2016

Breast Cancer Classification of Mammographic Masses Using Circularity Max Metric, A New Method

Tae Keun Heo
South Dakota State University

BREAST CANCER CLASSIFICATION OF MAMMOGRAPHIC MASSES USING CIRCULARITY MAX METRIC, A NEW METHOD

BY

TAE KEUN HEO

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Computer Science

South Dakota State University

2016

Breast Cancer classification of Mammographic Masses with

Circularity Max

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Computer Science degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Dr. Sung Y. Shin, Ph.D.
Thesis Advisor Date

Dr. Steven Hietpas, Ph.D.
Head, Department of Electrical Engineering
and Computer Science Date

Dean, Graduate School Date

CONTENTS

ABBREVIATIONS

DDSM	Digital Database for Screening Mammography
FN	False Negative
FP	False Positive
MCC	Matthews Correlation Coefficient
MRI	Magnetic Resonance Imaging
RDM	Radial Distance Measure
ROI	Region of Interest
SVM	Support Vector Machine
TP	True Positive
TN	True Negative

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

BREAST CANCER CLASSIFICATION OF MAMMOGRAPHIC MASSES USING CIRCULARITY MAX METRIC, A NEW METHOD

TAE KEUN HEO

2016

Breast cancer classification can be divided into two categories. The first category is a benign tumor, and the other is a malignant tumor. The main purpose of breast cancer classification is to classify abnormalities into benign or malignant classes and thus help physicians with further analysis by minimizing potential errors that can be made by fatigued or inexperienced physicians. This paper proposes a new shape metric based on the area ratio of a circle to classify mammographic images into benign and malignant class. Support Vector Machine is used as a machine learning tool for training and classification purposes. The improved performance of the proposed shape metric was used to evaluate and to compare the performances between existing method, which is called Circularity Range Ratio and proposed method, which is called Circularity Max. The result shows that the proposed Circularity Max method improves the Matthews Correlation Coefficient, specificity, sensitivity and accuracy. Therefore, the shape metric can be a promising tool to provide preliminary decision support information to physicians for further diagnosis.

1. INTRODUCTION

Cancer is a disease related to abnormal cell growth that has potential to invade and spread to other parts of the body. However, all anomalies are not considered as dangerous. Benign tumors do not have possibility to spread to other organs, and they are not considered as cancerous cases. Malignant tumors, on the other hand, spread through other organs inside human. There are many different types of cancer cases depending where they develop. Breast cancer is the most common cancer among women, and about 1 in 8 U.S. women (about 12%) is estimated to develop invasive breast cancer over the course of her lifetime according to the American Cancer Society. In 2014, an estimated 232,670 new cases of invasive breast cancer were expected to be diagnosed in women in the U.S., along with 62,570 new cases of non-invasive (in situ) breast cancer [1]. It is the second-most common and leading cause of cancer deaths among women in US [2]. At current, since the cause of breast cancer remains unknown there is no effective ways to prevent it. Hence, it is very important to detect breast cancer at an early stage which gives women a better chance of full recovery causing a high survival rate [3].

Several imaging techniques exist for breast cancer examination such as magnetic resonance imaging (MRI), Ultrasound imaging, X-ray imaging and mammography. Mammography uses a low dose X-ray system to examine the breast, and is one of the most effective method for detection of breast cancer [4]. To use mammography to detect tumors, the cancer needs to be classified to different classes. One of the key problems in the classification process is the choice of features for differentiating classes. An effective shape descriptor is one of the key components in classification, as shape is one of the

basic properties present in the image [5]. Irregular shape is one of the most frequently appearing features for the malignant masses which can be used to identify breast tumor as benign or malignant. Most benign masses are characterized by well-defined edges and are regular, oval and smooth in shape with possible macrolobulations whereas malignant tumors have ill-defined, fuzzy and rough contours with microlobulations, spiculations, and concavities [6]. Analysis of shape of mammographic masses using global shape measures convexity, circularity and compactness has been done in many of the research studies. The tumors from ultrasonic images were segmented using level set method at first and six morphologic features were extracted including convexity and roundness. These features were used along with support vector machine (SVM) to classify tumors into benign and malignant [7]. Major portions of mammographic tumor boundaries were separated using boundary segmentation method and two features speculation index and fractional concavity along with modified compactness was found using iterative procedure for polygonal modelling of the mass boundaries. These features were used later for the classification process [8]. A feature vector based on boundary analysis to get three features Radial Distance Measure (RDM), convexity and angular measure was proposed and k-Nearest Neighbor was used as a classifier to distinguish healthy from pathological records [9]. The author used Circularity along with texture features and Radial Angle which is the smaller included angle between radial and gradient direction of the edge [10]. A turning angle function was demonstrated [11], which included calculating turning angle, the angle formed by intersection between tangent function and the horizontal. Tumor circularity and surface roughness was used to classify breast tumors [12]. Hence compactness, circularity and convexity have been used as shape

features for breast tumor classification which show good results when used in combination with other shape features. These global descriptors when used as a standalone shape descriptors or features lead to an average classification performance and hence are usually combined with other shape descriptors to discriminate shapes [13].

The goal of this paper is to further improve classification results using the improved proposed shape features. Support Vector Machine (SVM) is a typical method that is based on a hyper-plane classifier, which classifies two classes by maximizing the margin [14]. Due to its remarkable generalization performance, support vector machine (SVM) has attracted attention and gained extensive application in many fields. Support vector machines (SVM) have been used widely used for face detection in images [15], object recognition [16], handwritten digit recognition [17]. The author used support vector machine to classify images based on histogram of images with good performance, because of the superior generalization ability of SVM's in high-dimensional spaces [18]. In this research, SVM used as a machine learning tool to classify breast cancer into two different classes - benign and malignant.

The main goal of this paper is to improve overall performance, specificity and sensitivity of breast cancer classification to predict malignancy of suspicious areas using proposed shape features. In this paper, I use extracted tumor binary images as input files for the proposed method. I propose one new shape features as an improvement of Circularity Range Ratio. Using this feature I can implement the classification into benign or malignant cases using SVM as a machine learning tool. Then, I compare Circularity Max with the existing shape measures Circularity Range Ratio. The organization of rest

of this paper is as follows. Section 2 shows overall system of Image processing of breast cancer. Section 3 explains the existing shape features Circularity Range Ratio and Support Vector Machine (SVM). Section 4 describes the proposed shape features. Section 5 and 6 present the experimental results and data analysis and finally, conclusions are presented in section 7.

2. MODEL DESCRIPTION

Figure 1 shows the different phases of the overall system schemes. A mammogram image contains both useful and non-useful information. The image segmentation part divides an image into two parts: the Region of Interest (ROI) that is an abnormal region on the mammogram and the healthy tissue region. There are some popular segmentation methods that focus on threshold, or gradients. In this research, expert observers and radiologists were consulted to find ROIs of mammogram images, and feature classification with proposed method Circularity Max is focused in this paper.

