Original Research Paper

A Hybrid Approach of Texture Feature and Gradient Orientation for Computer Aided Diagnosis System Based on Breast Density Classification

[1*]**Nujum Alabdulali**, [2]**Alanod Bin Dris**, [2]**Fatimah Alqahtani and** [2]**Aseel Bin Othman**

[1]*Department of Computer Science, Qassim University, Qassim, Saudi Arabia*
[2]*Department of Computer Science, College of Computer and Information Sciences,*
King Saud University, Riyadh, Saudi Arabia

Abstract: A Computer-Aided Diagnosis (CAD) system can perform an accurate diagnosis and help radiologists by presenting a second opinion about breast density. However, the development of a robust CAD system for breast density classification is still an open problem. In this study, we proposed a CAD system based on hybrid intelligent machine learning technique for automatic classification of breast density on mammogram images. The proposed technique employs gradient orientation pattern HOG and texture descriptor CLBP-HF as features and K Nearest Neighbor (KNN) as classifier. The experiments were carried out on benchmarks public domain MIAS and DDSM datasets. The classification accuracy is 96.4% whereas recall and precision are 96.59 and 96.75% on MIAS dataset. Moreover, the comparison with the state-of-the-art breast density classification methods shows that the proposed method outperforms the existing methods on both MIAS and DDSM datasets, the improvement is significant on both datasets. The proposed method will help radiologists in assessing the breast density, which is important for breast cancer diagnosis.

Keywords: CAD, Histogram of Orientation, HOG, Complete Local Binary Pattern, CLBP, Breast Density

Introduction

The Based on the report of American Cancer Society, breast cancer is the most common type among women worldwide (Siegel *et al*., 2016). Around 15% of the total number of cancers is breast cancer. It is the most common type of cancer (King Faisal Specialist Hospital and Research Centre). Nowadays, breast density mammograms are getting high attention due to its relation to breast cancer risk (Kumar *et al*., 2015). However, Breast density is the amount of fat compared to the amount of tissue in the mammogram image. A dense breast is more likely to develop cancer (CDC, 2020). An early diagnosis provides the doctors a chance to handle it and saves lives (Basu, 2018). Machine learning and pattern recognition techniques can play a role in this area in order to help radiologists to measure the breast tissue.

From another perspective, there are demands for developing Computer Aided Diagnosis (CAD) systems

in order to engage it as a second reader for classifying the abnormalities in the mammogram. Also, it has been proven in many studies the high accuracy results and the significant performance that presented by CAD system (Hiba *et al*., 2016; Li *et al*., 2004; Mustra *et al*., 2010; Vaidehi and Subashini, 2015a). Additionally, CAD system can be used to recognize the degree of density in the breast profile (Vaidehi and Subashini, 2015a). By taking the advantages of CAD system, radiologists then can improve their medical decisions and presented their opinion in breast mammogram images with help of CAD system based on medical images analysis and by using machine learning techniques (Vaidehi and Subashini, 2015a).

In term of learning and recognition of breast diseases, CAD system should have similar abilities to the radiologists. For this reason, development CAD system mainly consists of pattern recognition techniques including machine learning. Pattern recognition is the act

of extracting features from objects (e.g., lesions) in raw data and making a decision based on a classifier output, such as classifying each object into one of the possible categories of various patterns (Apt, 2003). In order to develop the CAD system, there are various techniques that have been summarized in many review purpose (Cram101 Textbook Reviews, 2011).

Generally, it is essential for creating CAD system to make an integration of different image processing operations such as image preprocessing filters, images segmentation, feature extraction and classification (Vaidehi and Subashini, 2015a). Computer Aided Diagnosis (CAD) analyzes mammogram images and classifies them based on the BI-RADS categories to help to discover breast cancer at an early stage and decrease mortality. CAD assists in building a content-based mammogram system. The system tries to automatically classify the mammographic tissue density and retrieve similar tissue density images based on the input (query) mammogram image (Vaidehi and Subashini, 2015b).

The methodology of classifying mammography in automated breast density system is analogical. It usually starts with preprocessing to enhance the noise, remove any labels and produce a clear image. This is followed by segmentation stage where the image is partitioned into its constituent parts or objects. Feature extraction is the next step; there are various texture descriptors each of them has a different usage. Finally, classification process which presents the accuracy of the method is used. Figure 1 presents a simple representation of the sequence flow of a CAD system (Gonzalez *et al.*, 2009; Sharma, 2017). For each step in CAD system, there are some techniques could be approached and each technique will have a different effect on CAD performance.

Studies in this field are aiming to register significant classification accuracy. However, the challenge remains to produce a CAD system that works in all cases regardless of the quality and the size of the database. Achieving maximum accuracy is the main motivation for CAD system in breast density characterization. The current work aims to improve the classification rate of not only in terms of accuracy but also in terms of recall and precision in order to improve the performance and robust of CAD system in breast density is our motivation.

