Characterization of Breast Masses Selecting Novel Shape and Texture Features with Multi-Objective Feature Selection Technique

Anmol Sharma[a], Harman Preet Kaur[b], Abhishek Midya[c], Jayasree Chakraborty [c,*]

[a]*Department of Information Technology, DAV Institute of Engineering and Technology, Jalandhar, India*
[b]*Department of Computer Science, DAV Institute of Engineering and Technology, Jalandhar, India*
[c]*Department of Surgery at Memorial Sloan Kettering Cancer Center New York, USA*

Abstract

Masses are one of the important yet challenging signs of breast cancer, visible in the mammogram. The paper presents a novel mass classification scheme via the introduction of new feature selection algorithm along with feature extraction technique. To capture complete and complex shape, we propose Translation, Rotation, and Shift (TRS) invariant Zernike moments as global shape descriptor. The extracted features are further clubbed with texture information. The discriminating features are then selected with a new wrapper-based feature selection scheme combined with multi-objective Non-dominated Sorting Genetic Algorithm (NSGA-II) where three objectives are optimized simultaneously. The experiments show that the proposed three objective functions allow the NSGA-II to reduce the feature dimensionality from 312 to four, while significantly outperforming classifiers trained on features with high dimensionality. With a set of four features, the method achieves the best area under the receiver characteristic curve of 0.95 and an accuracy of 89.89% using an artificial neural network for 270 randomly selected images from the DDSM database.

Keywords: Mass classification, mammograms, feature selection, multi-objective optimization

1. Introduction

Breast cancer is the most common cancer among women worldwide with nearly 1.7 million new cases estimated in 2012 [1]. It is also the most common cause of cancer related

*chakrabj@mskcc.org

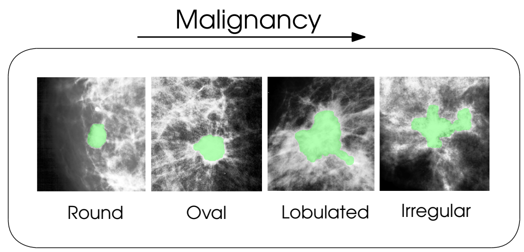
mortality among women in developing countries (such as Cambodia, Nepal, and Rwanda) and the second most common cause of cancer mortality, followed by lung cancer, among women in developed countries [1]. This can be largely attributed to non-availability of early detection facilities in the form of screening examinations [2]. Mammography has long been a tool of choice to perform screening examination for early detection of breast cancer [3, 4, 5]. Abnormalities seen in mammograms are of four types – masses, calcifications, architectural distortion, and bilateral assymetry. Often radiologists falsely flag an abnormality as cancer or fail to detect sign of cancer due to fatigue causing from inspection of large number of images daily, imperfect imaging, or subtle nature of abnormalities [5]. Since 65-90% of surgeries of suspected cancers turn out to be benign [5], the need is evident for an accurate Computer Aided Diagnosis (CADx) system to assist radiologists as second reader in distinguishing between benign and malignant abnormalities. It has been shown that detection sensitivity without CAD is 80%, while with CAD it is 90% [6]. Among all the mammographic-abnormalities, masses are one of the common signs of breast cancer [7]. However, the detection and diagnosis of masses is a challenging task due to their subtlety and variable appearance [8]. A large number of algorithms have been developed for various stages in automatic screening like segmentation, feature extraction, and classification [9], but still more research is needed in this area to further improve detection and classification accuracy. The most current research is targeted towards development of a CADx system for diagnosing breast masses in mammograms with a fine focus on the feature extraction stage [8].

In this research, we propose a mammographic mass classification system. The contribution in this work lies in feature extraction and selection phases. In the *feature extraction* stage we utilize translation, rotation and scale invariant (TRS invariant) Zernike Moments as shape descriptors, the use of which to our knowledge is not proved extensively in the literature. Although the investigation of Zernike Moments in earlier works [10, 11, 8, 12, 13] for the classification of masses into benign and malignant have been performed by using rotation invariant magnitudes of moments extracted from translation and scale normalized input images, these extracted moments do not form a complete and complex set of features [14]. In our work, we investigate the use of Zernike Moments as global shape descriptors which are made invariant to TRS transform using the methods described in [15], discussed in the subsequent sections. Apart from the shape-based features, texture and orientation features are also extracted in the form of Haralick's features calculated from Gray Level Co-occurrence Matrix (GLCM) [16] and Angle Co-occurrence Matrix (ACM) [17]. *Feature selection* has been shown to improve classifier performance while preserving robustness [18, 5]. The important features are then selected with a newly proposed feature selection scheme based upon a multi-objective evolutionary algorithm which minimizes a novel set of three objective functions—*feature dimensionality*, *classification* *error*, and a newly developed mutual information based objective function designed to ensure maximum relevancy and minimum redundancy of the selected feature subset using Non-dominated Sorting Genetic Algorithm II (NSGA-II) [19]. The introduction of three objectives for feature selection, solving them with NSGA-II, and evolving a pareto optimal solution find highly representative and non-redundant feature subsets that hold significant representation power for accurate classification of benign and malignant masses. Hence, our contribution in this study can be summarized as the investigation of TRS invariant Zernike Moments as global shape descriptors, analysis of the combination of TRS invariant Zernike Moments with texture and orientation features, and introduction of a new feature selection method with three objective functions for the classification of mammographic masses as benign and malignant.

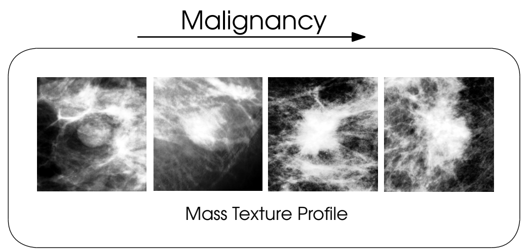
The rest of the paper is organized as follows; Section 2 provides a brief discussion about the existing features followed by the recent state-of-the-art feature selection techniques proposed in connection with mammographic mass classification, Section 3 describes the database used for this study, Section 4 describes our methodology in detail. Section 5 presents the results and corresponding discussion of the proposed approach, and lastly Section 6 concludes the paper with a summary of the work undertaken in this study.

2. Related Work

According to BIRADS [20], the most prominent clues of malignant mass lies in its shape, texture, and sometimes the directional patterns of its boundary. A benign mass is generally of oval or round in shape with well defined margin and low density, whereas a malignant mass may generally be of ill-defined shape with spiculated margin and high density as shown in Figure 1. Based on these characteristics different texture [21, 22]-, shape [23, 24, 25]-, and margin [26, 27]-based features have been proposed by the researchers to categorize masses as benign and malignant. Liu et al. extracted Haralick's features from Gray Level Co-occurrence Matrix (GLCM), obtained from mass region [28]. An area under the receiver operating characteristic (ROC) curve (A_z value) of 0.98 was obtained using Haralick's feature extracted from GLCM and Optical Density Co-occurrence Matrix (ODCM) [29]. A mass classification scheme using texton features was reported in [30]. A set of fifteen first and second order statistical textural features were calculated by Vani et al. [31], which delivered an accuracy (A_{cc}) of 91%. Muramatsu et al. analyzed radial local ternary patterns and obtained an A_z value of 0.90 with 376 regions of interest (ROIs) [32]. Recently, Rabidas et al. introduced Discriminative Robust Local Binary and Discriminative Robust Local Ternary Pattern for distinguishing benign and malignant masses [33]. Some researchers utilized oriented tissue patterns for the characterization of masses as benign and malignant. Using Gabor filter to extract features at different



(a)



(b)

Figure 1: (a) Typical shapes and (b) texture exhibited by masses seen in mammograms [20].

orientations and frequencies, Buciu and Gacsadi obtained an A_z value of 0.78 with 322 normal and abnormal cases [34]. Analyzing the oriented tissue patterns of three regions around masses using Angle Co-occurrence Matrix (ACM) [35], Chakraborty et al. classified masses [36, 37]. Zhang et al. [38] built an ensemble system of classifiers based upon shape features extracted from the mammographic masses. Tahmasbi et al [8] used Zernike Moments as shape descriptors for the classification of masses in mammograms. Many researchers clubbed shape- and texture-based features for mass classification. An A_z value of 0.96 for 160 ROIs using Zernike moments and local binary pattern has been reported [39]. Azizi et al. obtained A_{cc} of 89.90% for malignant cases and 87.40% for benign massses with 200 test images [40]. Sahiner et al. combined different texture-features and morphological features and reported an A_z value of 0.91 utilizing 249 mammograms [41].