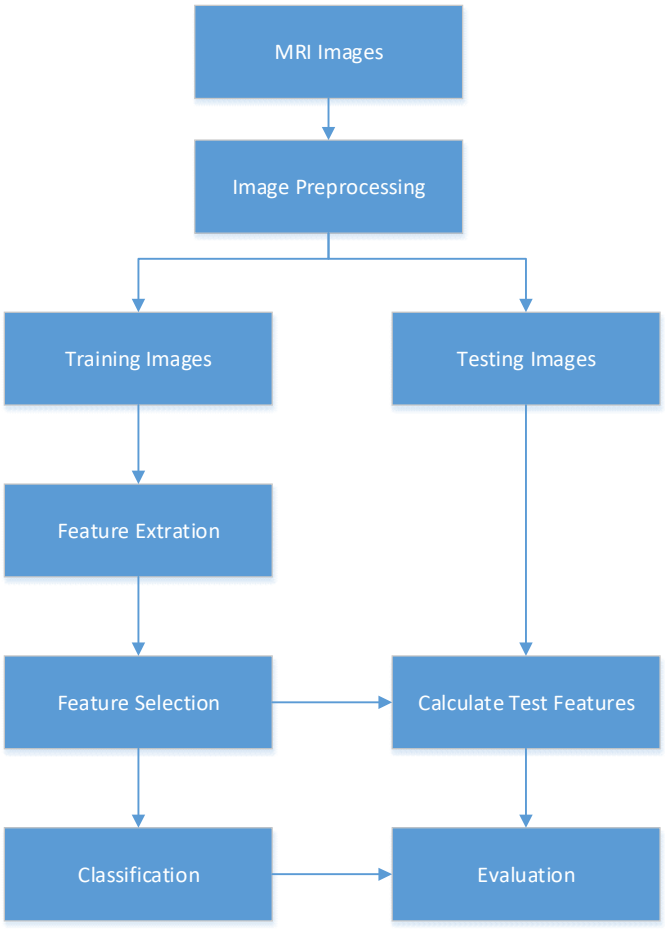

Figure 1 Overall System

First, preprocessing of the digitized mammogram image can suppress noise and improve the contrast of the image. Second, image dataset is divided into two groups: training data and test data. In the third phase, features are extracted using training data, and the best subset of them are selected as classifier input. Finally, the detection/classification of masses will be conducted.

2.1 Image Pre-Processing

The pre-processing stage is used to enhance image quality and to represent tumor objects more reliably, by removing unwanted parts and enlarging the intensity difference between objects and background. This work is used an intensity based pre-processing method to remove unwanted objects in mammogram images. A Wiener filter is used to remove noise. The Wiener filter is an adaptive local linear low pass filter that minimizes the Mean Squared Error (MSE). The Wiener filter tries to build an optimal estimate of the original image by enforcing a minimum Mean Square Error constraint between the estimate and the original image [4],[6].

2.2 Feature Extraction

The feature extraction step calculates all possible features that are expected to be effective in finding abnormalities in mammogram images [10]. In this research, ROI features extraction, including shape and texture features were used.

Shape features are based on the shapes of ROIs. Perimeter, area, and circularity are three different shape features in this study. Circularity shows how closely the shape of

the ROI approaches a circle and is calculated according to following equation. The circularity value is 1 for a circle and 0 for a line.

$$Circularity = \frac{4\pi * Area}{Perimeter^2} \quad (1)$$

The circularity is explained in section 3.

2.3 Feature Selection

Feature selection is the process of selecting an optimum subset of features in order to remove redundant features and reduce the complexity of the classifier [6], [10].

The correlation based method which is used in this study evaluates a different subset of features and ranks them based on Pearson correlation value. Good feature subsets contain features highly correlated with the class, yet uncorrelated with each other. At the first step, the feature-feature and feature-class correlation values were calculated from the training data. Then, a score of a subset of features was assigned base on following equation [26].

2.4 Classification

Breast cancer detection without determining tumor can be considered a binary classification. During training, a feature extractor is used to convert each input value to a feature value. These feature sets are fed into the binary classification algorithm. During prediction, the same feature extractor is used to convert test inputs to feature sets. These feature sets are fed into the model that generates the detected class.

3. EXISITING METHODS

This section introduces some existing methods of shape features such as Convexity, Circularity, and Compactness which have been used for classification.

3.1 Convexity

A convex hull can be considered as an elastic ribbon that stretches around the contour of an object and is the minimal convex covering of an object. Convexity can be defined as the ratio of perimeter of convex hull that wraps around the tumor shape to that of the perimeter of the original tumor shape [19].

$$Convexity = Perimeter_{convexhull} / Perimeter_{shape} \quad (2)$$

where $Perimeter_{convexhull}$ is the perimeter of the convex hull enclosing the tumor shape and $Perimeter_{shape}$ is the actual tumor shape perimeter. If a tumor mass is round, its convexity tends to be near 1 where as a mass with speculated edge will have a convexity smaller than 0.5[6]. Several studies have been done using convexity as a global shape feature and it has been used with other shape features for good classification results [6-7] [20].

The goal of this research is to propose a new shape feature based on convex hull that utilizes both the area information of convex hull shape that wraps the tumor shape and the tumor shape itself as well as contour points information from convex hull rather than just the perimeter information so that the new proposed shape improves the classification results in comparison to the traditional convexity shape measure.

3.2 Circularity

Circularity ratio is another global shape based feature that describes how the tumor shape is similar to that of a circle and can be helpful in determining the regularity of a given mass [19]. The higher the circularity is, the more circular the object tends to be. When circularity is higher the probability of masses as being benign is higher. Circularity is given by

$$Circularity = Area_{shape} / Area_{cicle} \quad (3)$$

where $Area_{shape}$ is the area of tumor and $Area_{circle}$ is the area of circle that has same perimeter as that of tumor. Since

$$Area_{circle} = (Perimeter_{shape})^2 / 4\pi \quad (4)$$

he formula can be changed

$$Circularity = 4\pi * Area_{shape} / (Perimeter_{shape})^2 \quad (5)$$

As 4π is constant, the equation is re-written as

$$\text{Circularity} = \text{Area}_{\text{shape}} / (\text{Perimeter}_{\text{shape}})^2 \quad (6)$$

Several studies have used circularity as a shape feature which has been used in conjunction with other shape features for breast tumor classification [6-7] [20]. A study suggests that circularity ratio did not serve as an efficient classifier [21]. So the goal of this research is to propose a new shape feature based on ratio of the largest and smallest circle that have largest distance and smallest distance from the centroid of the tumor as radius of circle respectively. An irregular shape will have a greater variation between the

maximum and minimum area and the variation would decrease as the shape becomes more oval and round.

3.3 Compactness

To measure the performance of the proposed shape feature irregularity ratio, a global shape feature called compactness used, which is further modified to restrict the range of value obtained from 0 to 1. Compactness is given by

$$C = P^2 / A \tag{7}$$

where P is the perimeter of the tumor boundary and A is the area of the tumor mass. A benign tumor mass is expected to have a less complexity value compared to that of a malignant tumor. The modified version of compactness, which is used to restrict the value from 0 to 1 [22] is given by

$$C = 1 - 4 * \pi * A / P^2 \tag{8}$$

A value of C=0 represents a circle. The value of C increases as the contour becomes more irregular and the shape complexity increases and can go up to a maximum value of 1. A study has indicated that modified measure of compactness provides efficient tumor classification accuracy compared comparable to those given by other shape factors based on Fourier descriptors, moments and others [23]. Hence compactness was chosen as a shape measure to compare the efficiency of the proposed shape feature irregularity ratio.

3.4 Circularity Range Ratio

Distance can be defined as the difference between the largest and smallest values for a set of numbers. As the shape is more regular, it must have an almost even distribution of distances which means that the difference between the largest and smallest distance is small. An irregular malignant tumor must have a greater difference between smallest and largest distance since there is more uneven distribution of distances. Using this idea, I propose a new shape feature called circularity range ratio. Circularity range ratio can be used to improve the performance of global shape feature circularity, which uses the variation in shape area rather than just the tumor shape area and boundary information used in circularity. In this research, I take the largest and smallest from the set of distances, dmax and dmin respectively. Using these two numbers as a radius, I can get area of two circles Amax and Amin. Circularity range ratio can then be defined as:

$$CR_{ratio} = A_{max} / A_{min} \quad (9)$$

Figure 1 shows an example of circularity range ratio extracted for two binary object images. For a benign image, since it has less difference between dmax and dmin there will be less difference between Amax and Amin which will lead to a less CRratio value. For malignant image the more irregular the shape the greater the CRratio becomes as the area difference is greater due to the big difference between dmax and dmin.