CAD system is still an open challenging problem. Motivated by this need, we make our contributions in the development of automatic and accurate CAD system for characterizing of breast density as fatty, fibro-glandular and dense. We created a CAD system that focused mainly on improving classification based on feature extraction techniques which employ texture descriptors CLBP-HF and gradient descriptors HOG for the representation of the region of interest in a mammogram image. Texture Features (CLBP-HF) and gradient orientations (HOG) have shown promising results (Busaleh *et al.*, 2016; Guo *et al.*, 2010; Liu *et al.*, 2014; Naresh and Vani, 2015; Rabidas *et al.*, 2016).

However, none of these studies tested these two techniques on the most common dataset in breast density MIAS that labeling three classes corresponding to BI-RADS categories that approached by ACR. Plus, none of these studies try to fuse (HOG) and (CLBP-HF) even of the high results produced by these techniques in fusion with different techniques (Busaleh *et al.*, 2016). This idea encourages us to examine (HOG and CLBP-HF) separately and examine their fusion as state of the art work and examine their results on MAIS dataset then compare the results with Digital Database for Screening Mammography (DDSM) dataset.

In this study, we presented a proposed a CAD system implementing a hybrid feature extraction technique for breast density measurement. By investigated both features extraction approaches (HOG and CLBP-HF) individually and hybrid. The methodology was developed using well-known CAD system methodology that first starts with acquiring mammogram images from the Medical Image Analysis Society (MIAS) benchmark mammograms database and Digital Database for Screening Mammography (DDSM) dataset. Secondly, the system segments the mammogram region of interest and extracts features using HOG and CLBP-HF techniques and forwarding the data into the k-Nearest Neighbor classifier to output a result showing the breast is either fatty, grandeur or dense. The evaluation revealed that the proposed performs the best out of the (Bosch *et al.*, 2006; Hiba *et al.*, 2016; Mustra *et al.*, 2010; Silva and Menotti, 2012; Vaidehi and Subashini, 2015a) created a benchmark in the literature and yields an accuracy of 96.4, 96.59% recall and 96.75% precision outperforming the benchmark. This study presented a robust system for measure the breast density with high accuracy and significant performance.
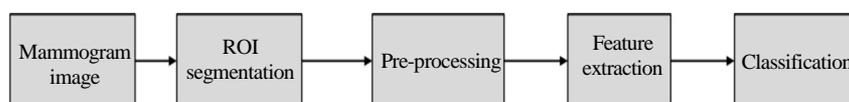

Fig. 1: Typical methodology of a CAD system

Related Work

In this section, first we present an overview of the recent work methods of Computer Aided Diagnosis (CAD) systems for breast density and then we briefly review the recent works which have been proposed for texture description HOG and CLBP-HF in mammogram images. We did not mean this review to be exhaustive. Our purpose is to provide a systematic review of only the indicative research concerning our subject.

Since (Wolfes', 1976) discovered that there is a high coloration between mammography breast tissue patterns and the risk of developing breast cancer, there has been an active interest in the fields of breast tissue density classification.

A newly proposed CAD system (Hiba *et al.*, 2016) aims to processed/enhanced then classified the digitized mammograms automatically into one of four categories in the density scale BI-RADS. The developed framework assisted radiologists by providing an automatic system for detecting and diagnosing possible cancers in mammograms. The authors applied Pectoral muscle removal for preprocessing step and histogram for ROI segmentation on the DDSM dataset. The proposed method used the Bag Of Features approach (BOF) for feature extractions for image representation. Then the resulting descriptor was compacted and used by SVM to carry out classification and learning. Performance comparison showed that their method achieved a 91.25% correct accuracy classification.

A hybrid approach (Vaidehi and Subashini, 2015a) was examined on statistical feature extraction techniques aiming to classify the ROI label into one of three breast tissue classes: Dense, fatty, glandular. For that, they had been tested on MIAS dataset. Also, they extracted Region Of Interest (ROI) of breast region by using statistical features. They classified using KNN classifier with different metrics (City-block, Euclidean and Cosine) and the majority of two of the distance measures output was returned as the final output. They reached about 91% accuracy results in the combined KNN classifier.

An automatic classification of breast density (Mustra *et al.*, 2010) is proposed using texture features that can be classified by obtaining feature selection process on MIAS dataset based on Haralick and Soh feature set with optimization for K-Nearest Neighbor (KNN) classifier. Furthermore, the feature selection was performed by three methods: The individual feature ranking, linear forward selection and finally using wrappers. The authors' idea was to obtain a possible correct classification rate and to figure out which texture features should be further developed. The author succeeded with 73.3% accuracy in using wrappers method and genetic search method.

A Content-Based mammogram Retrieval (CBIR) system was built (Vaidehi and Subashini, 2015b). CBIR systems tried to automatically classify the mammographic tissue density into (dense, glandular and fatty) and retrieved the similar tissue density images based on the input (query) image. They used about 11 features of haralick texture for extracting purpose. Finally, SVM with polynomial and Gaussian kernel are used for classifying the mammograms. The classified image retrieved using K-NN algorithm with Euclidean distance. The proposed work uses (MIAS) dataset and the classification accuracy obtained is 91.51%.