To obtain better classification accuracy, some recent techniques have been developed in transform domain [42, 43, 23]. Using wavelet and curvelet transform, an A_{cc} of 97.30% with 161 patients was reported in [43]. Gorgel et al. [23] proposed a local seed region growing for the detection of regions of interest in mammogram, followed by spherical wavelet transform for feature extraction. Beura et al. [44] proposed to use GLCM and 2D-DWT as features along with a t-test based feature selection method. Dhahbi et al. [45] proposed a feature extraction method by converting the image to discrete curvelet transformed domain and calculating the first four order moments from the distribution of curvelet coefficients.

In most of the existing mass classification techniques, after the extraction of features, non-redundant and discriminative features were selected using several feature selection schemes, such as mutual information based feature ranking [46], stepwise regression [47], forward and backward Selection based methods [48, 49, 50], and ReliefF [51], Genetic Algorithms (GA) [52], Particle Swarm Optimization (PSO) [53, 54], Cuckoo Search (CS) [55], and support vector machine (SVM) based Recursive Feature Elimination (SVM-RFE) [56]. The performance of classifier in many folds is dependent on the proper selection of feature set. Liu et al. introduced a features selection scheme using SVM and recursive feature elimination after the extraction of geometric and texture features to classify the masses [56]. Mencattini et al. [57] extracted geometrical features like area, perimeter of boundary, radius, circularity etc. along with textural features to classify the masses after selecting the discriminating features via ranking them using ROC and A_z metrics. A feature selection scheme using particle swarm optimization was reported in [53]. A semi supervised relief based feature selection scheme for mass classification was introduced by Liu et al. [58]. Dong et al. [59] extracted shape, margin, texture, and intensity features from the segmented mass and its surroundings regions and use multiple classifiers like SVM, GA-SVM, PSO-SVM to compare their classification performance. As feature extraction as well as feature selection schemes, reported hitherto, delivers a mixed bag of performance, scope for further improvement exists.

3. Database

The Digital Database for Screening Mammography (DDSM) [60], collected at the University of South Florida, is a publicly available mammography database containing 43 volumes of total 2,620 cases with both mediolateral-oblique and cranialcaudal views of each breast. All the cases are supplied with associated patient information (age at time of study, ACR breast density rating, subtlety rating, and approximate boundary of the abnormality, etc.) and image information (scanner, spatial resolution, etc.) [60]. We randomly selected 270 images from the database, of which 144 images were diagnosed with benign and 126 with malignant masses. The selected images are with spatial resolutions of 50 μm/pixel, 42 μm/pixel, and 43.5 μm/pixel. All the selected images were converted from LJPEG format to PGM and normalized against different scanners using our own software which was released as open source software at `http://www.github.com/trane293/DDSMUtility`.

4. Proposed Method

Based on the observation that shape, texture, and margin characteristics of masses carry important information to distinguish between benign and malignant masses, in this study, we investigate these characteristics by extracting three different feature sets—TRS invariant Zernike Moments as global shape descriptors, Haralick's features from GLCM and ACMs to capture texture and orientation patterns, from different regions associated with mass. After feature extraction, the most informative and non-redundant features are selected using the proposed NSGA-II-based feature selection scheme minimizing three objective functions. The selected features are then used to classify benign and malignant masses. A schematic diagram of the proposed mass classification method is shown in Figure 2.

4.1. Pre-processing

Since, the performance of features, specially shape and margin-based descriptors, highly depends on proper segmentation of mass region, the proposed method uses Chan-Vese [61] algorithm, due to its effective performance, to obtain fine boundary considering database provided boundary as the initial contour. For the application of Chan-Vese algorithm, an ROI is selected by cropping the input mammogram with an offset of δ mm on all sides of the initial boundary marked by the radiologist. Let Y be the set of all pixels in the approximate mass boundary given by the radiologist, X be the set of all pixels on the image boundary. Also let $ib \in X$ be a pixel on image boundary such that $ib \perp ob$, where ob is a pixel $\in Y$, then δ is given as:

$$\delta = \begin{cases} x \text{ mm}, & \text{if } |ib - ob| > x \\ & \forall\ ob \in Y and \\ & \forall\ ib \in X \\ |ib - ob|, & \text{otherwise.} \end{cases}$$

In this study, x is empirically set to 12.5 mm. Once the ROI is obtained, histogram equalization [5] is performed to address the issue of low contrast, which hinders accurate analysis and segmentation of masses, followed by suppression of high-frequency noise using simple median filter [62]. The final ROI obtained after contrast enhancement and median filtering is shown in Figure 3. The image, marked as A_1171_1.LEFT_CC in the DDSM database, has been used to illustrate our approach throughout the paper.

4.2. Mass boundary detection using Chan-Vese Active Contour without Edges

The Chan-Vese algorithm relies on the internal homogeneity, instead of edge information. At the very basic, the model evolves a contour by minimizing an energy function

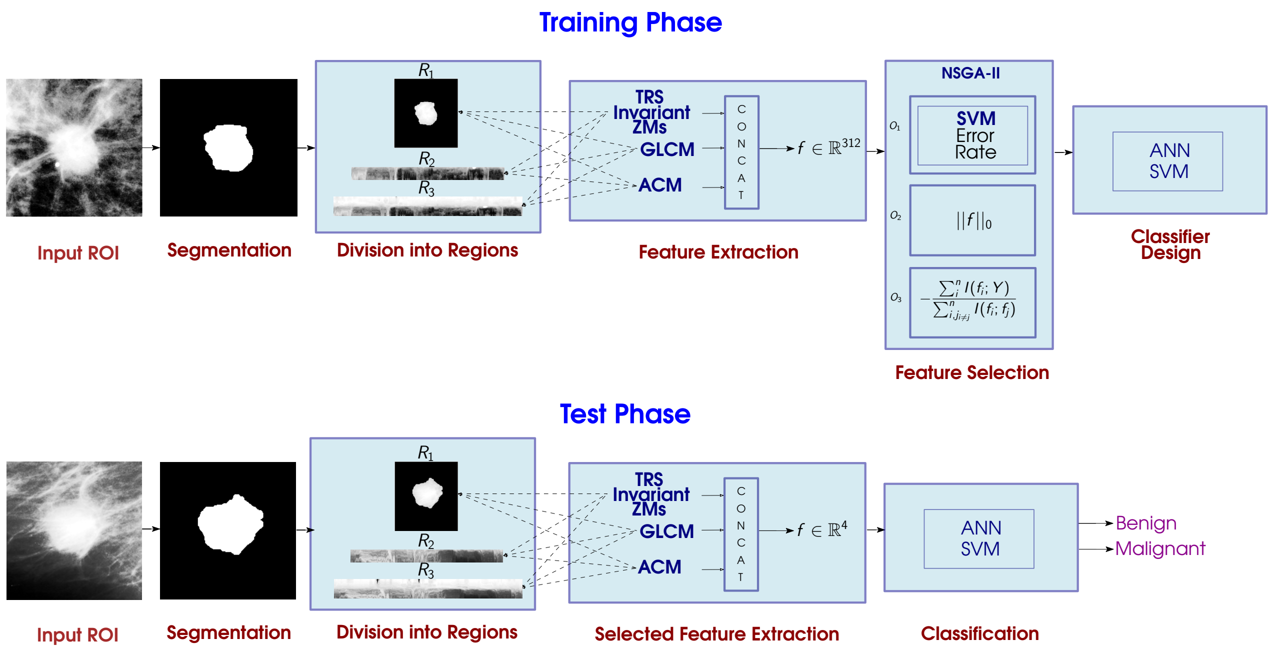


Figure 2: Flowchart illustrating our approach. Top row shows training phase. During test phase, only the selected features are extracted from the three regions, and feature selection step is not present.