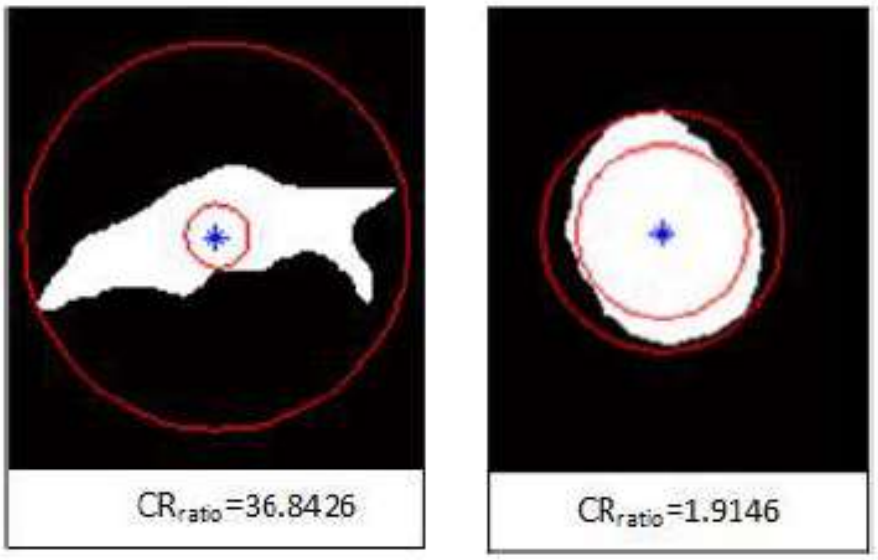

Figure 2 A malignant tumor (left) with CR_{ratio} = 36.8426 and a benign tumor with CR_{ratio} = 1.9146

3.5 Support Vector Machines (SVM)

In machine learning, support vector machines are learning models that have been used extensively to analyze data and recognize patterns and are used for classification and regression analysis [15-18]. Support vector machines are also powerful classification method which has been used successfully in many real-world problem such as medical diagnosis. In this research, SVM is used because of its good generalization performance even in case of high-dimensional data and a small set of training patterns and their ability to find non-linear solutions efficiently using the kernel functions where the data is mapped into a high-dimensional space in which the problem becomes linearly separable. SVM takes a set of input data (xi), does training using SVM training algorithm and classifies whether the new data belongs to two possible classes (yi). It then constructs a hyper-plane (H) to separate these classes, which can be used for classification, regression, or other tasks. Linear SVM as most simple type can be formulated [24]:

Find **w** and b such that $\mathbf{\Phi}(\mathbf{w}) = \mathbf{w}^T\mathbf{w}$ is minimized and for all

$$(\mathbf{x}i, yi),\ i=1..n : yi\ (\mathbf{w}^T\mathbf{x}i + b) \geq 1 \qquad (10)$$

w is the vector of coefficients, b is a constant, and the index i labels the n training cases and yi represents the class labels and xi represents the independent variables. The kernel **Φ** is used to transform data from the input to the feature space. But in most cases when the ratio between the number of the target (positive) and non-target (negative) training instances significantly differs from the 1:1 ratio, the dataset becomes unbalanced as the number of the negative data instances is much higher than the number of the positive data instances and hence slack variables ξi and

parameter C are added to allow misclassification and to control over fitting respectively and the formula can be rewritten as [25]:

Find **w** and b such that $\mathbf{\Phi}(\mathbf{w}) = \mathbf{w}^T\mathbf{w} + C\Sigma\xi i$ is minimized and for all

$$(\mathbf{x}i, yi),\ i=1..n : yi\ (\mathbf{w}^T\mathbf{x}i + b) \geq 1 - \xi i, \ , \xi i \geq 0 \qquad (11)$$

w is the vector of coefficients, b is a constant, The index i labels the n training cases. Here, y represents the class labels and xi represents the independent variables. The kernel **Φ** is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting. High values of C will largely penalize misclassified examples, and therefore the resulting hyper-plane will be one that strongly avoids classification errors, even when sacrificing generalization. Ultimately, a will lead into a hard-margin SVM behavior. On the other hand, low values only lightly penalize misclassifications, and the result might be an erroneous separation [26].

Since most of the training sets that I work with are linearly non separable, even if I introduce slack variables classification results might not be optimal. So instead of using slack variables the data can be transferred from low dimensional feature space into a high dimensional feature space where the training set is separable. Kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyper plane in the high-dimensional feature space is chosen. In this research, nonlinear SVM used with Gaussian radial basis kernel as classifier where C and σ are 3 and 0.074 respectively.

4. PROPOSED METHOD

In this paper, Support Vector Machine (SVM) method is used as an extraction tool to classify mammogram into fatty, glandular, muscle, and tumor tissue and convert those images as binary object image with one object inside [27]. And then the obtained binary object images are used to calculate the proposed shape features.

Fiugre 3 Examples of extracted binary object images with centroid

4.1 Distance Pixel Set

A distance pixel set is simply a collection or set of all distances from the centroid of the breast tumor to each boundary pixels. We assume that an extracted benign tumor contour would have less number of pixels and its distance variation would be more regular in Figure 4, whereas a malignant one would have more number of pixels and the distance would be more uneven and randomly spread out in Figure 5. From the histogram, all the features extracted for this research, namely Mean and Distances. Distance can be defined as the difference between the largest and smallest values for a set of numbers. It is an indication of statistical dispersion. Hence, distance for a malignant tumor will have greater value than that for a benign tumor.

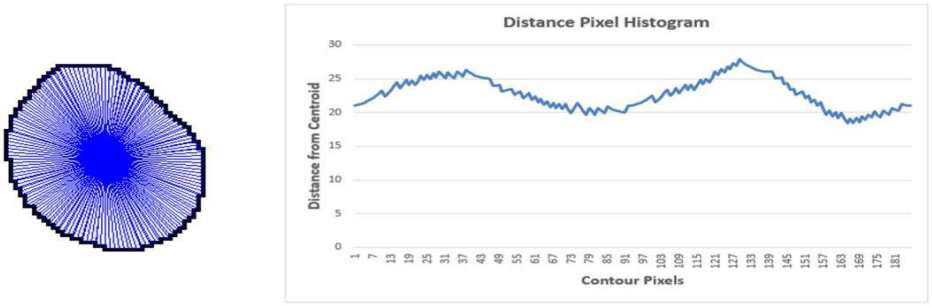

Figure 4 A Benign Tumor with its Distance Pixel Histogram

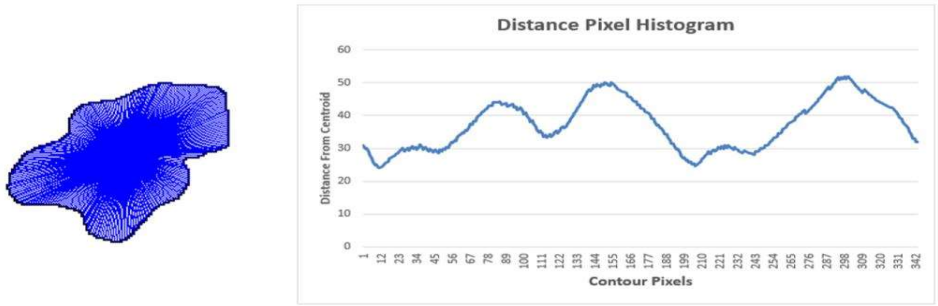

Figure 5 A Malignant Tumor with its Distance Pixel Histogram

If there are 0….N pixels that make up the tumor boundary then there will be (d0,….,dN) set of distances. This information is used to calculate features using mostly the maximum and minimum distances dmax and dmin.

4.2 Circularity Max

In this paper, circularity max method is proposed to classify if the shape is irregular or circular and to make up for weakness of circularity range ratio. Maximum and minimum distances of irregular shape from the centroid by using Distance Pixel Set is used as a radius of shapes. When using this data, the area of irregular shape and circle can be calculated in the matlab with built-in functions. Figure 6 shows that distance of irregular shape from its centroid and it can be a radius of the circle. Now I can calculate the Cmax value with the following formula:

$$Cmax = \frac{Area\ of\ shape}{Area\ of\ circle} \quad (12)$$

If Cmax value is getting smaller, it means the irregular shape would be irregular. However, if Cmax value is getting bigger up to 1, the shape is close to a circle. It means the shape is more close to a circle. The important thing is that Cmax value cannot exceed 1.