The mini-MIAS dataset has been investigated in (Sharma, 2017), where fatty glandular and dense glandular mammograms are treated as one group of dense mammograms and this resulted in a two-class classification problem (fatty, dense). The paper proposed a comparison of different ROI sizes for breast density classification. They implemented various models SGLCM, GLDS, FoS, SFM, Law's TEM, Fractal and FPS to extract texture features. They used Correlation-based Feature Selection (CFS) in order to select highly discriminating features. SVM has been applied and about 93.5% accuracy result considered as the highest result they get with 125×125 ROI size.

A new methodology was presented (Bosch *et al.*, 2006) to model and classify breast parenchymal tissue. Textons and SIFT measures have been proposed as features extraction. KNN and SVM were used to classify MIAS and DDMS dataset in many stages. In the first stage, the results obtained when classifying the MIAS dataset using its own annotation: Fatty, glandular and dense. The SVM outperformed KNN with 91.39%. Furthermore, they showed the results when BI-RADS annotation is used over both the MIAS and the DDSM databases in the second stage. Again, SVM outperformed KNN with 95.42% in MIAS and 84.75% in DDSM.

HOG based texture descriptor had not investigated in breast density field. However, A HOG texture approach (Busaleh *et al.*, 2016) was proposed in breast mass detection since it's one of the best descriptors for gradient orientation patterns. They also used SVM classifier on DDMS dataset to classify the data into two classes (mass or normal tissue). In order to determine the local distributions of DRLBP-HOG and DRLTP-HOG descriptors, they divided each image into different blocks of ROI and calculated the accuracy locally. The highest accuracy of applying HOG separately is 92.35%. They also investigated the combined of HOG with (DR-LBP) and also combined HOG with (DRLTP+HOG), the results achieved was 99.80 and 98.23% accuracy respectively. This high result suggested to a brilliant methodology for calculating accuracy locally.

Another study (Krishnaveni *et al.*, 2014) analyzed an efficient method by diagnosing the mammogram and used HOG in the feature extraction step. They used MIAS dataset and classified it using Naive Bayes

classification into (benign or malign) as two-class tumor problem. After ROI segmentation applied, they started with reprocessing step that consist of Contrast Enhancement, Smoothing by Gaussian filter, Adaptive thresholds and top hat filtering on grey scale image. Then they presented HOG feature extraction followed with several of texture features (mean, standard deviation, energy, contrast, correlation and homogeneity) which are calculated to measure the performance of the classification. The work was outperformed alternative pervious work with 96.25%.

Authors in (Hiba *et al.*, 2016) were the first to propose the technique for CLBP as a local feature extractor to generalize and complete LBP. They used the Outex and CUReT databases. They made their investigation by tested the operators: CLBP-Sign (CLBP_S), CLBP-Magnitude (CLBP_M) and the original image as its center gray level (C) with its the local difference (CLBP_C). The chi-square metrics is obtained with the nearest neighborhood classifier. CLBP has proven to obtain much better texture classification accuracy when combining the three mentioned operators than the state-of-the-arts LBP algorithms by 93.05% comparing to 84.65% that can be applied only the sign operator.

A recent paper proposed by (Naresh and Vani, 2015) worked on detecting breast cancer by using LBP and CLBP methods. The texture feature extraction is performed on the mammogram image taken from MIAS database. A morphological operation was used in preprocessing step. By using LBP, they extracted the texture features, but LBP uses only sign parameters and only using the parameter causes feature loss in an image. In comparison with CLBP that considers both sign and magnitude parameters, the result was robust for breast cancer detection, as CLBP is used for most recognition techniques. SVM classifier is obtained in order to identify the normal and normal cases. However, authors did not provide clear measurements of the results in the paper.

A comparative analysis (Rabidas *et al.*, 2016) was made on different texture features based on LBP, LBP Variance (LBPV) and Completed LBP (CLBP) features as a features extraction in mass classification. Experiments were performed on DDSM database considers as benign and malignant mass textures. In order to decrease the high dimensionality, stepwise logistic regression method has been employed. Then that followed by Fisher Linear Discriminant Analysis (FLDA) for benign-malignant mass classification. LBP, LBPV and CLBP have been achieved different results with four different block sizes where $w = 2,3,4$ and 5. However, LBP, LBPV and CLBP get around 92.25, 87.70 and 90.60% accuracy respectively.

A fusion of texture features was proposed (Gardezi and Faye, 2015) to improve classification accuracy by false positive reduction in mammograms. Their method

combined CLBP with grey level texture features where it obtained from the curvelet sub-bands. They experimented on MIAS and Image Retrieval in Medical Applications (IRMA) dataset. Removed the background and selected the Region Of Interest (ROI) is implemented in the preprocessing step to limit the search for abnormalities by before any classification process could be applied on the mammograms. The idea was to fuse both feature extraction methods and then using the nearest neighbor classifier to evaluate the features obtained from CLBP and curvelet and to classify benign and malignant classes. Even though, they tested separately but the fusion method had the highest result in 96.68% accuracy. The result suggested that fusion of features is significant and improved the performance of the system.

Different texture features and gradient orientation separately have been used for breast density and have shown significant results. In the present work, we intend to improve the accuracy of our CAD system by focusing on applying HOG and CLBP-HF separately and in fusion and determined their ability in extracted the most discriminate features of breast mammogram images.