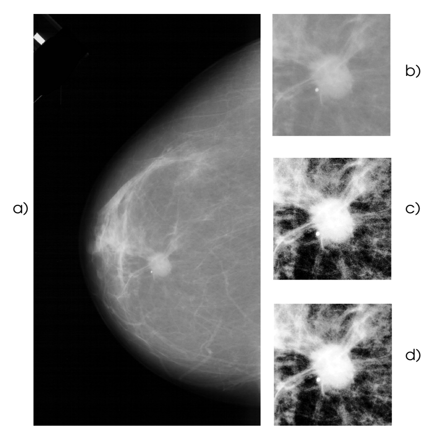


Figure 3: Illustration of the Preprocessing method. *a)* Input mammogram, *b)* Cropped ROI, *c)* ROI after histogram equalization, *d)* ROI after median filtering

$F(\phi)$ via a level set method [63], where $\phi(i,j,t)$ represents the current state of the contour in the sense that the pixel (i,j) of the image plane belongs to the contour if $\phi(i,j,t) = 0$, where t is the "iteration".

The "fitting energy" function that forms the core of the Chan-Vese active contour model can be written as:

$$\begin{gathered} F(\phi) = \mu\bigg(\int_{\Omega} |\Delta H(\phi)| dx\bigg)^p \; + \; v \int_{\Omega} H(\phi) dx \; + \\ \lambda_1 \int_{\Omega} |I - c_1|^2 H(\phi) dx \; + \; \lambda_2 \int_{\Omega} |I - c_2|^2 (1 - H(\phi)) dx. \end{gathered} \tag{1}$$

where μ, v, λ_1, λ_2, and p are user defined parameters; H is the Heaviside function; I is the image to be segmented; and Ω is the domain of the image. c_1 and c_2 are averages of the image I in the regions where $\phi \geq 0$ and $\phi < 0$, respectively, given by

$$c_1 = \frac{\int_{\Omega} I.H(\phi) dxdy}{\int_{\Omega} H(\phi) dxdy}, \quad c_2 = \frac{\int_{\Omega} I.(1 - H(\phi)) dxdy}{\int_{\Omega} (1 - H(\phi)) dxdy}. \tag{2}$$

The first term of the "fitting energy" function $\mu\big(\int_{\Omega} |\Delta H(\phi)| dx\big)^p$ is incorporated to penalize the total length of the edge contour for a given segmentation. If a smooth boundary is expected, this term may be weighed more heavily to avoid finding a complex (and in turn long) perimeter. Similarly, the second term $v \int_{\Omega} H(\phi) dx$ is a penalty on the total area of the foreground region found by the segmentation. The third term, $\lambda_1 \int_{\Omega} |I - c_1|^2 H(\phi) dx$, is directly proportional to the variance of image gray level in the foreground region and provides a measures of how "uniform" the particular region is in terms of pixel intensities. The fourth term does the same for background region. Minimization of the sum of these terms leads to segmentation of an image into maximum possible uniform foreground and background region. The set of parameters were empirically selected as $\mu = 0.5$, $v = 0$, $\lambda_1 = \lambda_2 = 1$, $p = 1$, and $dt = 0.1$. The extracted boundary and segmentated mass region obtained with the Chan-Vese algorithm for the image shown in Figure 3d is provided in Figure 4b and 4c.

For smoothing of the mass boundary, obtained using Chan-Vese algorithm, the segmented image is subjected to erosion and dilation using a disc shaped structuring element, given by:

$$I \circ s = (I \ominus s) \oplus s, \tag{3}$$

where I is the image and s is the disc shaped structuring element with radius $r = 6$; $\ominus$

denotes erosion and $\oplus$ denotes dilation operation, respectively.

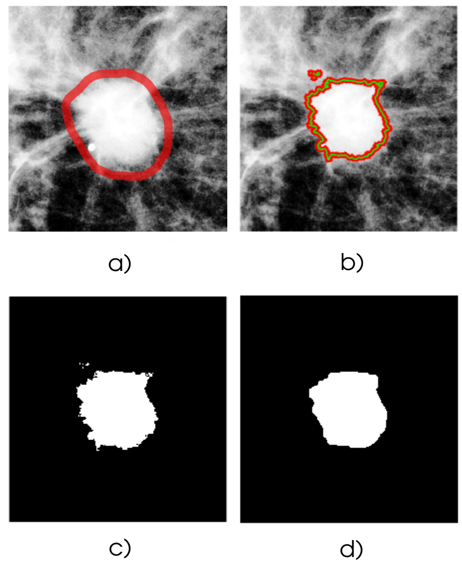


Figure 4: *a)* Initial contour, *b)* Evolved contour with Chan-Vese active contour without edges method, *c)* Raw Chan-Vese mask, *d)* Morphologically operated mask

4.3. Selection of ROIs for feature extraction

Tumor texture plays an important role in mass characterization. Moreover, it has been consistently observed that the boundary of the masses carry important information of malignancy [20, 64, 36]. The presence of spiculations over mass boundaries becomes a strong evidence of malignancy [65, 3]. To effectively capture these characteristics, we define three regions: R_1, R_2, and R_3, associated with mass [36], as illustrated in Figure 5, where

- R_1— The entire mass region,
- R_2— A band of pixels outside the mass region.
- R_3— A band of pixels enclosing the boundary, equi-distant towards the inner and outer direction from the mass boundary. The width of R_3 is double to that of R_2.

The extraction of texture- and gradient-based features directly from regions, surrounding the mass boundary, is error prone since the tissue patterns inside the mass region is oriented radially. Similarly, capturing of spiculations arising from the mass boundary is complicated because of the fact that the direction of spiculations changes with the shape of mass and the curvature of its margin [21]. To overcome these problems, we employ Rubber Band Straightening Transform (RBST) on R_2 and R_3, proposed by Sahiner et al. in [21], which maps the band of pixels onto the Cartesian plane in such a way that the boundary of mass appears approximately as a horizontal line and the spiculations as vertical. Due to these properties, RBST images are proved to be a better alternative for calculation of features from the surrounding regions of the mass, as compared to direct extraction of features from the radial image [21].

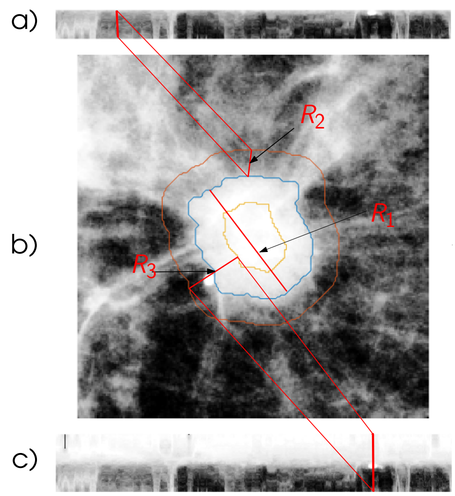


Figure 5: Selection of regions R_1, R_2 and R_3. a) RBST transform of region R_2, b) ROI with illustration of regions R_1, R_2 and R_3, c) RBST transfrom of region R_3.

However, instead of selecting same band width for all masses [21], here we automatically calculated the width of R_2 as a function of the area of segmented mass as proposed in [36], given by:

$$W = r_c \left(\frac{A_r}{A_c} \right), \tag{4}$$

where r_c = radius of the bounding circle, A_r = area of the object, and A_c = area of the bounding circle. Hence, the width of R_3 is $2W$.