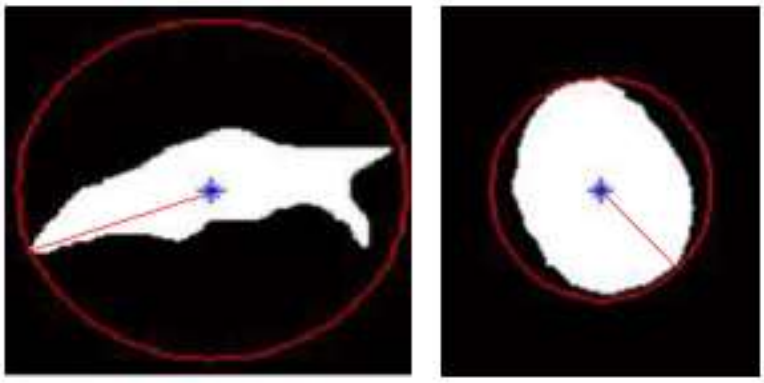

Figure 6 The distance of the irregular shape from its centroid of Circularity Max

Circualrity range ratio has a problem that has to be made up. Circularity max can make up for its weakness point. Figure 7 shows a weakness of the circularity range ratio with CRratio value. The image is a benign tumor that a doctor diagnosed. However, CRratio value is 28.4371 in this case. It means that the shape is a malignant. Even if the shape in this image looks like a circular shape that a doctor diagnosed a benign, the CRratio value says that it is close to an irrgular shape with circularity range ratio. However, Cmax value is 0.8933 and it does not exceed 1. It means that there is high probability of the benign tumor case. From this result, we can see that circularity range ratio have a flaw. This flaw can be solved with circularity max. The threshold of

circularity max is 0.6500. 100 benign images and 100 malignant images in database were randomly picked for training data sets. These images were trainied in SVM to find a proper threshold. When the threshold was set 0.5000, the error factor was 15.8125%. However, when the threshold was set 0.6500, the error factor was 3.5813%. Therefore, 0.6500 was set as a threshold for circularity max.

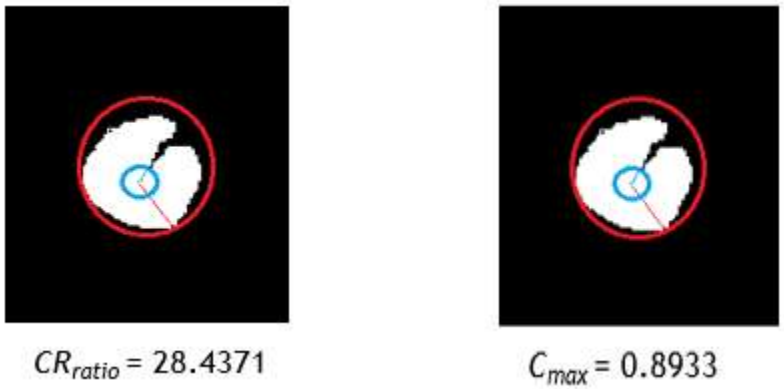

Figure 7 CRratio value (left) and Cmax value (right)

5. EVALUATION

In this section, a total of 1000 mammographic images were used to evaluate the performance of the proposed shape features. All of the files of training and test data are from DDSM [28] which also gives the information such as abnormality, lesion type, pathology and so on. From the pathology information, I know whether the breast cancer is benign or malignant. Out of those 1000 mammographic images, 350 images were used for benign training case and 350 were used for malignant test cases using SVM as learning tool. Then after the model was trained, 150 images were used for benign test cases and 150 for malignant test cases. The proposed shape feature was extracted for every single image and the exact same database consisting of 1000 mammographic images were used to extract feature for Circularity Range Ratio and Circularity Max. To evaluate the performance of breast cancer classification, MCC, Specificity, Sensitivity and Accuracy were used which are given as:

- The Matthews Correlation Coefficient (MCC) is used as a measure of the quality of binary classifications. The MCC is in essence a correlation coefficient between the observed and predicted binary classifications; it returns a value between −1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than a random prediction, and −1 indicates a total disagreement between prediction and observation.

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+F\ \ )*(TP+FN)*(TN+FP)*(TN+\ \ \ )}} \qquad (13)$$

- Sensitivity relates to the test's ability to identify positive results.

$$Sensitivity = \frac{TP}{TP+FN} \quad (14)$$

- Specificity relates to the test's ability to identify negative results.

$$Specificity = \frac{TN}{TN+FP} \quad (15)$$

- Accuracy is the proportion of true results (both positive and negative) in the population.

$$Accuracy = \frac{TP+T}{TP+FN+FP+TN} \quad (16)$$

Here, true positives (TP) is the data set that has been correctly detected as benign by the algorithm and true negatives (TN) is the data set that has been correctly identified as malignant by the algorithm. Similarly false negatives (FN) is the data set that has been incorrectly detected as malignant by the algorithm while they are in fact benign, and false positives (FP) is the number of samples that has been incorrectly detected as benign by the algorithm while they are normal.

MCC is a correlation coefficient between the observed and predicted binary classifications and a value of 1 corresponds to a perfect correlation. It is one of the best matrixes to describe the binary classification results. Specificity indicates how much probability I can say the tested negative class is actually negative class and sensitivity indicates the probability that the tested positive class is actually positive class. Sensitivity gives the fraction of the actual positives and specificity denotes the fraction of the actual negatives that has been correctly predicted. For the confusion matrix, each column of the

matrix represents the instances in a predicted class, while each row represents the instances in an actual class.

		Condition as determined by *Dr*		
		TRUE	FALSE	
Test Outcome	Positive	True Positive (TP)	False Positive (FP)	Precision
	Negative	True Negative (TN)	False Negative (FN)	Negative predictive value
		Sensitivity or Recall	Specificity	Accuracy

Figure 8 Confusion Matrix

6. RESTULT DATA

Table 1 and 2 gives the confusion matrix for Circularity Max and circularity range ratio shape feature. From table 8, Circularity Max improves the MCC, sensitivity, specificity and accuracy by 6%, 3%, 3.3% and 3% over the exisiting method circularity range ratio shape feature (700 Training images and 300 Testing images). From the results, the result data is related to the number of training data set and testing data set.

	Benign	Malignant
Benign	72	78
Malignant	93	57

	Benign	Malignant
Benign	65	85
Malignant	87	63

Table 1 Circularity Max (left) and Circularity Range Ratio (right): 100 Training images and 300 Testing images

	Benign	Malignant
Benign	99	51
Malignant	80	70

	Benign	Malignant
Benign	92	58
Malignant	75	75

Table 2 Circularity Max (left) and Circularity Range Ratio (right): 200 Training images and 300 Testing images

	Benign	Malignant
Benign	115	35
Malignant	65	85

	Benign	Malignant
Benign	111	39
Malignant	68	83

Table 3 Circularity Max (left) and Circularity Range Ratio (right): 300 Training images and 300 Testing images

	Benign	Malignant
Benign	120	30
Malignant	61	89

	Benign	Malignant
Benign	117	33
Malignant	64	86

Table 4 Circularity Max (left) and Circularity Range Ratio (right): 400 Training images and 300 Testing images

	Benign	Malignant
Benign	127	23
Malignant	54	96

	Benign	Malignant
Benign	124	26
Malignant	61	89

Table 5 Circularity Max (left) and Circularity Range Ratio (right): 500 Training images and 300 Testing images

	Benign	Malignant
Benign	129	21
Malignant	51	99

	Benign	Malignant
Benign	126	24
Malignant	56	94

Table 6 Circularity Max (left) and Circularity Range Ratio (right): 600 Training images and 300 Testing images

	Benign	Malignant
Benign	132	18
Malignant	48	102

	Benign	Malignant
Benign	128	22
Malignant	53	97

Table 7 Circularity Max (left) and Circularity Range Ratio (right): 700 Training images and 300 Testing images

