Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines

Jacob Levman,

Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada (phone: 416-480-6100 x3390; fax: 416-480-5714)

Tony Leung,

Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

Petrina Causer,

Radiologist with Sunnybrook Health Sciences Centre, Toronto, ON, Canada

Don Plewes, and

Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Anne L. Martel

Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Abstract

Early detection of breast cancer is one of the most important factors in determining prognosis for women with malignant tumours. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has been shown to be the most sensitive modality for screening high-risk women. Computer-aided diagnosis (CAD) systems have the potential to assist radiologists in the early detection of cancer. A key component of the development of such a CAD system will be the selection of an appropriate classification function responsible for separating malignant and benign lesions. The purpose of this study is to evaluate the effects of variations in temporal feature vectors and kernel functions on the separation of malignant and benign DCE-MRI breast lesions by support vector machines. We also propose and demonstrate a classifier visualization and evaluation technique. We show that support vector machines provide an effective and flexible framework from which to base computer-aided diagnosis techniques for breast MR imaging, and that the proposed classifier visualization technique has potential as a mechanism for the evaluation of classification solutions.

Index Terms

Classification; Breast Imaging; Magnetic Resonance Imaging; Support Vector Machines; Visualization

I. Introduction

Genetic mutations on the BRCA1/2 genes can result in an 85% lifetime risk of developing breast cancer [1]. Regular screening has been identified as key for improving survival rates [2]. The research presented in this paper was conducted at Sunnybrook Health Sciences Centre where ongoing screening trials are comparing the performance of dynamic contrast enhanced magnetic resonance imaging (DCE-MRI), ultrasound, mammography and clinical breast examination as diagnostic screening methodologies for this population [3]. Results have indicated that magnetic resonance imaging is the most sensitive modality evaluated. Furthermore, MRI does not expose the patient to ionizing radiation. This paper will be limited to consideration of dynamic contrast-enhanced magnetic resonance imaging only.

In order to further improve the DCE-MRI based breast cancer screening process significant research has been conducted towards developing a computer aided diagnosis system to assist radiologists in image analysis (see figure 1 for sample DCE-MR subtraction images of malignant and benign lesions). Given that the aforementioned screening trials [3] employ an imaging protocol that generates five three dimensional (3D) volumes each containing 28–32 slices per breast, a motivating factor for the development of a computer-aided diagnostic system is to reduce the complexity of the data for radiological analysis. A key component of such a computer-aided diagnostic system will be the selection of an appropriate classification algorithm responsible for separating malignant and benign lesions.

Significant research has been conducted on the application of computer classification methods to the analysis of breast MR images. Artificial neural networks have been one of the most common approaches for researching the classification of malignant and benign breast MR lesions [4]–[10]. Jacobs *et al.* have proposed the use of self organizing maps, an unsupervised learning approach using artificial neural networks, to assist in the separation of malignant and benign breast MR lesions [11]. Linear discriminant analysis, a classical classification approach has also been applied to breast MR image analysis [12]–[13]. Research into the classification of breast MR lesions based on a signal intensity time curve's wash-in and wash-out characteristics [14]–[15] has led to commercial products in use by radiologists (Confirma Inc., Kirkland WA USA).

Support vector machines (SVMs) have been shown to perform well as a computer-aided diagnostic classification mechanism for breast cancer screening in ultrasound [16] and mammography [17]. More recently it has been shown that SVMs outperform a variety of other machine learning techniques when applied to the separation of malignant and benign DCE-MR breast lesions [18]–[19].

In addition to classification research, this paper is also concerned with the evaluation of given classifiers through visualization techniques. Some related research has been conducted on this topic. Nattkemper and Wismuller have researched the use of self organizing maps for tumour feature visualization from breast MR data [20]. Komura *et al.* have proposed multidimensional support vector machines as a mechanism for the visualization of high dimensional data sets and demonstrated the technique with gene expression data [21].

Somorjai and Dolenko have proposed the visualization of high dimensional data onto a special plane called the relative distance map [22].