4.4. Feature Extraction

4.4.1. Extraction of Zernike Moments

Zernike Moments (ZM) [66], effective shape descriptor of an object [67, 68], are the mapping of an image onto a set of complex orthogonal Zernike polynomials which represent the image with minimum redundancy of information [67]. For a digital image with intensity function $I(x, y)$ at pixel position (x, y), the ZM of order n and repetition l ($|l| \leq n$)

is given by [67]:

$$A_{n\ell} = \frac{n+1}{\lambda_N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) V^*_{n\ell}(x,y), \tag{5}$$

where λ_N is the number of pixels located in the unit circle, used as normalization factor; $V_{n\ell}$ represent Zernike ploynomial, orthogonal on a unit disk $x^2 + y^2 \leq 1$, and is defined as –

$$V_{n\ell} = R_{n\ell}(r) e^{i\ell\theta}, \quad |r| \leq 1, \tag{6}$$

with

$$R_{n\ell} = \sum_{s=0}^{(n-|\ell|)/2} (-1)^s \frac{(n-s)!\ r^{n-2s}}{s!((n+|\ell|)/2-s)!((n-|\ell|)/2-s)!}. \tag{7}$$

The difference $n - |\ell|$ is always even and the asterisk(*) in equation denotes complex conjugate. Substituting equation (7) in (6) and then the resulting equation in (5), we get

$$A_{n\ell} = \frac{n+1}{\lambda_N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) R_{n\ell}(r_{xy}) e^{i\ell\theta}. \tag{8}$$

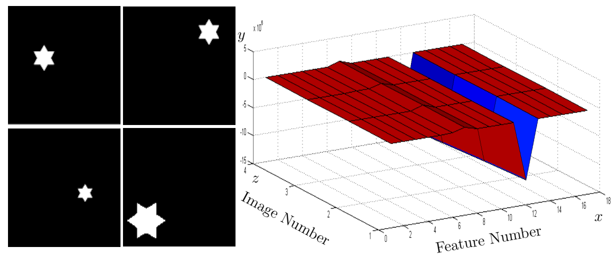
However, these features are sensitive to translation and scaling of ROI [66, 14]. In earlier works, this problem was tackled by normalizing the images by translating the mass into the center of the ROI and scaling the ROI to a fixed radius, while to preserve the rotation invariance, only the magnitude of the moments was used, discarding the complex part of moments [11, 10, 8, 12, 13].

In this study, we investigate the use of a complete and complex set of ZMs invariant to translation, scaling, and rotation (TRS) which eliminates the need of normalizing the ROI, before feature extraction. The detailed description of steps involved in computing the proposed TRS invariant moments are described in the subsequent sections.

In this study, invariance to TRS transforms is achieved using the following methods:

Invariance to Translation. Instead of normalizing images against translation by shifting the mass at the image center as in earlier works, the present work calculates translation invariant ZMs considering origin at the centroid of the mass lesion (x_c, y_c) to reduce the computational complexity [15].

Invariance to Scaling. The scaling of mass to a predefined radius leads to the loss of shape information as the co-scaling procedure incorporates re-sampling and re-quantization of the image. To address this issue, we normalize the ZMs with respect to central moment as



a) b)

Figure 6: Invariance of ZMs upto order $n = 5$ towards TRS transformations. a) 4 test images with applied transformations, b) graph showing negligible change in value of moments (y-axis) with respect to changing images (z-axis).

proposed by Ye et al. [14] which can be expressed as

$$A^s_{n\ell} = \frac{A_{n\ell}}{\mu_{00}}, \quad \text{where } \mu_{00} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} I(x,y)dxdy$$

Invariance to Rotation. In many applications, rotation invariant ZMs have been extracted by using magnitudes of the moments as they are invariant to rotation [69, 14]. But the extracted set of moments, having limited recognition power, do not form a complex set of features. [15]. In this study, the rotation normalization method proposed by Flusser et al. [15] has been used to extract complete rotation invariant ZMs. The ZMs, extracted from the ROI, are normalized using a proper nonzero normalizing moment, found by searching the ZMs with repetition $\ell = 1$, ie. A_{31}, $A_{51}...A_{n_{max}1}$. If they are found to be under the chosen threshold (in this study $thres = 1e - 3$), we suppose the object to be rotationally symmetric and then search the moments with successively increased repetitions 2, 3, etc. for the first non-zero moment. For nonsymmetric objects, we choose the normalizing moment as A_{31} as proposed in [15]. Now, if the normalizing moment $A_{m_r\ell_r}$ has a phase –

$$\phi = \frac{1}{\ell_r} arctan\left[\frac{Im(A_{m_r\ell_r})}{Re(A_{m_r\ell_r})}\right] \tag{9}$$

then:

$$\overline{Z}_{n\ell} = \frac{A_{n\ell}}{(A_{m_r\ell_r})^{\ell/\ell_r}} \tag{10}$$

is the required rotation invariant [15].

The invariance of our extracted ZMs is illustrated in Figure 6 using four synthetic images, containing different versions of the same underlying image (star), after application of

Table 1: Features and their respective indices in the concatenated feature vector

Feature Name	$GLCM$	$GLCM$	$GLCM$	$ACM1$	$ACM1$	$ACM1$	$ACM2$	$ACM2$	$ACM2$	$INVTS$	$INVTS$	$INVTS$
Region	R_1	R_2	R_3	R_1	R_2	R_3	R_1	R_2	R_3	R_1	R_2	R_3
# Features	14	14	14	14	14	14	14	14	14	62	62	62
Index in F	1–14	15–28	29–42	43–56	57–70	71–84	85–98	99–112	113–126	127–188	189–250	251–312

transformations in the form of translation, rotation, scaling, and all together. As illustrated in the Figure 6b, the values of moments (y axis) show negligible change with respect to different images (z axis). We explore TRS invariant ZMs upto order $n = 10$ i.e. a total of 62 ZM features are extracted from each region.

4.4.2. Gray-Level Co-occurrence Matrix (GLCM)

To capture the intensity-based texture characteristics of mass and its surrounding regions, we compute Gray-Level Co-occurrence Matrix (GLCM), proposed by Haralick et al. [16], from all the three regions. The image gray levels are quantized to 256 levels. Four GLCM matrices are computed with angles 0^o, 45^o, 90^o, 135^o, and pixel distance $d = 1$, then averaged to form a final matrix, from which 14 Haralick's features are computed.

4.4.3. Angle Co-occurrence Matrices

Angle Co-occurrence Matrices (ACMs) were first introduced by Chakraborty et al. [17, 35] to quantify oriented edge patterns of mammogram for the detection of architectural distortion via computing joint occurrences of orientation angles of tissue structures. The matrices are defined as:

$$ACM1_{(\ell,\theta)}(i,j) = \frac{P_a(i,j)}{\sum_{i=1}^{N_\theta}\sum_{j=1}^{N_\theta} P_a(i,j)}, \tag{11}$$

$$ACM2_{(\ell,\theta)}(i,j) = \frac{P_m(i,j)}{\sum_{i=1}^{N_\theta}\sum_{j=1}^{N_\theta} P_m(i,j)}, \tag{12}$$

where $P_a(i,j)$ counts the number of occurrences of a pixel pair with the orientation angles i and j, separated by (ℓ,θ); $P_m(i,j)$ is the sum of the gradient magnitude responses of all the pixel pairs having angles i and j, separated by a distance ℓ at an angle θ. N_θ is the number of quantized angle levels.

In this study, ACMs are computed from gradient magnitude and angle information, obtained by applying a 3×3 Sobel filter on the image. The average of four ACMs, computed with $l = 1$ and $\theta = 0^o$, 45^o, 90^o, and 135^o after quantizing the orientation angle into $N_\theta = 64$ bins, are considered as the final rotation invariant ACM matrices to extract 14 haralick's features.

A total set of 312 features, extracted from all the regions are concatenated to form one single vector $F \in \mathbb{R}^{312}$ as shown in Table 1.

4.5. Feature Selection

The feature vector, obtained by concatenating all features from R_1, R_2, and R_3 regions is of high dimension ($F \in \mathbb{R}^{312}$) which increases the computational complexity of the classifier. Moreover, some features may carry redundant information. Hence, selection of optimum feature set is important. In this study, we propose feature selection scheme to optimize three objectives simultaneously, designed to select discriminating features, using a Non-dominated Sorting Genetic Algorithm (NSGA-II).

The first objective function used is the classification error, which is an accepted performance measure with low values denoting better performance. We wrap an SVM classifier with ten-fold cross-validation in our proposed feature selection technique to measure the classification error, which can be written as

$$O_1 = \frac{FP + FN}{TP + FP + TN + FN}, \tag{13}$$

where TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively.