100 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	-0.147	0.433	0.42	0.427	11.55
CMax	-0.141	0.48	0.38	0.43	11.43

200 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.114	0.613	0.5	0.557	13.13
CMax	0.129	0.66	0.467	0.563	12.81

300 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.295	0.740	0.550	0.645	15.122
CMax	0.340	0.766	0.567	0.667	14.241

400 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.360	0.780	0.573	0.670	18.475
CMax	0.402	0.800	0.593	0.677	18.392

500 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.4319	0.827	0.593	0.710	21.844
CMax	0.4974	0.847	0.64	0.743	20.371

600 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.4777	0.840	0.627	0.733	26.121
CMax	0.5307	0.860	0.660	0.760	26.002

700 / 300	MCC	Sensitivity	Specificity	Accuracy	Execution Time
Circularity Range Ratio	0.511	0.853	0.647	0.75	30.65
CMax	0.5715	0.88	0.68	0.78	29.782

Table 8 Results of Evaluation with expended image set with execution time

In table 8 and figure 9, when the training set and testing set were 100 and 300, you can see MCC value was negative. That means if we have fewer images than testing data sets, we might have worse result than we thought. From this, we have to set up much more training data set than testing data sets. In figure 10, proposed method Circularity Max has better results than existing method circularity range ratio. The more circularity

max that is proposed in this research has training data sets, the performance and accuracy will be closed to 90%. For the execution time, there is no differences between circularity range ratio and circularity max.

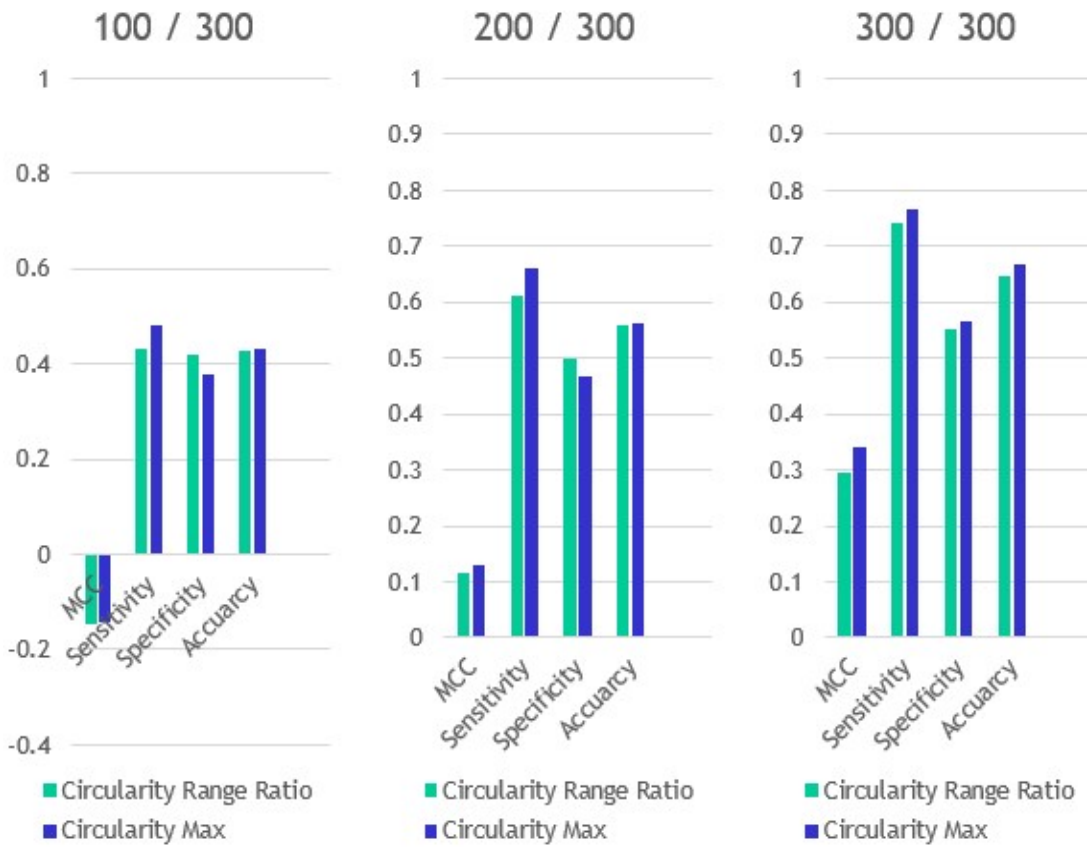

Figure 9 Graphs of Result Comparison between Circularity Range Ratio and Circularity Max

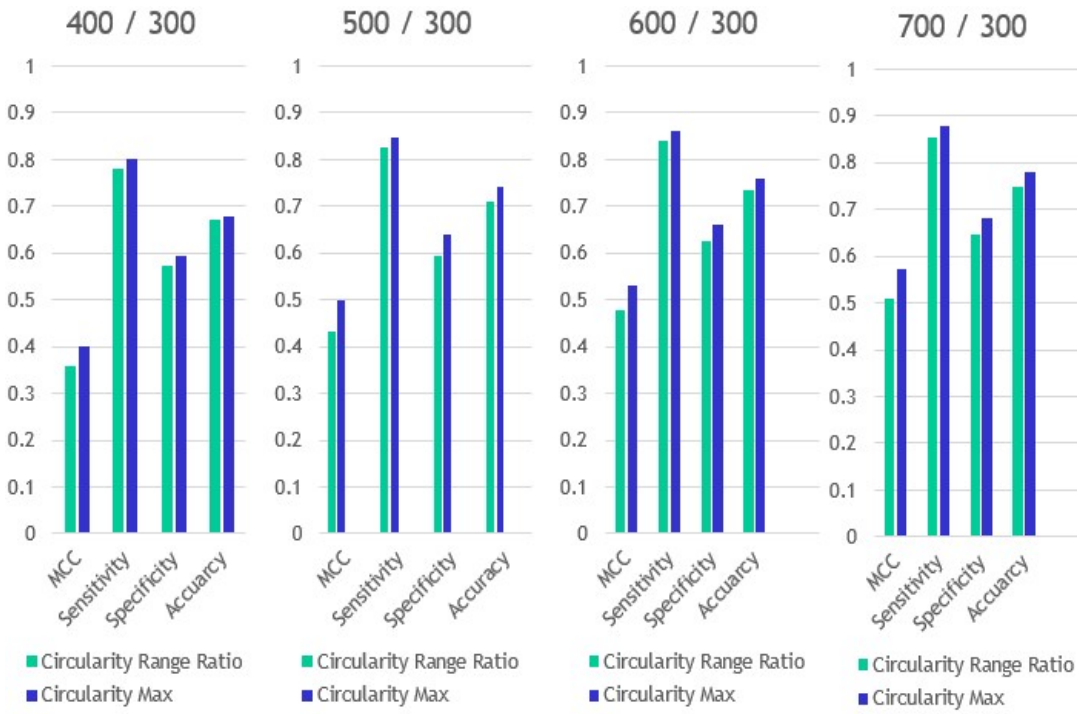

Figure 10 Graphs of Result Comparison between Circularity Range Ratio and Circularity Max

7. CONCLUSION

Breast cancer is the most common cancer among women. It is estimated that about one in 8 US women (about 12%) will develop invasive breast cancer over the course of her lifetime according to the American Cancer Society [1]. It is very important to detect breast cancer at an early stage. Breast cancer can be classified abnormalities into benign or malignant classes. Classifying breast cancer into benign and malignant categories help physicians to minimize the potential errors that can occur, thus improving breast cancer prognosis. In this paper, one new shape feature, circularity max was proposed and investigated to classify mammographic images into benign and malignant class. I utilized these shape features to perform classification into benign and malignant class using SVM. Existing shape feature which is called circularity range ratio was used to compare the efficiency of the new proposed shape feature.