Methodology

The main objective of this paper is to create a suitable CAD system implementing a hybrid approach of two well-known feature extraction algorithms, HOG and CLBP-HF, to help detect and accurately classify breast density into the BI-RADS categories.

An excellent feature set should have effective and discriminating features. Therefore, if more than one feature extraction technique is used in a mammographic image, then it will help to get higher accuracy. The feature extraction methods focus on converting images into feature vectors. Only effective features are extracted and presented to the classifier to classify the breast density into the predefined categories.

In designing our method, we followed the main steps that any CAD system follows as indicated (Vaidehi and Subashini, 2015a). The architecture of the proposed method for our proposed system is shown in Fig. 2.

Region of Interest Extraction

The ROI size is very effective in analyzing and calculating accuracy results. Thus the 'size and location' of the ROI is critical. It is a challenging task to find an optimal size for breast density classification. Cropping operation was applied to the dataset images to extract the region of interest which contained the breast area with its abnormalities and re-size the image to 256×256 in MIAS dataset and 512×512 in DDSM dataset. The images were cantered with only the breast region to prepare for further processing (Sharma, 2017; Silva and Menotti, 2012).
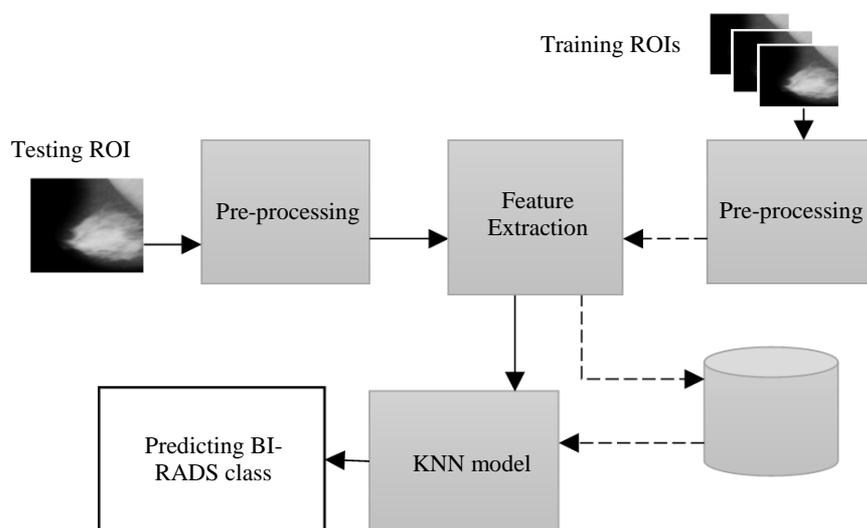
Fig. 2: An architecture of the our proposed system

Pre-Processing

In digital mammography, there is plenty of noise such as the pectorals muscle, tags and any other object unrelated to the breast. Also, previous works on breast density classification noticed that feature extraction process is affected by this kind of noise. As a result, pre-processing is a critical step to correct and adjust mammogram images for further analysis and classification (Silva and Menotti, 2012; Yadav *et al.*, 2014). This step covers the following process:

Noise Reduction

The median filter usually used to reduce noise in an image. Median filter is one of the simplest and most effective as it removes thin artifacts while preserving sharp edges (Sahakyan and Sarukhanyan, 2012). Furthermore, the Gaussian filter was tested in this experiment with a 3-by-3 neighborhood connection as approached in (Deng and Cahill, 1993):

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\left(x^2 + y^2\right)/2\sigma^2\right)$$

where, σ^2 is the variance of Gaussian filter and the size of the filter kernel 1 ($-1 =< x, y =< 1$) is often determined by omitting values lower than 5% of the maximum value of the kernel.

Image Enhancement

Since Mammograms have non-uniform background and very little contrast in the area above the core breast tissue region. Therefore, image enhancement is required (Silva and Menotti, 2012). Our proposed system applied image enhancement using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique as indicated in (Yadav *et al.*, 2014). The idea is that the CLAHE algorithm limited the local histogram of a block of pixels in order to limit the amount of contrast enhancement for each block. CLAHE algorithm can be processed using the following steps:

- Divide the original image into blocks
- Obtain a local histogram for each pixel in the block
- Clip the histogram based on the clip level
- Redistribute the histogram using binary search
- Obtain the enhanced pixel value by histogram integration

Image Segmentation

There are many techniques can be applied in order to segment the image. Global threshold technique and morphological operations applied on the bi-level images to remove imperfections introduced during segmentation (Sahakyan and Sarukhanyan, 2012). Figure 3 presented examples of each step in the applied preprocessing.

Candidate Block Splitting

This step is not one of the main steps in CAD system but it is inspired by (Busaleh *et al.*, 2016). To extract local features that lead to higher classification accuracy, we split the image into blocks as shown in Fig. 4. It consists of two regions; background and breast. For extracting local features, we will split the image into 4×4 and 16×16 blocks.