Using a subset $f \subset F$ for classification can keep the performance robust. Moreover, small feature dimensionality reduces computational complexity in both train and test phases. Hence, for keeping the feature dimensionality minimum while retaining a good classification performance, we propose to use the cardinality of the evolved feature subset as our second objective function as given below,

$$O_2 = ||f||_0. \tag{14}$$

In addition to these criteria, a new objective function is designed to select a non-redundant and highly representative subset of features via maximizing the mutual information between the features and the class variable (maximum relevancy) and minimizing the mutual information between the features themselves (minimum redundancy) and is defined as:

$$O_3 = -\frac{\sum_i^m MI(f_i; Y)}{\sum_{i,j_{i \neq j}}^m MI(f_i; f_j)}, \tag{15}$$

where $Y = [C_1, C_2 .... C_k]$ is the set of classes (in our study $k = 2$), f_i is the i^{th} feature variable; m is the number of features, and mutual information between two variables X and Y is

$$MI(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right). \tag{16}$$

These three objective functions ($\mathbb{O} = \left\{ O_1(\overrightarrow{f}), O_2(\overrightarrow{f}), O_3(\overrightarrow{f}) \right\}$) are then simultaneously minimized. However, in multi objective optimization (MOO) problem, a single solution, satisfying all objectives, may not be always possible. In this case, the objective functions are said to be conflicting and are generally solved by determining a number of Pareto optimal solutions. The MOO setting can be formally stated as follows:

Let $\mathbb{O}$ be a set of m objective functions that are required to be simultaneously optimized,

$$\mathbb{O} = \left\{ O_1(\overrightarrow{f}), O_2(\overrightarrow{f}), ...O_m(\overrightarrow{f}) \right\}, \tag{17}$$

where $\overrightarrow{f*}$ is the vector of decision variables given as

$$\overrightarrow{f*} = [f_1^*, f_2^*, ..., f_r^*] \tag{18}$$

and r is the dimensionality of the variables (number of features in this case). Constraints in an MOO problem define a feasible region $\mathcal{F}$ containing all admissible solutions. In general, the scalar concept of optimality does not apply for MOO. An objective vector $\overrightarrow{\mathbb{O}_1}$ is said to *dominate* another objective vector $\overrightarrow{\mathbb{O}_2}$ (i.e., $\overrightarrow{\mathbb{O}_1} < \overrightarrow{\mathbb{O}_2}$) if no component of $\overrightarrow{\mathbb{O}_1}$ is greater than the corresponding components of $\overrightarrow{\mathbb{O}_2}$. Henceforth, a solution $\overrightarrow{f^*}$ is called as Pareto optimal solution of the given set of objective functions, $\mathbb{O}$, if and only if there is no other $\overrightarrow{f^\#}$ that dominates $\overrightarrow{f^*}$.

The NSGA-II [19] is a widely used MOO algorithm due to its good spread of solutions with convergence near the true Pareto-optimal front, less-niching, simple constraint handling strategy [70], and lower computational complexity of $O(MN^2)$, where M is the number of objectives and N is the population size. The algorithm uses an evolutionary process, where at each iteration a population of candidate solutions, known as chromosomes, is evolved towards better solution via generating child population using selection, crossover, and mutation operations. In this study, chromosomes are represented by bit stream where each bit represents the selection (1) or rejection (0) of a feature. Therefore, the length of chromosome is 312, same as the number of total features (N_f). Once mutated, the next population is constructed by combining the parent and child population and carrying out the non-dominated sorting based on the objective functions and crowding distance. Similarity between members of each sub-group is evaluated on the Pareto front, and the resulting groups and similarity measures are used to promote a diverse front of non-dominated solutions [71]. The method is stopped when a stopping criteria (function improvement tolerance $\delta = 1e^{-4}$) or maximum number of generations is reached and the highest ranked Pareto solutions are considered as the final set of solutions. The proposed feature selection method is described in Algorithm 1. A population (pop) of size 200, denoted by N_{pop}, is randomly initialized from a uniform distribution. The number of generations (T), crossover probability (P_c), and mutation probability (P_m) are empirically set to 200, 0.8, and 0.01, respectively. The selected features are then used to perform classification between benign and malignant masses and the best performing Pareto solution is chosen.

Algorithm 1 NSGA-II based Feature Selection

Input: N_{pop}, P_c, P_m, δ, T, $Train$ (Training data), C_L (Class Labels)
Output: optimum set of features, $\overrightarrow{f*} = [f_1^*, f_2^*, \ldots.., f_r^*]$

generation t←0;
$pop_0 \leftarrow$ Initialize parent population of size N_{pop}
Evaluate objective functions ($O_1,\ O_2,\ O_3$) for each chromosome $\in pop_0$
Sort pop_0 based on non-domination sorting
while tol$>\ \delta$ & $t <= T$ **do**
    Create child population ($Child$) using i) tournament selection, ii) crossover, and iii) mutation
    Evaluate objective functions for each chromosome $\in Child$
    Merge parent and child population ($Union = Pop \cup Child$)
    Construct all non-dominated front sets $Fronts$ using $Union$
    $pop_{t+1} \leftarrow \phi$
    $Front_L \leftarrow \phi$
    for all ($Front_i \in Fronts$) **do**
        if (Size(pop_{t+1}) + Size($Front_i$) $< N_{pop}$) **then**
            $pop_{t+1} \leftarrow pop_{t+1} \cup Front_i$
        else
            calculate crowding distance in $Front_i$
            sort $Front_i$ based on crowding distance
            $pop_{t+1} \leftarrow pop_{t+1} \cup N_{pop} - |pop_{t+1}|$ elements of $Front_i$
        end if
    end for
    $t \leftarrow t + 1$
end while
return $Children$ representing $\overrightarrow{f*}$

4.6. Classification

In this study we use two classifiers namely Artificial Neural Network (ANN) and Support Vector Machine (SVM) for the classification of masses as benign or malignant. A feed forward neural network with one hidden layer of five neurons is used. The number of hidden layers and neurons in the network are selected using random search. The weights and bias values of the network are updated according to the Levenberg-Marquardt optimization rule. For SVM, we use an RBF kernel. To avoid bias, the average results obtained with ten repetitive runs of the ten-fold cross-validation is considered for the performance evaluation.

5. Results and Discussion

The proposed method of benign-malignant mass classification is implemented in MATLAB® 2015a on a PC with Intel Core i5 4200U processor of 2.30GHz, 4GB RAM, and Windows 10 operating system. The 10-fold cross-validation is repeated 10 times and the mean (μ) and standard deviation (σ) of some well established metrics – Classification Accuracy, False Positive Rate (FPR), False Negative Rate (FNR), and Area under the ROC Curve (A_z) are used for performance evaluation. Since, the method emphasizes on both feature extraction and selection, we present the results in two parts. First, without using any feature selection to evaluate the performance of different features in different regions (E1 Experiments), followed by utilization of our proposed feature selection technique (E2 Experiments).

5.1. E1 Experiments

As discussed above, in E1 experiments, different sets of features, extracted from different regions, are analyzed separately as well as in combination to evaluated their performance in classification using ANN and SVM classifiers directly without applying any feature selection algorithm. The results are listed in Table 2 along with the total number of features used in each experiment. The best result is obtained when all the features from all regions are combined to train the classifiers. The combined features from all regions delivers an accuracy of $83.26 \pm 1.87\%$ with FPR $17.28 \pm 2.43\%$, FNR 15.80 ± 3.91, and A_z value of 0.94 ± 0.01 using ANN. Also, it is observed that the both texture- and orientation-based features work well, whereas shape based features (Zernike Moments) perform considerably poor in all observations. From Table 2, it can be inferred that textural and orientation features extracted from Region R_1 perform better than any of the other combinations.

Table 2: E1 Experiments using ANN and SVM as classifiers. Best results are highlighted in boldface.