The results demonstrate that the proposed circularity max method outperforms existing circularity range ratio method. The results show that circualrity max improves MCC by 6%, sensitivity by 3%, specificity by 3.3% and accuracy by 3% over the existing method circularity range ratio shape feature (700 Training images and 300 Testing images). From these results, we know that circularity max could solve the problem that circularity range ratio has. Important thing for the number of data set is that setting more training data set than testing data set should be suggested because when the trainning data set is less than testing data sets, the MCC value of result could be negative number in table 8 and Figure 9. I believe that the more circularity max that is proposed in this research has training data sets, the performance and accuracy will be close to 90%.

Execution time for every result data is calculated as well, but the execution time is not considered in this research. In table 8, there is no difference between the existing method circularity range ratio and proposed method circularity max. However, for example, if we have 10 million image data for training and testing data sets in the future, it might take one day or more for training and testing data sets. For this reason, the execution time should be considered and improved for the performance.

Therefore, by using this proposed shape feature method, a more efficient benign/malignant classification can be obtained and compared to the currently widely used. This improved efficiency can further help physicians with preliminary decision support information for further diagnosis.

8. APPENDIX: IMPLEMENTATION

A. Centerofmass.m

```
function varargout = centerOfMass(A,varargin)

narginchk(0,1);

nargoutchk(0,1);

fname = 'centerOfMass';


% Checked required inputs

validateattributes(A,{'numeric'},{'real','finite'},fname,'A',1);


%% INITIALIZE VARIABLES

A(isnan(A)) = 0;

if ~(strcmpi(class(A),'double') || strcmpi(class(A),'single'))

    A = single(A);

end

if any(A(:)<0)

    warning('MATLAB:centerOfMass:neg','Array A contains negative
values.');

end


%% PROCESS

sz = size(A);

nd = ndims(A);

M = sum(A(:));

C = zeros(1,nd);

if M==0

    C = [];

else
```

```
    for ii = 1:nd
        shp = ones(1,nd);
        shp(ii) = sz(ii);
        rep = sz;
        rep(ii) = 1;
        ind = repmat(reshape(1:sz(ii),shp),rep);
        C(ii) = sum(ind(:).*A(:))./M;
    end
end

% Assemble the VARARGOUT cell array
varargout = {C};

end % MAIN
```

B. SVMRange.m

```
%read with deliminated character from training data
dphTraining=dlmread('Training-DPH.txt',';');

group = cell(300,1);
for i=1:150
    group{i} = ['Benign'];
end
for i=151:300
    group{i} = ['Malignant'];
end

%Copy training values into their own matrix
meanTraining=dphTraining(:,1);
```

```
varianceTraining=dphTraining(:,2);

rangeTraining=dphTraining(:,3);

compactnessTraining=dphTraining(:,4);


xdata=rangeTraining;   %change matrices here


%SVM Training

svmStruct = svmtrain(xdata,group,'Kernel_Function', 'rbf', 'RBF_Sigma',
0.074,'ShowPlot',true);


benigndata=cell(100,1);

malignantdata=cell(100,1);


dphTest=dlmread('Test-DPH.txt',';');


%Copy training values into their own matrix

meanTest=dphTest(:,1);

varianceTest=dphTest(:,2);

rangeTest=dphTest(:,3);

compactnessTest=dphTest(:,4);


benigndata=rangeTest(1:100);    %change matrices here

malignantdata=rangeTest(101:200);  %change matrices here


%SVM Test

disp('Benign Test Case');

speciesbenign = svmclassify(svmStruct,benigndata,'ShowPlot',true);

disp('Malignant Test Class');

speciesmalignant =
svmclassify(svmStruct,malignantdata,'ShowPlot',true);
```

```
benigncounta=0;
malignantcounta=0;

for i=1:100
if(strcmp(speciesbenign(i),'Benign'))
    benigncounta=benigncounta+1;
else
    malignantcounta=malignantcounta+1;
end
end

benigncountb=0;
malignantcountb=0;
for i=1:100
if(strcmp(speciesmalignant(i),'Benign'))
    benigncountb=benigncountb+1;
else
    malignantcountb=malignantcountb+1;
end

end

disp('True Positive');
disp(benigncounta);
disp('False Negative');
disp(malignantcounta);
disp('False Positive')
disp(benigncountb);
disp('True Negative')
```

```
disp(malignantcountb);
```

C. SVMTraining.m

```
myFolder = 'Testing/'; %change Folder here


if ~isdir(myFolder)

    errorMessage = sprintf('Error: The following folder does not
exist:\n%s',myFolder);

    uiwait(warndlg(errorMessage));

    return;

end


filePattern = fullfile(myFolder,'*.png');

overlayFiles = dir(filePattern);

number_file = length(overlayFiles);


disp(filePattern);

disp(overlayFiles);

disp(number_file);

cella=1;

disp('number_file');

disp(number_file);

for FileNum = 1:number_file;

    baseFileName = overlayFiles(FileNum).name;

    image = fullfile(myFolder,baseFileName);

    fprintf(1,'Now reading %s\n',image);

cellb=1;


im = imread(image);

[rows columns depth]=size(im);
```

```
I = imread(image);

L = bwlabel(I);

s  = regionprops(L, 'centroid');

centroids = cat(1, s.Centroid);


%For getting convex Hull of a Binary Image

ycenter=centroids(2);

xcenter=centroids(1);

[y,x] = find(im);

dx = [-0.5 -0.5  0.5  0.5];

dy = [-0.5  0.5 -0.5  0.5];

x_corners = bsxfun(@plus, x, dx);

y_corners = bsxfun(@plus, y, dy);

x_corners = x_corners(:);

y_corners = y_corners(:);

imshow(im, 'InitialMagnification', 'fit')

hold on

hold off

imshow(im, 'InitialMagnification', 'fit')

hold on

hold off

k = convhull(x_corners, y_corners);

x_hull = x_corners(k); %x coordinate of pixel that touch the shape and
form the convex hull around the shape

y_hull = y_corners(k); %y coordinate of pixel that touch the shape and
form the convex hull around the shape

hold on

plot(x_hull, y_hull, 'r', 'LineWidth', 4) %plot convex hull around the
shape
```

```
hold off


%create a mask around the convex hull

convhullmask = poly2mask(x_hull, y_hull,rows,columns);


%gives the area of objects in binary image

roiareamask=bwarea(convhullmask);


%for calculating the convex hull perimeter

convexhullperimeter = regionprops(convhullmask,'perimeter');


C=[];

convexhullmat=[];

convexhullmat=[x_corners(k) y_corners(k)] %x_corners(k) and
y_corners(k) are the x and y coordinate of pixel that touch the shape
and form the convex hull around the shape


distance_convex=[];

all_distance_convex=[];


irrratiomat=[];

irrratiomat=[x_corners,y_corners]; %x_corners and y_corners are the
boundary pixels that form up the binary image

size_irrratiomat=size(convexhullmat);

row=1;


%calculate all the distance from centroid to the points forming up
convex hull

while(row~=size_irrratiomat(1))

    line([xcenter convexhullmat(row,1)], [ycenter
convexhullmat(row,2)], 'LineWidth', 1);
```

```
     all_distance_convex(row)=sqrt((ycenter-
convexhullmat(row,1))^2+(xcenter-convexhullmat(row,2))^2);

      row=row+1;

end


%maximum distance forming convex hull

max_d=max(all_distance_convex);


%normalizing each of the convex hull distance

row=1;

while(row~=size_irrratiomat(1))

 line([xcenter convexhullmat(row,1)], [ycenter convexhullmat(row,2)],
'LineWidth', 1);

 distance_convex(row)=sqrt((ycenter-convexhullmat(row,1))^2+(xcenter-
convexhullmat(row,2))^2)/max_d;

 row=row+1;

end


size_mean=size(distance_convex);

row=1;

sum=0;

disp(size_mean);

while(row~=size_mean(2))

    sum=sum+distance_convex(row);

    disp(sum);

    row=row+1;

end


mean_convex=sum/size_mean(2);

display(mean_convex);
```

```
noofpixels=size(all_distance_convex);

irregularindex=noofpixels(2)*mean_convex;

%EricCmax
r2 = max(distance_mat);
x = xcenter;
y = ycenter;
th = 0:pi/50:2*pi;
xunit = r2 * cos(th) + x;
yunit = r2 * sin(th) + y;
hold on
h2 = plot(xunit, yunit,'red');
hold off
area2=pi*r2*r2;
perim2=4*pi*r2;

circularityratio=area1/area2;

%Calculating the distance pixel set
BW=I;
BW_filled = imfill(BW,'holes');
boundaries = bwboundaries(BW_filled);
for k=1:1
   b = boundaries{k};
end

total=size(b);
hold on
```

```
plot(centroids(:,1), centroids(:,2), 'b*')

hold off

distance_mat=[];

for i=1:size(b)

    dist=sqrt((ycenter-b(i,1))^2+(xcenter-b(i,2))^2);

    distance_mat(i)=dist;

i=i+1;

end



%Save all the data to a file for machine learning phase

file_1 = fopen('Testing-ERIC.txt','at'); %change Filename here


%fprintf(file_1,'%s ; %f ; %f ; %f ; %f ; %f ;
%f\n',image,ccr,circularityratio,irrratio,compactness,convexity,circula
rityratio); %write to result file


fprintf(file_1,'%f;%f;%f;%f;%f;%f\n',ccr,circularityratio,irrratio,comp
actness,convexity,circularityratio); %write to result file


%fprintf(file_1,'%s ;%f ; %f\n',image,check1,check2); %write to result
file


%fprintf(file_1,'%s ; %f ; %f\n',image,convexity,circularityratio);
%write to result file


cella=cella+1;

fclose('all');


end
```