Feature Extraction

Extracting features plays a significant role in the result of the classifier. Due to that, we tested these two features extraction as fusion and separately.
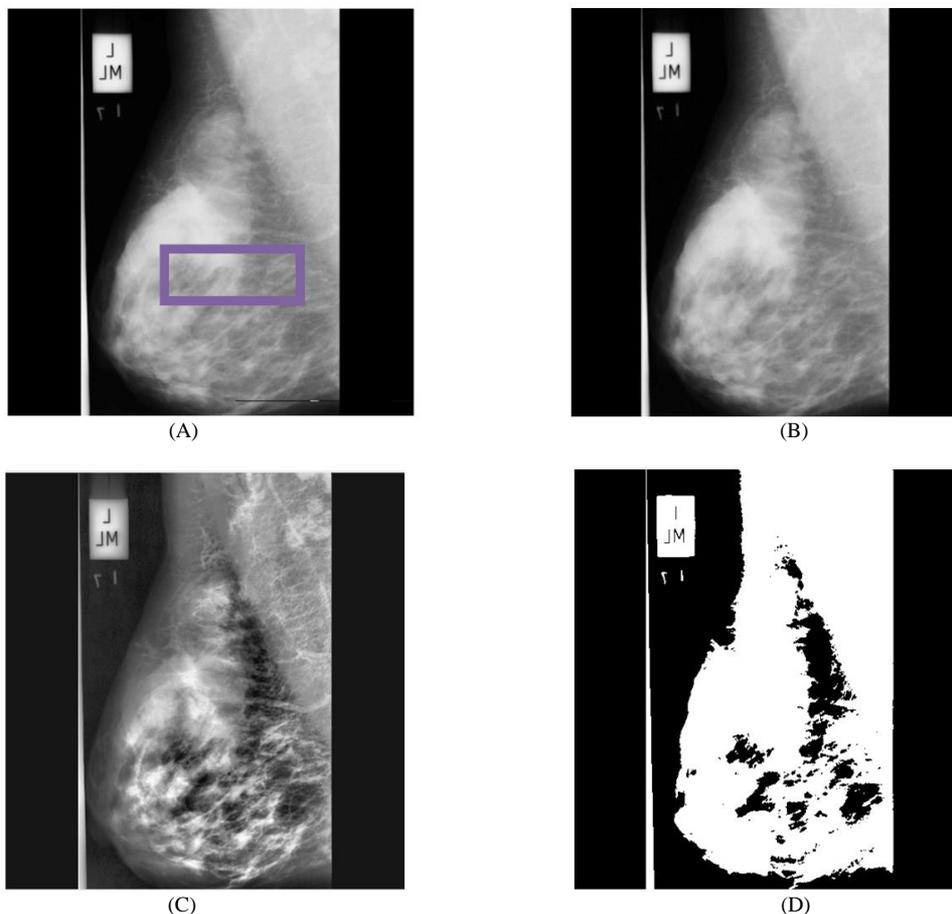
(A)

(B)

(C)

(D)

Fig. 3: Examples of preprocessing step (A) Original image; (B) Image after applying median filter; (C) Image after enhancement with CLAHE; (D) Image after threshold Technique
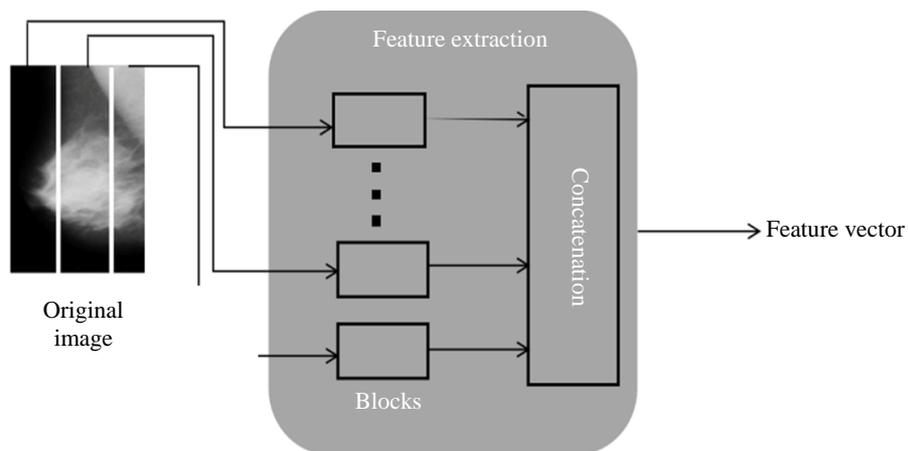

Fig. 4: Extracting local features

Histogram Of Gradient (HOG)

HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image. Implementation of HOG can be done as follows (Wang *et al.*, 2015a).

Gradient Computation

After image preprocessing, it is required to first preprocess the image and then move on to calculating the vertical and horizontal gradients of an image before calculating their histogram. Dalal and Triggs (2005) used

larger masks and found that they always decrease performance and smoothing damages it significantly.

Spatial/Orientation Binning

The image is split into small-connected predefined regions called cells. Each pixel in the cell calculates a weighted vote (gradient magnitude) for an edge orientation histogram channel. These channels are based on the values of orientation of the gradient element. The combined histogram entries are used as the feature vector describing the object.

Normalized Blocks

The gradient strength must be locally normalized in each cell, cells are then grouped together to form a larger cell or block. The blocks overlap working like a "windows" moving through all the cells, which means cells contribute to more than one block. The HOG descriptor result is the concatenated vector of the normalized components of each block.