Region	Descriptor	Features	ANN				SVM			
			Accuracy (%)	FPR (%)	FNR (%)	A_z	Accuracy (%)	FPR (%)	FNR (%)	A_z
R_1	$GLCM$	14	82.41±1.33	23.34±3.16	10.52±1.03	0.92±0.01	84.44±1.71	16.95±2.04	14.50±3.16	0.92±0.01
R_1	$ACM1$	14	81.74±1.62	18.47±4.28	17.77±2.72	0.91±0.01	81.79±1.55	23.54±3.09	13.53±2.71	0.91±0.01
R_1	$ACM2$	14	82.07±1.17	17.56±1.70	17.84±1.73	0.90±0.02	79.96±1.74	29.47±4.34	11.75±2.15	0.89±0.01
R_1	$Zernike$	62	65.22±1.24	24.56±1.58	45.83±2.35	0.69±0.01	66.28±3.51	32.50±3.87	34.86±4.91	0.70±0.00
R_1	$Combined$	104	82.15±1.69	20.57±3.93	14.66±5.71	0.91±0.01	70.35±2.96	60.98±4.09	3.13±1.33	0.90±0.01
R_2	$GLCM$	14	81.81±1.41	22.52±2.67	12.61±3.45	0.91±0.01	75.07±2.18	23.82±4.59	25.53±2.95	0.91±0.00
R_2	$ACM1$	14	81.70±1.29	18.01±2.82	19.27±1.90	0.91±0.01	81.10±2.00	22.11±3.15	16.16±2.46	0.91±0.01
R_2	$ACM2$	14	78.59±1.12	19.89±2.89	22.64±2.49	0.88±0.01	67.83±1.95	42.84±2.65	22.61±3.70	0.75±0.01
R_2	$Zernike$	62	56.59±1.80	38.07±2.47	48.19±2.88	0.61±0.02	58.84±2.62	37.07±2.68	44.71±3.99	0.60±0.02
R_2	$Combined$	104	79.74±2.07	22.15±3.25	17.28±4.37	0.90±0.01	69.71±3.38	59.69±6.03	5.14±1.90	0.90±0.01
R_3	$GLCM$	14	82.00±1.50	22.14±2.68	12.90±2.08	0.89±0.02	80.18±1.42	22.53±2.16	17.60±1.66	0.89±0.01
R_3	$ACM1$	14	80.33±1.32	17.87±2.66	22.26±2.32	0.89±0.01	80.69±2.18	27.23±3.27	12.78±3.17	0.89±0.00
R_3	$ACM2$	14	79.00±1.33	19.52±2.64	21.82±2.57	0.88±0.01	76.29±2.64	30.91±4.60	17.56±3.31	0.86±0.00
R_3	$Zernike$	62	67.11±1.19	53.12±1.42	8.96±1.27	0.70±0.01	64.66±2.32	19.35±3.38	49.60±3.02	0.67±0.01
R_3	$Combined$	104	80.67±2.01	20.45±3.49	17.59±2.78	0.91±0.01	70.02±2.35	59.54±3.75	2.52±1.94	0.85±0.01
$Combined$	$GLCM$	42	82.70±1.98	21.54±2.74	12.42±2.31	0.91±0.01	82.78±1.66	29.44±2.47	5.16±0.99	0.91±0.00
$Combined$	$ACM1$	42	82.44±1.21	16.61±3.16	18.95±2.67	0.92±0.01	72.31±1.82	55.12±3.59	4.11±2.04	0.91±0.01
$Combined$	$ACM2$	42	82.22±2.07	16.54±3.19	18.99±4.17	0.91±0.02	69.43±1.98	60.98±2.13	2.45±1.18	0.78±0.00
$Combined$	$Zernike$	186	65.26±2.20	41.90±5.13	25.87±7.28	0.76±0.01	60.39±5.42	54.52±7.48	26.11±4.80	0.74±0.01
$Combined$	$Combined$	312	**83.26±1.87**	**17.28±2.43**	**15.80±3.91**	**0.94±0.01**	**82.98±1.99**	**18.87±0.10**	**14.16±0.51**	**0.93±0.01**

5.2. E2 Experiments

The E2 experiments are conducted to observe the classification performance after selecting the most discriminative and non-redundant subset of features using our proposed feature selection scheme. Since E1 experiments show that the best performance is achieved by using a combined feature set extracted from all regions, we evolve the subset from the pool of all extracted features from all regions ($F \in \mathbb{R}^{312}$). The objective functions are analyzed in different combinations to evaluate their effect on feature selection and thus in classification. The classification results are provided in Table 3 along with total number of features selected. For conciseness, the combinations of the objective functions used in the table are abbreviated, and are defined as follows:

$$\begin{aligned} OF_1 &= \{O_1\} \\ OF_{1,2} &= \{O_1, O_2\} \\ OF_{1,3} &= \{O_1, O_3\} \\ OF_{1,2,3} &= \{O_1, O_2, O_3\} \end{aligned}$$

It must be noted that for OF_1, we use simple binary Genetic Algorithm. It is clearly evident that Parteo solutions evolved using $OF_{1,3}$ and $OF_{1,2,3}$ yield the best performance. Also, these solutions have a significantly small feature dimensionality (11, 11, 6, 4) which accounts to dimensionality reduction by over 95%. Table 4 reports the features selected

Table 3: E2 Experiments with two best evolved Pareto Solutions from full feature set f, using ANN and SVM as classifier

Objective Functions	# Features	ANN				SVM			
		Accuracy (%)	**FPR (%)**	**FNR (%)**	$\boldsymbol{A_z}$	**Accuracy (%)**	**FPR (%)**	**FNR (%)**	$\boldsymbol{A_z}$
OF_1	86	88.81±1.17	13.90±2.45	7.83±2.97	0.95±0.01	86.14±2.09	6.42±1.79	22.86±3.91	0.94±0.01
$OF_{1,2}$	6	87.33±1.43	18.43±1.79	5.90±1.31	0.92±0.00	88.53±1.85	15.48±2.53	6.99±2.59	0.92±0.01
$OF_{1,2}$	5	87.19±0.86	18.19±1.96	6.38±2.16	0.93±0.01	89.73±1.42	13.90±2.18	6.00±1.45	0.94±0.02
$OF_{1,3}$	11	88.56±0.91	16.78±0.97	5.02±1.37	0.94±0.00	87.01±2.53	13.22±2.60	12.85±3.26	0.93±0.00
$OF_{1,3}$	11	88.81±0.82	15.31±1.80	5.89±2.39	0.94±0.01	87.60±2.21	12.15±3.06	12.67±1.88	0.93±0.01
$OF_{1,2,3}$	**6**	**89.44±1.58**	**13.46±2.67**	**9.08±2.70**	**0.96±0.01**	**87.17±1.52**	**12.28±1.94**	**11.44±2.16**	**0.95±0.00**
$OF_{1,2,3}$	**4**	**89.89±0.81**	**11.35±0.99**	**6.56±1.38**	**0.95±0.00**	**89.83±1.95**	**14.59±3.35**	**9.64±2.06**	**0.95±0.01**

with different objective functions. It can be observed that most of the features selected when objective function combination $OF_{1,2,3}$ is used are GLCM and ACM features, whereas when the combination $OF_{1,3}$ is used the invariants dominate the Pareto solutions. This shows that Zernike Moments of higher order (nearly 10) have ample representation power for the classification of masses as benign or malignant, and can be coupled with texture- and orientation-based (GLCM and ACM) features for good performance. However, according to the results in Table 3, the combination $OF_{1,2,3}$ performs the best with a fair margin as compared to using $OF_{1,3}$ and achieved an accuracy 89.89 ± 0.81, FPR 11.35 ± 0.99, FNR 6.56 ± 1.38, and A_z value 0.95 ± 0.01 for ANN with reducing feature dimensionality by 98.71%. Similar performance was achieved by SVM using the same Pareto solution, with accuracy 89.83 ± 1.95, FPR 14.59 ± 3.35, FNR 9.64 ± 2.06 and A_z 0.95 ± 0.01.