D. CircularityMax.m

```
%read with deliminated character from training data
ISFMatTraining=dlmread('Training-ERIC.txt',';');



group = cell(300,1);
for i=1:150
    group{i} = ['Benign'];
end
for i=151:300
    group{i} = ['Malignant'];
end



%Copy training values into their own matrix
circularityRangeTraining=ISFMatTraining(:,2);


xdata=circularityRangeTraining; %change matrices here


svmStruct = svmtrain(xdata,group,'Kernel_Function', 'rbf', 'RBF_Sigma',
0.074,'ShowPlot',true);


benigndata=cell(100,1);
malignantdata=cell(100,1);


ISFMatTest=dlmread('Testing-ERIC.txt',';');


%Copy training values into their own matrix
circularityRangeTest=ISFMatTest(:,2);


benigndata=circularityRangeTest(1:100); %change matrices here
```

```
malignantdata=circularityRangeTest(101:200);    %change matrices here


disp('Benign Test Case');

speciesbenign = svmclassify(svmStruct,benigndata,'ShowPlot',true);

disp('Malignant Test Class');

speciesmalignant =
svmclassify(svmStruct,malignantdata,'ShowPlot',true);

benigncounta=0;

malignantcounta=0;


for i=1:100

if(strcmp(speciesbenign(i),'Benign'))

    benigncounta=benigncounta+1;

else

    malignantcounta=malignantcounta+1;

end

end


benigncountb=0;

malignantcountb=0;

for i=1:100

if(strcmp(speciesmalignant(i),'Benign'))

    benigncountb=benigncountb+1;

else

    malignantcountb=malignantcountb+1;

end


end


%to get the perimeter of Binary image
```

```
P = regionprops(im,'perimeter');
perimeter=P(end);


%Global Shape Features Convexity,Circularity Ratio and Compactness

%disp('Global Shape Features');
%Global Shape Feature Convexity
convexity=conhullperimeter.Perimeter/perimeter.Perimeter;
%disp('Convexity = ');
%disp(convexity);
disp('True Positive');
disp('   127');
disp('   ');
%Global Shape Circularity
circularityratio=(roiarea)/(perimeter.Perimeter)^2;
%disp('circularityratio = ');
%disp(circularityratio);

disp('True Positive');
disp(benigncounta);
disp('False Negative');
disp(malignantcounta);
disp('False Positive')
disp(benigncountb);
disp('True Negative')
disp(malignantcountb);
```

E. SVMvariance.m

```
%read with deliminated character from training data

dphTraining=dlmread('Training-DPH.txt',';');




group = cell(300,1);

for i=1:150

    group{i} = ['Benign'];

end

for i=151:300

    group{i} = ['Malignant'];

end




%Copy training values into their own matrix

meanTraining=dphTraining(:,1);

varianceTraining=dphTraining(:,2);

rangeTraining=dphTraining(:,3);

compactnessTraining=dphTraining(:,4);
```

```
xdata=varianceTraining;   %change matrices here


%SVM Training
svmStruct = svmtrain(xdata,group,'Kernel_Function', 'rbf', 'RBF_Sigma',
0.074,'ShowPlot',true);


benigndata=cell(100,1);
malignantdata=cell(100,1);


dphTest=dlmread('Test-DPH.txt',';');


%Copy training values into their own matrix
meanTest=dphTest(:,1);
varianceTest=dphTest(:,2);
rangeTest=dphTest(:,3);
compactnessTest=dphTest(:,4);


benigndata=varianceTest(1:100);    %change matrices here
malignantdata=varianceTest(101:200);  %change matrices here
```

```
%SVM Test

disp('Benign Test Case');

speciesbenign = svmclassify(svmStruct,benigndata,'ShowPlot',true);

disp('Malignant Test Class');

speciesmalignant =
svmclassify(svmStruct,malignantdata,'ShowPlot',true);

benigncounta=0;

malignantcounta=0;



for i=1:100

if(strcmp(speciesbenign(i),'Benign'))

    benigncounta=benigncounta+1;

else

    malignantcounta=malignantcounta+1;

end

end



benigncountb=0;

malignantcountb=0;

for i=1:100
```

```
if(strcmp(speciesmalignant(i),'Benign'))
    benigncountb=benigncountb+1;
else
    malignantcountb=malignantcountb+1;
end


end


disp('True Positive');
disp(benigncounta);
disp('False Negative');
disp(malignantcounta);
disp('False Positive')
disp(benigncountb);
disp('True Negative')
disp(malignantcountb);
```

F. Evaluation.cpp

```
#include <iostream>
#include <cmath>

using namespace std;

void M_C_C(int , int , int , int);
void Sensitivity(int , int);
void Specificity(int , int);
void Accuracy(int , int , int , int);
int tp, tn, fp, fn;

int main()
{

	cout<<"Ture Positive: ";
	cin>>tp;
	cout<<"False Nagative: ";
	cin>>fn;
	cout<<"False Positive: ";
	cin>>fp;
	cout<<"Ture Nagative: ";
	cin>>tn;
	cout<<endl;
	M_C_C(tp, tn, fp, fn);
	Sensitivity(tp, fn);
	Specificity(tn, fp);
	Accuracy(tp, tn, fp, fn);
	return 0;
}
```

```
void M_C_C(int a, int b, int c, int d)
{
	float mcc=0;
	float num, deno, sqaure;

	num = (tp*tn)-(fp*fn);
	deno = ((tp+fp)*(tp+fn)*(tn+fp)*(tn+fn));
	sqaure=sqrt(deno);
	mcc = num / sqaure;
	cout<<"MCC: "<<mcc<<endl;

	return;
}

void Sensitivity(int a, int b)
{
	float sen=0;
	float num, deno;
	num = tp;
	deno = tp+fn;
	sen = (num/deno);
	cout<<"Sensitivity: "<<sen<<endl;

	return;
}

void Specificity(int a, int b)
{
```

```
	float spe=0;
	float num, deno;
	num = tn;
	deno = tn+fp;
	spe = (num/deno);
	cout<<"Specificity: "<<spe<<endl;

	return;
}

void Accuracy(int a, int b, int c, int d)
{
	float acc=0;
	float num, deno;
	num = tp+tn;
	deno = tp+fn+fp+tn;
	acc = (num/deno);
	cout<<"Accuracy: "<<acc<<endl;

	return;
}
```

REFERNCES

[1] U.S. Breast Cancer Statistics. (2011). [Online]. Retrieved August 11, 2016 from http://www.breastcancer.org/symptoms/understand_bc/statistics.jsp

[2] Althuis, M.D., Dozier, J.M., Anderson, W.F., Devesa, S.S., and Brinton, L.A., 2005. Global trends in breast cancer incidence and mortality 1973-1997. Int J Epidemiol Volume 34, Issue 2 (Apr), 405-412. DOI: http://dx.doi.org/10.1093/ije/dyh414.