The purpose of this study is to thoroughly evaluate the use of support vector machines as a mechanism for the design of a classifier that separates malignant and benign breast lesions from DCE-MR images. We also describe a new classifier visualization technique that will assist in the evaluation of given classification design options.

II. Materials and Methods

A. Image Acquisition

The screening protocol used is as follows. Simultaneous bilateral magnetic resonance imaging was performed using a 1.5T magnet (GE Signa, version 11.4). Sagittal images were obtained with a phased-array coil arrangement using a dual slab interleaved bilateral imaging method [23]. This provided 3D volume data over each breast obtained with an RF spoiled gradient recalled sequence (SPGR, scan parameters: TR/TE/angle=18.4/4.3/30°, 256×256×32 voxels, FOV: 18×18×6–8cm). Imaging is performed before and after a bolus injection of 0.1 mmol/kg of Gd-DTPA. Each bilateral acquisition was obtained in 2 minutes and 48 seconds. Slice thickness was 2 to 3 mm.

A total of 94 DCE-MRI breast examinations from high risk patients were obtained containing lesions pathologically proven to be malignant (24 cases) or benign (70 cases). The quantity and pathological diagnosis of the different lesions addressed in this study are provided in Table I. Ground truth is based on the diagnosis of the histopathologist, who analyzes the tissue biopsies. In cases where a patient with a suspicious lesion did not receive a biopsy but returned to screening for greater than one year without observed changes to the lesion, a benign diagnosis is accepted. These cases are included in Table I as having received a pathological diagnosis of “Benign by Assumption”. For the purposes of this study we have selected only those cases where the radiologist has ordered a follow-up imaging examination after attempting to diagnose the screening case. Thus we have selected a set of exams that a radiologist found difficult to separate.

B. Image Registration

Image registration is the process of aligning images that vary in position over time. This is performed in order to compensate for any patient motion that takes place during the examination. For this study we have used a 3D non-rigid registration technique for breast MR images [24].

C. Data Preprocessing and Feature Vectors

The boundary of each radiologically identified lesion was manually delineated and all of the pixels within the region of interest (ROI) were averaged together to form a single signal intensity time-series vector per lesion. The ROI was drawn around the most enhancing area of the lesion in two dimensions on the slice where the lesion is most visible in order to avoid non-lesion and necrotic tissues. The resultant curves were used in the leave-one-out validation trials presented in section III.A. Those same regions of interest were also used to

extract each signal intensity time curve within the ROI for use in the more clinically viable approach discussed in section III.B. In this approach 5×5 median neighbourhood filtering was performed prior to extracting each time-curve in order to suppress noise. In both cases each vector has five signal intensity (SI) values corresponding to the SI obtained in the pre-contrast image and the 4 post-contrast images. Sample time curves of malignant and benign lesions plotted from our data set are provided in figure 2. We obtained a total of 94 vectors (24 cancer, 70 benign) for use in our leave-one-out classification study (section III.A) and 2544 vectors (417 cancer and 2127 benign) for use in our randomized trials (section III.B).

The input to any classification algorithm consists of a specific feature vector extracted from the available data. We have experimented with 4 different feature vectors related to this problem each of which is based on an individual voxel's signal intensity (SI) time curve, where $s(n)$ denotes the signal intensity at time $t_n = n * 168$, $n = \{0, 1, 2, 3, 4\}$ in seconds. The four feature vectors are as follows:

1st Feature: Relative Signal Intensities. The available 5 time points are extracted and each of the terms in the resultant vector are divided by the signal intensity of the first time point. The redundant first variable is removed and so the first feature $s(n)$ vector

consists of 4 variables and is defined as $f1 = \frac{s(n)}{s(0)}$ for $n=1, \dots, 4$. This feature was selected as it is the most straight forward approach to use.