Completed Local Binary Pattern Histogram Fourier Features (CLBP-HF)

Local Binary Pattern (LBP) is a method for texture feature extraction popular in image processing and recognition approaches (Zhao *et al.*, 2011). Guo *et al.* (2010) was the first to introduce the technique. Despite that, the original LBP descriptor (Ahonen and Pietikainen, 2007) has some disadvantages: Presenting long histograms which are not rotation invariant; capturing only the very local texture structure and being unable to exploit long-range information; limited discriminative capability based purely on local binaries differences and lacking noise robustness. In order to solve these problems, there are many LBP variations have been developed in order to cover LBP descriptor with a different aspect. One of them is discriminative power where CLBP-HF is advisable in order to improve discriminative power.

Complete Local Binary Pattern (CLBP), first proposed by (Zhao *et al.*, 2011) in 2010 is similar to the technique of LBP, CLBP applies three distinct methods. First CLBP considers the sign difference similar to LBP. In addition, the magnitude difference between the center pixel and its neighbors. Also, the center pixel is encoded into a binary code. Joint of these methods will provide us with CLBP histogram:

$$d_b = S_b * M_b \, and \begin{cases} S_b = sign(d_b) \\ M_b = |d_b| \end{cases}$$

where, S_b is the sign of d_b, S_b equals to 0 if $d_b >= 0$ and equals to 1 otherwise. While M representing the magnitude.

A static-texture, rotation invariant descriptor based on uniform LBP computed from Discrete Fourier Transforms (DFT) of Local Binary Pattern (LBP) histograms. Unlike other histograms, based invariant texture descriptors, which normalize rotation locally, the proposed invariants are constructed globally for the whole region to be described (Ahonen *et al.*, 2009). Also, this approach is generalized to embed sign and magnitude components of the LBP into this framework to improve the description ability (Wang *et al.*, 2015a; Zhao *et al.*, 2011).

We had tested extracting features in global and local form. In order to extract local features which could lead to getting high classification accuracy results, we were splatted the image into different (Busaleh *et al.*, 2016).

Classification

Classification is the technique for classifying the input patterns into analogous classes. One of the simplest classifiers is K-Nearest Neighbor (KNN) (Wang *et al.*, 2015b). It consists of the assignment of an unclassified vector using the closest k neighbors found in the training set (Vaidehi and Subashini, 2015a). For measuring the distance, there are many metrics, for instance, the Euclidean and City-block distance. Defining *K* is a challenging task since a small value leads to noise with high influence on the result and large value make it computationally expensive.

Results and Discussion

In this section, we present the results and discuss them. For the validation on the proposed method, we analyzed it using 250 ROI taken from Mammographic Image Analysis Society (MIAS) dataset which is a standard database with mammogram images used in scientific research (Vaidehi and Subashini 2015b). Every mammogram image is in size 1024×1024 pixels. We used 250 mammograms from the mini MIAS database. The classes are defined as Dense-glandular, Fatty-glandular and Fatty. Also, we compared the results with 240 mammograms DDSM where the classes are Fatty, Fibro-glandular, Heterogeneously dense and Extremely dense (Hiba *et al.*, 2016). The simulations have been carried out in the MATLAB environment.

Our independent variables are; the numbers of features extraction that has been used, the distance measure *k* in kNN classifier, the number of bin in HOG and the number of blocks for the division of an ROI. Two feature extraction technique has been used in our methodology, HOG applied with Gaussian filter mask of $\partial = 0.5$ and 32 number of bins and a simple 1-D [−1, 0, 1] masks assuming that this provided the best feature extraction result as indicated in (Wang *et al.*, 2015a). As for CLBP-HF, we applied it to window size 8 and radius 1 as mentioned in (Guo *et al.*, 2010). We tested the methodology globally and locally on the mammogram.

For local measurement, we divided ROIs into numbers of blocks N × N, N = 4 and 16 blocks to find the best division splitting as inspired from (Busaleh *et al.*, 2016). The classifier kNN was applied; defining value *k* is a challenging task. Therefore, we used three possible values of *k* which are 1,5 and 8 with the shortest distance. The measurement between classes will be done by both Euclidean distance metric and city block distance metric.

For performance evaluation, the 10-fold cross-validation technique was applied to partition the dataset into 10 equal subsets (folds) wherein each stage one-fold works as a validator and the remaining are training. This helps in guaranteeing that all patterns in the dataset will participate in both training and testing phases. It was considered the well-known performance measures accuracy, recall and precision are the dependent variables.

We implement the fusion of the two feature extraction methods globally and locally. However, the highest results were presented in the global form as shown in Table 1 and 2. The proposed system achieved significant improvement with different metrics. This implies that our assumption about hybrid texture descriptor and gradient orientation and its significant results were justified. The accuracy result was 96.4% according to the *k* = 1 in both Euclidean and city block metrics in KNN. It noticed that MIAS dataset has higher results than DDSM.