[3] American Cancer Society, "Breast Cancer", American Cancer Society, Inc., Surveillance and Health Policy Research, Feb 2011

[4] Champaign, J.L. and Cederbom, G.J., 2000. Advances in Breast Cancer Detection with Screening Mammography. Ochsner J, Volume 2, Issue 1 (Jan), 33-35.

[5] Mercimek, M., Gulez, K., and Mumcu, T., 2005. Real object recognition using moment invariants. Sadhana 30, 6 (2005/12/01), 765-775. DOI: http://dx.doi.org/10.1007/BF02716709

[6] Boujelben, A., Tmar, H., Abid, M. and Mnif J. 2012. Automatic Diagnosis of Breast Tissue, Advances in Cancer Management, Prof. Ravinder Mohan (Ed.), ISBN: 978-953-307-870-0, InTech, DOI: 10.5772/22565. Retrieved August from http://www.intechopen.com/books/advances-in-cancer-management/automatic-diagnosis-of-breast-tissue

[7] Chang, R.-F., Wu, W.-J., Moon, W., and Chen, D.-R., 2005. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. Breast Cancer Research

[8] Rangayyan, R.M., Mudigonda, N.R., and Desautels, J.E.L., 2000. Boundary modelling and shape analysis methods for classification of mammographic masses. Medical and Biological Engineering and Computing 38, 5 (2000/09/01), 487-496. DOI: http://dx.doi.org/10.1007/BF02345742.

[9] Boujelben, A., Chaabani, A.C., Tmar, H., and Abid, M., 2009. Feature Extraction from Contours Shape for Tumor Analyzing in Mammographic Images. In Digital Image Computing: Techniques and Applications, 2009. DICTA '09., 395-399. DOI: http://dx.doi.org/10.1109/DICTA.2009.71.

[10] Yang, S.-C., Wang, C.-M., Chung, Y.-N., Hsu, G.-C., Lee, S.-K., Chung, P.-C., and Chang, C.-I., 2005. A Computer-Aided System for Mass detection and Classification in Digitized Mammograms. Biomedical Engineering: Applications, Basis and Communications on Volume 17, Issue 05, 215-228. DOI: http://dx.doi.org/doi:10.4015/S1016237205000330

[11] Guliato, D., De Carvalho, J.D., Rangayyan, R.M., and Santiago, S.A., 2008. Feature Extraction from a Signature Based on the Turning Angle Function for the Classification of Breast Tumors. Journal of Digital Imaging 21, 2 (10/31 05/16/received 07/22/revised 08/05/accepted), 129-144. DOI: http://dx.doi.org/10.1007/s10278-007-9069-9.

[12] Pohlman, S., Powell, K., Obuchowski, N., Chilcote, W., and Grundfest-Broniatowski, S., 1995. Classification of breast lesions based on quantitative measures of tumor morphology. In Engineering in Medicine and Biology Society, 1995. IEEE 17th Annual Conference, 537-538 vol.531. DOI: http://dx.doi.org/10.1109/IEMBS.1995.575238.

[13] ZHANG, D. and LU, G., 2004. Review of shape representation and description techniques. Pattern Recognition 37, 1 (1//), 1-19. DOI: http://dx.doi.org/http://dx.doi.org/10.1016/j.patcog.2003.07.008.

[14] Chun-fu, L. and Sheng-de, W., 2002. Fuzzy support vector machines. Neural Networks, IEEE Transactions on Volume 13, Issue 2, 464-471. DOI: http://dx.doi.org/10.1109/72.991432.

[15] Osuna, E., Freund, R., and Girosi, F., 1997. Training support vector machines: an application to face detection. In Computer Vision and Pattern Recognition, 1997. Proceedings, 1997 IEEE Computer Society Conference on, 130-136. DOI: http://dx.doi.org/10.1109/CVPR.1997.609310.

[16] Pontil, M. and Verri, A., 1998. Support vector machines for 3D object recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 20, Issue 6, 637-646. DOI: http://dx.doi.org/10.1109/34.683777.

[17] Neves, R.P., Zanchettin, C., and Filho, A.G.L., 2012. An Efficient Way of Combining SVMs for Handwritten Digit Recognition. In Artificial Neural Networks and Machine Learning – ICANN 2012, A.P. Villa, W. Duch, P. Érdi, F. Masulli and G. Palm Eds. Springer Berlin Heidelberg, 229-237. DOI: http://dx.doi.org/10.1007/978-3-642-33266-1_29.

[18] Chapelle, O., Haffner, P., and Vapnik, V.N., 1999. Support vector machines for histogram-based image classification. Neural Networks, IEEE Transactions on Volume 10, Issue 5, 1055-1064. DOI: http://dx.doi.org/10.1109/72.788646.

[19] Yang Mingqiang, Kpalma Kidiyo and Ronsin Joseph (2008). A Survey of Shape Feature Extraction Techniques, Pattern Recognition Techniques, Technology and Applications, Peng-Yeng Yin (Ed.), ISBN: 978-953-7619-24-4, InTech, DOI: 10.5772/6237. Retrieved from http://www.intechopen.com/books/pattern_recognition_techniques_technology_and_applications/a_survey_of_shape_feature_extraction_techniques

[20] Huang, Y.L., Chen, D.R., Jiang, Y.R., Kuo, S.J., Wu, H.K., and Moon, W.K., 2008. Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. Ultrasound Obstet Gynecol on Volume 32, Issue 4 (Sep), 565-572. DOI: http://dx.doi.org/10.1002/uog.5205.

[21] Lim, M., Hou, A., Congdon, N., and Chua, J., 2013. Feature Identification for Colon Tumor Classification. SIAM Undergraduate Research Online (SIURO) Volume 6 DOI: http://dx.doi.org/10.1137/13S012212

[22] Shen, L., Rangayyan, R.M., and Desautels, J.E.L., 1993. Detection and Classification of Mammographic Calcifications. International Journal of Pattern Recognition and Artificial Intelligence on Volume 07, Issue 06, 1403-1416. DOI: http://dx.doi.org/doi:10.1142/S0218001493000686.

[23] Rangayyan, R.M., El-Faramawy, N.M., Desautels, J.E.L., and Alim, O.A., 1997. Measures of acutance and shape for classification of breast tumors. Medical Imaging, IEEE Transactions on Volume 16, Issue 6, 799-810. DOI: http://dx.doi.org/10.1109/42.650876.

[24] Moon, J., Kim, H.I., Choi, H.D., and Jeon, S.I., 2012. A support vector machine based classifier to extract abnormal features from breast magnetic resonance images. In

Proceedings of the Proceedings of the 2012 ACM Research in Applied Computation Symposium (San Antonio, Texas2012), ACM, 2401637, 148-152. DOI: http://dx.doi.org/10.1145/2401603.2401637.

[25] Damaševičius, R., 2010. Optimization of SVM parameters for recognition of regulatory DNA sequences on Volume 18, Issue 2 (2010/12/01), 339-353. DOI: http://dx.doi.org/10.1007/s11750-010-0152-x.

[26] Gaspar, P., Carbonell, J., and Oliveira, J.L., 2012. On the parameter optimization of Support Vector Machines for binary classification. J Integr Bioinform on Volume 9, Issue 3, 201. DOI: http://dx.doi.org/10.2390/biecoll-jib-2012-201.