2nd Feature: Derivative of Signal Intensities. This feature vector is comprised of the time based derivative of the relative signal intensity time curve (from feature #1) before the removal of the first redundant variable. This feature vector consists of 4

variables and is defined as $f2 = \frac{d}{dt} f1(n) = \frac{s(n+1) - s(n)}{t_{n+1} - t_n}$ for $n=0, \dots, 3$. This feature vector was selected as we hypothesize that this might reflect the actual physiological differences between malignant and benign lesions more closely. Cancerous lesions are fed by blood vessels that are characteristically leaky. It is thought that this leakiness translates into a higher rate of Gd-DTPA diffusing out of the vasculature into the lesion tissue. Since the concentration of Gd-DTPA is indirectly related to the observed signal intensity, it was thought that the rate of increase of signal intensity in the lesion (the derivative of the SI-time curve) might more closely reflect differences in permeability between the malignant and benign tumours.

3rd Feature: Relative Signal Intensities and their derivatives in one vector. It was thought that the second feature may provide useful separation information, however this approach ignores the information from the first feature vector. This feature thus combines the information from the first two vectors to form a longer feature vector that consists of 8 variables and is defined as $f3 = \{f1, f2\}$. Each of the variables in this feature vector were scaled from zero to one.

4th Feature: The fourth feature vector consists of three variables: Maximum signal intensity enhancement (as a percentage) from pre-contrast to any post-contrast image, time of maximum enhancement in seconds, maximum washout (as a percentage). Each of these 3 variables were scaled from zero to one. This feature was selected as

its variables are similar to the parameters used by commercially available approaches to the computer-aided diagnosis of breast MR images [14].

D. Support Vector Machine Based Classification

Support vector machines (SVMs) are an emerging area of research in machine learning and pattern recognition [25]. SVMs are a machine learning method for creating a classification function from a set of labeled training data. Support vector machines operate by locating a hypersurface that attempts to split the training data into two categories. The hypersurface is selected such that its distance to the nearest training data on either side of the surface is maximized. If no hypersurface is capable of linearly separating the data, a kernel transformation function is used to map the data into a different dimensional space (called a feature space) so that it can be linearly separated using standard SVM hypersurface techniques. Multiple types of kernels have been developed to map data into differing dimensions. For the purpose of this study, we have compared a linear kernel (equation 1), a polynomial kernel (equation 2) and a radial basis function kernel (equation 3):

$$K(x_i, x_j) = x_i \bullet x_j \quad (1)$$

$$K(x_i, x_j) = (\gamma(x_i \bullet x_j) + a)^d \quad (2)$$

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (3)$$

where x_i and x_j are input vectors comprised of one of the previously mentioned feature vectors, $\bullet$ is the dot product operation and γ , a and d are kernel parameters. Non-linear kernel functions (equations 2 & 3) provide flexibility to the SVM hyperplane calculation. These equations allow the definition of the classification function to be non-linear in the input space. If the kernel transformation function does not fully separate our data, a slack error variable is used to create a soft margin hyperplane for data separation. This error variable is calculated as the weighted sum of the misclassified training set data points. For receiver operating characteristic (ROC) curve analysis, the relative error penalty term for our benign class was kept at 1, and the error penalty term for our malignant class was varied from 1 to 7.

Classification of lesions has been performed as a 2-class problem where the 2 classes are cancer and benign. This classification problem has also been addressed in terms of 3-classes where the cancerous cases are separated into two classes (invasive and non-invasive cancers). For the 3-class problem, two separating hyperplane classification functions are created instead of just one. The first separates invasive cancers from the third class (benign lesions). The second hyperplane separates non-invasive cancers from benign lesions. If either of the two hyperplanes classify a given lesion as cancerous, that algorithm's prediction is considered to be cancer. Only if both hyperplanes classify the lesion as non-

cancerous do we consider the prediction to be benign. For ROC curve generation for the 3-class problem the error penalty term for the two cancerous classes (invasive and non-invasive) were varied in unison (thus they were always identical to each other). Support vector machine based classification has been implemented using the libsvm open source library [26].