6. Conclusion

In this paper we investigate the use of TRS invariant Zernike Moments as global shape descriptors in combination with texture and directional edge information for classification of mammographic masses into benign and malignant, and also propose an NSGA-II based feature selection method with a novel set of three objective functions, optimized simultaneously. The experiments show that although ZMs perform poorly as individual features, they perform much better in tandem with other shape and texture features, where some high order invariants proved to be highly representative and effective descriptors of the shape. The proposed feature selection algorithm is also shown to be effective in reducing feature dimensionality by over 95%, while still managing to increase classification performance of the classifier when compared to using all extracted features.

Table 4: Features selected in the best performing Pareto solutions obtained using $OF_{1,2}$, $OF_{1,3}$, and $OF_{1,2,3}$

Objective Function	**#Feat- -ures**	**Selected Features** index(region)(feature_name)
$OF_{1,2}$	6	4(R_1)(GLCM), 6(R_1)(GLCM), 53(R_1)(ACM1), 93(R_1)(ACM2), 147(R_1)(INVTS), 275(R_3)(INVTS)
$OF_{1,2}$	5	4(R_1)(GLCM), 64(R_1)(GLCM), 53(R_1)(ACM1), 93(R_1)(ACM2), 147(R_1)(INVTS)
$OF_{1,3}$	11	4(R_1)(GLCM), 6(R_1)(GLCM), 79(R_3)(ACM1), 93(R_1)(ACM2), 169(R_1)(INVTS), 225(R_3)(INVTS), 246(R_3)(INVTS), 261(R_3)(INVTS), 292(R_3)(INVTS), 300(R_3)(INVTS), 312(R_3)(INVTS)
$OF_{1,3}$	11	4(R_1)(GLCM), 6(R_1)(GLCM), 29(R_3)(GLCM), 79(R_3)(ACM1), 93(R_1)(ACM2), 169(R_1)(INVTS), 225(R_2)(INVTS), 246(R_2)(INVTS), 261(R_3)(INVTS), 292(R_3)(INVTS), 312(R_3)(INVTS)
$OF_{1,2,3}$	6	4(R_1)(GLCM), 7(R_1)(GLCM), 34(R_3)(GLCM), 74(R_3)(ACM1), 83(R_3)(ACM1), 93(R_1)(ACM2)
$OF_{1,2,3}$	4	4(R_1)(GLCM), 29(R_3)(GLCM), 74(R_3)(ACM1), 93(R_1)(ACM2)

Acknowledgment

The authors would like to thank Prof. P.S Mann, DAVIET Jalandhar for their constructive comments and suggestions.

7. References

[1] J. Ferlay, I. Soerjomataram, M. Ervik, et al., GLOBOCON 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012, International Agency for Research on Cancer.

[2] L. N. Shulman, W. Willett, A. Sievers, F. M. Knaul, Breast cancer in developing countries: opportunities for improved survival, Journal of Oncology 2010.

[3] American Cancer Society, Breast cancer facts & figures 2013-2014, American Cancer Society, Atlanta.

[4] J. Tang, R. Rangayyan, J. Xu, I. El Naqa, Y. Yang, Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances, IEEE Transactions on Information Technology in Biomedicine 13 (2) (2009) 236–251. doi:10.1109/TITB.2008.2009441.

[5] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, H. Du, Approaches for automated detection and classification of masses in mammograms, Pattern Recognition 39 (4) (2006) 646–668.

[6] R. M. Rangayyan, F. J. Ayres, J. Leo Desautels, A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs, Journal of the Franklin Institute 344 (3) (2007) 312–348.

[7] American Cancer Society, Breast cancer facts & figures 2015-2016, American Cancer Society, Atlanta.

[8] A. Tahmasbi, F. Saki, S. B. Shokouhi, Classification of benign and malignant masses based on Zernike moments, Computers in Biology and Medicine 41 (8) (2011) 726–735.

[9] A. Oliver, J. Freixenet, J. Marti, E. Pérez, J. Pont, E. R. Denton, R. Zwiggelaar, A review of automatic mass detection and segmentation in mammographic images, Medical Image Analysis 14 (2) (2010) 87–110.

[10] A. Tahmasbi, F. Saki, S. B. Shokouhi, Mass diagnosis in mammography images using novel FTRD features, in: IEEE 17th Iranian Conference of Biomedical Engineering (ICBME), 2010, pp. 1–5.

[11] A. Oliver, X. Llado, J. Marti, J. Freixenet, Applying Zernike moments for automatic mass diagnosis, International Journal of Computer Assisted Radiology and Surgery 5 (Suppl 1) (2010) 200–206.

[12] F. Saki, A. Tahmasbi, H. Soltanian-Zadeh, S. B. Shokouhi, Fast opposite weight learning rules with application in breast cancer diagnosis, Computers in Biology and Medicine 43 (1) (2013) 32 – 41.

[13] S. Sharma, P. Khanna, Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM, Journal of Digital Imaging 28 (1) (2015) 77–90.

[14] Y. Bin, P. Jia-Xiong, Invariance analysis of improved Zernike moments, Journal of Optics A: Pure and Applied Optics 4 (6) (2002) 606.

[15] J. Flusser, B. Zitova, T. Suk, Moments and moment invariants in pattern recognition, John Wiley & Sons, 2009.

[16] R. M. Haralick, K. Shanmugam, I. H. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics (6) (1973) 610–621.

[17] J. Chakraborty, R. M. Rangayyan, S. Banik, S. Mukhopadhyay, J. L. Desautels, Detection of architectural distortion in prior mammograms using statistical measures of orientation of texture, in: Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis, Vol. 8315, 2012, pp. 831521–831521–8.

[18] M. L. Giger, Z. Huo, M. A. Kupinski, C. J. Vyborny, Computer-aided diagnosis in mammography, Handbook of Medical Imaging 2 (2000) 915–1004.

[19] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197.

[20] F. Narvez, G. Daz, E. Romero, Automatic BI-RADS description of mammographic masses, in: Digital Mammography, Vol. 6136 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2010, pp. 673–681. doi:10.1007/978-3-642-13666-5_91.

[21] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis, Medical Physics 25 (4) (1998) 516–526.

[22] L. Nanni, A. Lumini, S. Brahnam, Survey on LBP based texture descriptors for image classification, Expert Systems with Applications 39 (2012) 3634 – 3641.

[23] P. Görgel, A. Sertbas, O. N. Ucan, Mammographical mass detection and classification using local seed region growing–spherical wavelet transform (LSRG–SWT) hybrid scheme, Computers in Biology and Medicine 43 (6) (2013) 765–774.

[24] A. Vadivel, B. Surendiran, A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories, Computers in Biology and Medicine 43 (4) (2013) 259 – 267.

[25] A. Serifovic-Trbalic, A. Trbalic, D. Demirovic, N. Prljaca, P. C. Cattin, Classification of benign and malignant masses in breast mammograms, in: 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)-2014, 2014, pp. 228–233.

[26] K. Bojar, M. Nieniewski, New features for classification of cancerous masses in mammograms based on morphological dilation, in: 5th International Conference on Visual Information Engineering (VIE)-2008, 2008, pp. 111–116.

[27] C.-H. Wei, S. Y. Chen, X. Liu, Mammogram retrieval on similar mass lesions, Computer Methods and Programs in Biomedicine 106 (3) (2012) 234 – 248.

[28] X. Liu, J. Liu, D. Zhou, J. Tang, A benign and malignant mass classification algorithm based on an improved level set segmentation and texture feature analysis, in: 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)-2010, 2010, pp. 1–4.

[29] S. C. Tai, Z. S. Chen, W. T. Tsai, An automatic mass detection system in mammograms based on complex texture features, IEEE Journal of Biomedical and Health Informatics 18 (2) (2014) 618–627.

[30] X. Z. Li, S. Williams, G. Lee, M. Deng, Computer-aided mammography classification of malignant mass regions and normal regions based on novel texton features, in: 12th International Conference on Control Automation Robotics Vision (ICARCV)-2012, 2012, pp. 1431–1436.

[31] G. Vani, R. Savitha, N. Sundararajan, Classification of abnormalities in digitized mammograms using extreme learning machine, in: 11th International Conference on Control Automation Robotics Vision (ICARCV)-2010, 2010, pp. 2114–2117.