Table 3 shows the classification accuracy comparison between pervious systems used MIAS dataset in investigating the three-class problem. It can be seen clearly that our system gives the highest accuracy result. The compared previous work that used kNN as a classifier get 91% where the others used SVM and get high accuracy results reached to 95.42%. However, our proposed method gets 96.4% with KNN classifier. Table 4 show comparison between our proposed system with other system. All of those systems apply their design into DDSM dataset. The highest accuracy was 95.83% with kNN.

Table 1: Hybrid (HOG + CLBP-HF) accuracy results on MIAS dataset

Metric and K	Global			Local [4 4]			Local [16 16]		
	Acc.	Recall	Perci.	Acc.	Recall	Perci.	Acc.	Recall	Perci.
Euclidean with 1	**96.4**	**96.37**	**96.59**	56	56.59	55.61	50.8	51.31	50.47
Euclidean with 5	73.2	73.70	73.83	57.2	57.75	57.96	53.2	53.67	53.93
Euclidean with 8	68.4	68.71	69.41	57.6	58.06	58.14	56.4	57.00	56.76
Cityblock with1	96.4	96.42	96.57	59.2	59.72	59.42	54.0	53.84	53.50
Cityblock with5	76.8	77.18	77.46	62.8	63.21	63.70	59.2	60.02	58.58
Cityblock with8	72.4	72.83	72.66	60	59.04	59.03	60.4	60.94	60.08

Table 2: Hybrid (HOG + CLBP-HF) accuracy results on DDSM dataset

Metric and K	Global			Local [4 4]			Local [16 16]		
	Acc.	Recall	Perci.	Acc.	Recall	Perci.	Acc.	Recall	Perci.
Euclidean with 1	**95.83**	**95.83**	**95.88**	45.0	45.0	46.08	42.08	42.08	41.69
Euclidean with 5	68.33	68.33	69.26	41.66	41.67	41.49	39.17	39.17	37.87
Euclidean with 8	65.83	65.83	66.36	43.33	43.33	43.26	42.08	42.08	40.98
Cityblock with1	95.42	95.42	95.42	46.67	46.67	47.13	44.58	44.58	44.50
Cityblock with5	74.58	74.58	75.45	41.67	41.67	41.93	42.92	42.92	45.08
Cityblock with8	69.17	69.17	70.28	44.17	44.17	44.20	42.08	42.08	46.60

Table 3: Comparison of classification accuracy using MIAS dataset with pervious studied

Study	Feature extractors	Classifier	Accuracy (%)
Our proposed method	HOG + CLBP-HF	KNN	96.4
Vaidehi and Subashini (2015a)	Hybrid statistical features	KNN	91.0
Mustra *et al.* (2010)	Haralick and Soh features	KNN	73.3
Vaidehi and Subashini (2015b)	Haralick texture	SVM/KNN	91.51
Sharma (2017)	SGLC/GLDS/FoS/SFM/Law'sTEM/Fractal/FPS Models	SVM	93.5
Silva and Menotti (2012)	Textons/SIFT	SVM/KNN	95.42

Table 4: Comparison of classification accuracy using DDSM dataset with pervious studied

Study	Feature extractors	Classifier	Accuracy (%)
Our proposed method	HOG + CLBP-HF	KNN	95.83
Hiba *et al.* (2016)	LBP	SVM	91.25
Kumar *et al.* (2015)	2D WPT+ wavelet filters	SVM	73.7

Conclusion

Breast density characterization plays a significant role in detecting breast cancer at early stage. With the advance of computational intelligence and machine learning techniques, computer-aided detection attracts more attention breast density measurement. It has become one of the major research subjects in medical imaging.

In this study, a hybrid technique for classification metod of breast mammogram images is proposed. A promising result was established by preliminary evaluation on mammogram images, which demonstrates the robustness of the proposed technique. The proposed method is efficient for automated measurement of breast density according to the experimental results. Our proposed method produced a classification accuracy of 96.4, 96.59 recall and 96.75% precision on MIAS dataset. This experiment results show that the proposed method can successfully differentiate between the different tissue in the breast according to BI-RADS and this increased the detection accuracy of human breast density. We evaluated our method on DDSM and MIAS datasets. The results were higher in MIAS dataset. We also compared the obtained results with other methods. The proposed hybrid method is efficient and robust indicated by the comparative analysis.

Acknowledgment

We would like to thank College of Computer and Information Science, King Saud University, Kingdom of Saudi Arabia.

Author's Contributions

Anjoom Alabdulali and Alanod Bin Dris: Designed the research plan and organized the study, helped shape the research and comments on writing and apply the programming on MIAS dataset and the evaluation part.

Fatimah Alqahtani: Designed the research plan and organized the study, helped shape the research and comments on writing and apply the programming on DDSM dataset.

Aseel Bin Othman: Designed the research plan and organized the study, helped shape the research and comments on writing.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Ahonen, T., Matas, J., He, C., & Pietikäinen, M. (2009, June). Rotation invariant image description with local binary pattern histogram fourier features. In Scandinavian conference on image analysis (pp. 61-70). Springer, Berlin, Heidelberg.

Ahonen, T., & Pietikäinen, M. (2007, August). Soft histograms for local binary patterns. In Proceedings of the Finnish signal processing symposium, FINSIG (Vol. 5, No. 9, pp. 1-4).

