.1-PED-2019-2607, within PNCDI III.

REFERENCES

- [1] ARORA, R., RAI, P. K., AND RAMAN, B. Deep feature-based automatic classification of mammograms. *Medical & Biological Engineering & Computing* 58, 6 (June 2020), 1199–1211.
- [2] BAJCSI, A., ANDREICA, A., AND CHIRA, C. Towards feature selection for digital mammogram classification. *Procedia Computer Science* 192 (2021), 632–641.
- [3] BEKTAS, B., EMRE, I. E., KARTAL, E., AND GULSECEN, S. Classification of mammography images by machine learning techniques. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)* (2018), pp. 580–585.
- [4] BOUDRAA, S., MELOUAH, A., AND MEROUANI, H. F. Improving mass discrimination in mammogram-cad system using texture information and super-resolution reconstruction. *Evolving Systems* 11, 4 (Dec 2020), 697–706.
- [5] CANDRA, D., NOVITASARI, R., LUBAB, A., SAWIJI, A., AND ASYHAR, A. H. Application of feature extraction for breast cancer using one order statistic, GLCM, GLRLM, and GLDM. *Advances in Science, Technology and Engineering Systems Journal* 4, 4 (2019), 115–120.
- [6] CHAIEB, R., AND KALTI, K. Feature subset selection for classification of malignant and benign breast masses in digital mammography. *Pattern Analysis and Applications* 22, 3 (Aug. 2019), 803–829.
- [7] HEATH, M., BOWYER, K., KOPANS, D., MOORE, R., AND KEGELMEYER, W. P. The digital database for screening mammography. In *Proceedings of the 5th International Workshop on Digital Mammography* (2001), M. Yaffe, Ed., Medical Physics, pp. 212–218.

- [8] LI, F., SHANG, C., LI, Y., AND SHEN, Q. Interpretable mammographic mass classification with fuzzy interpolative reasoning. *Knowledge-Based Systems 191* (2020), 105279.
- [9] MAITRA, I. K., NAG, S., AND BANDYOPADHYAY, S. K. Technique for preprocessing of digital mammogram. *Computer Methods and Programs in Biomedicine 107, 2* (2012), 175–188.
- [10] MALEBARY, S. J., AND HASHMI, A. Automated breast mass classification system using deep learning and ensemble learning in digital mammogram. *IEEE Access 9* (2021), 55312–55328.
- [11] MATHAPATI, M., CHIDAMBARANATHAN, S., NASIR, A. W., VIMALARANI, G., RANI, S. S., AND GOPALAKRISHNAN, T. An intelligent internet of medical things with deep learning based automated breast cancer detection and classification model. In *Cognitive Internet of Medical Things for Smart Healthcare*. Springer International Publishing, Oct. 2020, pp. 181–193.
- [12] MOHANTY, F., RUP, S., DASH, B., MAJHI, B., AND SWAMY, M. N. S. Digital mammogram classification using 2D-BDWT and GLCM features with FOA-based feature selection approach. *Neural Computing and Applications 32, 11* (June 2020), 7029–7043.
- [13] PEDRO, R. W. D., MACHADO-LIMA, A., AND NUNES, F. L. Is mass classification in mammograms a solved problem? - a critical review over the last 20 years. *Expert Systems with Applications 119* (2019), 90–103.
- [14] SANTOS, G. PHOC descriptor applied for mammography classification. *Revista de Informática Teórica e Aplicada 27, 1* (2020), 26–35.
- [15] SUCKLING, J., PARKER, J., AND DANCE, D. The mammographic image analysis society digital mammogram database. In *International Congress Series* (01 1994), vol. 1069, pp. 375–378.
- [16] XIE, L., ZHANG, L., HU, T., HUANG, H., AND YI, Z. Neural networks model based on an automated multi-scale method for mammogram classification. *Knowledge-Based Systems 208* (2020), 106465.

BABEȘ-BOLYAI UNIVERSITY, FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, 1 MIHAIL KOGĂLNICEANU, CLUJ-NAPOCA 400084, ROMANIA

Email address: adel.bajcsi@ubbcluj.ro

Email address: camelia.chira@ubbcluj.ro

Email address: anca.andreica@ubbcluj.ro