[32] C. Muramatsu, T. Hara, T. Endo, H. Fujita, Breast mass classification on mammograms using radial local ternary patterns, Computers in Biology and Medicine 72 (2016) 43–53.

[33] R. Rabidas, A. Midya, A. Sadhu, J. Chakraborty, Benign-malignant mass classification in mammogram using edge weighted local texture features, in: SPIE Medical Imaging, Vol. 9785, 2016, pp. 97851X–97851X–6.

[34] I. Buciu, A. Gacsadi, Directional features for automatic tumor classification of mammogram images, Biomedical Signal Processing and Control 6 (4) (2011) 370–378.

[35] J. Chakraborty, R. M. Rangayyan, S. Banik, S. Mukhopadhyay, J. L. Desautels, Statistical measures of orientation of texture for the detection of architectural distortion in prior mammograms of interval-cancer, Journal of Electronic Imaging 21 (3).

[36] J. Chakraborty, A. Midya, S. Mukhopadhyay, A. Sadhu, Automatic characterization of masses in mammograms, in: IEEE 6th International Conference on Biomedical Engineering and Informatics (BMEI), 2013, pp. 111–115. doi:10.1109/BMEI.2013.6746917.

[37] A. Midya, J. Chakraborty, Classification of benign and malignant masses in mammograms using multi-resolution analysis of oriented patterns, in: IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015, pp. 411–414. doi:10.1109/ISBI.2015.7163899.

[38] Y. Zhang, N. Tomuro, J. Furst, D. S. Raicu, Building an ensemble system for diagnosing masses in mammograms, International Journal of Computer Assisted Radiology and Surgery 7 (2) (2012) 323–329.

[39] M. G. Laroussi, N. Ben Ayed, A. Masmoudi, D. Masmoudi, Diagnosis of masses in mammographic images based on Zernike moments and Local binary attributes, in: World Congress on Computer and Information Technology (WCCIT), 2013, pp. 1–6.

[40] N. Azizi, N. Zemmal, Y. Guiassa, N. Farah, Kernel based classifiers fusion with features diversity for breast masses classification, in: 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), 2013, pp. 116–121.

[41] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, L. M. Hadjiiski, Improvement of mammographic mass characterization using spiculation measures and morphological features, Medical Physics 28 (7) (2001) 1455–1465.

[42] M. M. Eltoukhy, I. Faye, B. B. Samir, Breast cancer diagnosis in digital mammogram using multiscale curvelet transform, Computerized Medical Imaging and Graphics 34 (4) (2010) 269–276.

[43] M. M. Eltoukhy, I. Faye, B. B. Samir, A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation, Computers in biology and medicine 42 (1) (2012) 123–128.

[44] S. Beura, B. Majhi, R. Dash, Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer, Neurocomputing 154 (2015) 1–14.

[45] S. Dhahbi, W. Barhoumi, E. Zagrouba, Breast cancer diagnosis in digitized mammograms using curvelet moments, Computers in Biology and Medicine 64 (2015) 79–90.

[46] I. Diamant, M. Shalhon, J. Goldberger, H. Greenspan, Mutual information criterion for feature selection with application to classification of breast microcalcifications, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2016, pp. 97841S–97841S.

[47] A. Kamra, V. Jain, S. Singh, S. Mittal, Characterization of architectural distortion in mammograms based on texture analysis using support vector machine classifier with clinical evaluation, Journal of Digital Imaging 29 (1) (2016) 104–114.

[48] S.-T. Luo, B.-W. Cheng, Diagnosing breast masses in digital mammography using feature selection and ensemble methods, Journal of Medical Systems 36 (2) (2012) 569–577.

[49] M. Tan, J. Pu, B. Zheng, Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model, International Journal of Computer Assisted Radiology and Surgery 9 (6) (2014) 1005–1020.

[50] V. Nguyen, D. Nguyen, T. Nguyen, V. Phan, Q. Truong, Filter-based feature selection and support vector machine for false positive reduction in computer-aided mass detection in mammograms, in: Seventh International Conference on Machine Vision (ICMV 2014), International Society for Optics and Photonics, 2015, pp. 94451H–94451H.

[51] A. Heshmati, R. Amjadifard, J. Shanbehzadeh, ReliefF-based feature selection for automatic tumor classification of mammogram images, in: 2011 7th Iranian Conference on Machine Vision and Image Processing, IEEE, 2011, pp. 1–5.

[52] A. K. Mohanty, M. R. Senapati, S. K. Lenka, A novel image mining technique for classification of mammograms using hybrid feature selection, Neural Computing and Applications 22 (6) (2013) 1151–1161.

[53] M. T. Wong, X. He, H. Nguyen, W. C. Yeh, Particle swarm optimization based feature selection in mammogram mass classification, in: International Conference on Computerized Healthcare (ICCH)-2012, 2012, pp. 152–157.

[54] G. Jothi, H. H. Inbarani, A. T. Azar, Hybrid tolerance rough set: PSO based supervised feature selection for digital mammogram images, International Journal of Fuzzy System Applications (IJFSA) 3 (4) (2013) 15–30.

[55] M. Sudha, S. Selvarajan, Feature selection based on enhanced cuckoo search for breast cancer classification in mammogram image, Circuits and Systems 7 (04) (2016) 327.

[56] X. Liu, J. Tang, Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method, IEEE Systems Journal 8 (3) (2014) 910–920.

[57] A. Mencattini, M. Salmeri, G. Rabottino, S. Salicone, Metrological characterization of a CADx system for the classification of breast masses in mammograms, IEEE Transactions on Instrumentation and Measurement 59 (11) (2010) 2792–2799.

[58] X. Liu, J. Liu, Z. Feng, X. Xu, J. Tang, Mass classification in mammogram with semi-supervised relief based feature selection, in: Fifth International Conference on Graphic and Image Processing, Vol. 9069, 2013.

[59] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, K. Wang, An efficient approach for automated mass segmentation and classification in mammograms, Journal of Digital Imaging 28 (5) (2015) 613–625.

[60] M. Heath, K. Bowyer, D. Kopans, R. Moore, The digital database for screening mammography, in: Proceedings of the 5th International Workshop on Digital Mammography, Citeseer, 2000, pp. 212–218.

[61] T. F. Chan, L. A. Vese, Active contours without edges, IEEE Transactions on Image Processing 10 (2) (2001) 266–277.

[62] A. Sharma, J. Singh, Image denoising using spatial domain filters: A quantitative study, in: IEEE 6th International Congress on Image and Signal Processing (CISP), Vol. 01, 2013, pp. 293–298. doi:10.1109/CISP.2013.6744005.

[63] S. Osher, J. A. Sethian, Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations, Journal of Computational Physics 79 (1) (1988) 12–49.

[64] R.-Y. Tang, W. Gao, L.-h. Ma, B.-y. Lin, G.-w. Xu, BI-RADS categorization and positive predictive value of mammographic features, Chinese Journal of Cancer Research 13 (3) (2001) 202–205.

[65] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, R. Sigal, BIRADS$^{\mathrm{TM}}$ classification in mammography, European Journal of Radiology 61 (2) (2007) 192–194.

[66] J. Wood, Invariant pattern recognition: a review, Pattern Recognition 29 (1) (1996) 1–17.

[67] M. R. Teague, Image analysis via the general theory of moments∗, Journal of the Optical Society of America 70 (8) (1980) 920–930. doi:10.1364/JOSA.70.000920.

[68] A. Wallin, O. Kubler, Complete sets of complex Zernike moment invariants and the role of the pseudoinvariants, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (11) (1995) 1106–1110. doi:10.1109/34.473239.

[69] A. Khotanzad, Y. H. Hong, Invariant image recognition by Zernike moments, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (5) (1990) 489–497. doi:10.1109/34.55109.

[70] J. Chakraborty, A. Konar, A. Nagar, H. Tawfik, A multi-objective pareto-optimal solution to the box-pushing problem by mobile robots, in: Second UKSIM European Symposium on Computer Modeling and Simulation (EMS'08), IEEE, 2008, pp. 70–75.

[71] J. Brownlee, Clever algorithms: nature-inspired programming recipes, Jason Brownlee, 2011.