Apt, K. (2003). Principles of constraint programming. Cambridge university press.

Basu, P. (2018). Cancer Treatment Centers of America. https://www.cancercenter.com/

Bosch, A., Munoz, X., Oliver, A., & Marti, J. (2006, June). Modeling and classifying breast tissue density in mammograms. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (pp. 1552-1558). IEEE.

Busaleh, M., Hussain, M., Aboalsamh, H. A., Zuair, M., & Bebis, G. (2016, December). False Positive Reduction in Breast Mass Detection Using the Fusion of Texture and Gradient Orientation Features. In International Symposium on Visual Computing (pp. 669-678). Springer, Cham.

CDC. (2020). Cancer Data and Statistics. https://www.cdc.gov/cancer/dcpc/data/index.htm

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

Deng, G., & Cahill, L. W. (1993, October). An adaptive Gaussian filter for noise reduction and edge detection. In 1993 IEEE conference record nuclear science symposium and medical imaging conference (pp. 1615-1619). IEEE.

Gardezi, S. J. S., & Faye, I. (2015). Fusion of completed local binary pattern features with curvelet features for mammogram classification. Applied Mathematics & Information Sciences, 9(6), 3037.

Gonzalez, R. C., Woods, R. E., & Masters, B. R. (2009). Digital Image Processing. Pearson Education India.

Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. IEEE transactions on image processing, 19(6), 1657-1663.

Hiba, C. H. O. U. G. R. A. D., Hamid, Z. O. U. A. K. I., & Omar, A. L. H. E. Y. A. N. E. (2016, October). An improved breast tissue density classification framework using bag of features model. In 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt) (pp. 405-409). IEEE.

Krishnaveni, S., Bhanumathi, R., & Pugazharasan, T. (2014). Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 3(3), 8274-8282.

Kumar, I., Bhadauria, H. S., & Virmani, J. (2015). Wavelet packet texture descriptors based four-class BIRADS breast tissue density classification. Procedia Computer Science, 70, 76-84.

Liu, L., Long, Y., Fieguth, P. W., Lao, S., & Zhao, G. (2014). BRINT: binary rotation invariant and noise tolerant texture classification. IEEE transactions on Image Processing, 23(7), 3071-3084.

Li, H., Giger, M. L., Huo, Z., Olopade, O. I., Lan, L., Weber, B. L., & Bonta, I. (2004). Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. Medical Physics, 31(3), 549-555.

Mustra, M., Grgic, M., & Delac, K. (2010, September). Feature selection for automatic breast density classification. In Proceedings ELMAR-2010 (pp. 9-16). IEEE.

Naresh, S., & Vani, S. (2015, August). Breast cancer detection using local binary patterns. In IJCA (Vol. 123, No. 16, pp. 6-9).

Rabidas, R., Midya, A., Chakraborty, J., & Arif, W. (2016, January). A study of different texture features based on local operator for benign-malignant mass classification. In 6th International Conference On Advances In Computing & Communications, IN Procedia Computer Science (Vol. 93, pp. 389-395).

Cram101 Textbook Reviews. (2011). Studyguide for Pattern Recognition by Sergios Theodoridis, ISBN: 9781597492720. Cram101 Incorporated.

Sahakyan, A., & Sarukhanyan, H. (2012). Segmentation of the breast region in digital mammograms and detection of masses. International journal of advanced computer science and applications, 3(2).

Sharma, V. (2017). Comparative analysis of region of interest of different sizes for breast density classification. International Journal of Medical Research & Health Sciences, 6(3), 76-84.

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016. CA: a cancer journal for clinicians, 66(1), 7-30.

Silva, W. R., & Menotti, D. (2012, July). Classification of mammograms by the breast composition. In Proceedings of the 2012 International Conference on Image Processing, Computer Vision and Pattern Recognition (pp. 1-6).

Vaidehi, K., & Subashini, T. S. (2015a). Breast tissue characterization using combined K-NN classifier. Indian Journal of Science and Technology, 8(1), 23.

Vaidehi, K., & Subashini, T. S. (2015b, January). Automatic classification and retrieval of mammographic tissue density using texture features. In 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO) (pp. 1-6). IEEE.

Wang, A., Dai, S., Yang, M., & Iwahori, Y. (2015a, August). A novel human detection algorithm combining HOG with LBP histogram Fourier. In 2015 10th International Conference on Communications and Networking in China (ChinaCom) (pp. 793-797). IEEE.

Wang, A., An, N., Chen, G., Li, L., & Alterovitz, G. (2015b). Accelerating wrapper-based feature selection with K-nearest-neighbor. Knowledge-Based Systems, 83, 81-91.

Wolfe, J. N. (1976). Breast patterns as an index of risk for developing breast cancer. American Journal of Roentgenology, 126(6), 1130-1137.

Yadav, G., Maheshwari, S., & Agarwal, A. (2014, September). Contrast limited adaptive histogram equalization based enhancement for real time video system. In 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2392-2397). IEEE.

Zhao, G., Ahonen, T., Matas, J., & Pietikainen, M. (2011). Rotation-invariant image and video description with local binary pattern features. IEEE transactions on image processing, 21(4), 1465-1477.