E. Signal Enhancement Ratio Based Classification

In order to properly evaluate support vector machines as a classification mechanism for the delineation of malignant and benign lesions from DCE-MR breast images, we must compare its performance against a well established technique. We have elected to compare our approach with the commercially available signal enhancement ratio (SER) method [14]. The SER algorithm operates by calculating the signal enhancement ratio which is defined as $SER = (s1 - pre) / (s2 - pre)$, where pre is the pre-contrast signal intensity. We have set s1 and s2 to be the signal intensities of our first post-contrast and last post-contrast volumes respectively. The algorithm classifies a signal intensity time curve as cancer if its SER is greater than a given cutoff value (typically set to 1.1). ROC curve generation for the SER method was accomplished by varying the SER cutoff at which a voxel is diagnosed as cancerous from 0.1 to 3.0 in steps of 0.1.

F. Visual Evaluation

It was thought that being able to visualize a classification function with respect to our data set could be a beneficial means of evaluating different design options. Unfortunately, since our data is three to eight dimensional in nature, visualization can be challenging. Here we are presenting a new technique for visualizing a classification function that separates data of a high dimensional nature.

The first step of this technique is to project our high dimensional data into two dimensions. This can be accomplished by a number of data projection techniques but we have used principal components analysis. Principal components analysis is a mathematical technique for rotating data such that the resultant orthogonal axes are aligned to the maximum variance in the data set. This rotation is performed and the two principal components with the highest corresponding eigenvalues are selected (referred to as the first and second principal components). Each rotated input vector is plotted on a grid representing these first two principal components with highest corresponding eigenvalues (see figure 3 part A for an example plot).

The second step of the technique is to sample this principal component (PC) space across the range of first and second principal component values (those with the highest eigenvalues) that our data set occupies. Thus we sample from the minimum PC value in our data to the maximum PC value, however, a larger sampling area can be used if deemed necessary. In the case of the plots presented in this paper, principal component space was sampled 400 times along each of the first two principal components. Each of these sampled points is reverse rotated back to the input space and the resultant signal-intensity time vector is compared against any given classifier prediction technique. This generates a binary image (in our case 400×400) where the regions represent how the given classifier predicts classes in principal

component space. An example binary image is provided in figure 3 part B. This binary image is processed through an edge detector (we used the Sobel edge detector) and the resultant edge is plotted in the original principal component space (where our input data has been plotted). An example plot is provided in figure 3 part C. The resultant contour line defines the border between different predictions for a given classifier in principal component space. Since we have based our projections on principal components analysis this visualization technique summarizes the total sample variance for classification interpretation.

The percentage of total data variance displayed in the resultant two-dimensional plot can be calculated as the sum of the first two principal component eigenvalues (the two largest eigenvalues) divided by the sum of all of the eigenvalues. The classification function can also be evaluated in the lower principal components by extending this technique to a set of pair-wise plots of principal component space (plotting not just PC1 against PC2 but also PC2 against PC3, PC3 against PC4 and so on). It should be noted that any method of projecting high-dimensional data into a two-dimensional space inherently involves the loss of some information. It is impossible to guarantee that any given projection is the ideal two-dimensional plot for classification evaluation. As such we strongly recommend sampling points in the projection space and viewing these points in the input space, as we have demonstrated in figures 5 and 6.

To assist in plot interpretation, input space axes can also be projected into the principal component space. We have accomplished this by creating a fixed number (we used 100) of input space data points whereby all of the variables are set to zero but for the axis we are projecting. The values of the remaining variable for each input data point are scaled between 0 and a fixed percentage (we used 35%) of that variable's total variance. This percentage is a figure formatting parameter that controls the size of axes that are projected onto these two-dimensional plots. This process is repeated for each axis to be projected. The resultant input space data points are rotated into the principal component space. The resultant axes can then be shifted to an appropriate viewing location as seen in figures 4, 5 and 7. Since each axis has been scaled to the same fixed percentage of data variance, the relative length of each projected axis reflects that variable's relative influence on total data variance.

III. Results

A. Leave-One-Out Trials

Leave-one-out cross-validation consists of training the machine learning algorithm with a training set formed from all but one case of the total data set. A separating hyperplane is calculated based on this training set. The validation set consists of the single remaining case and is compared with the calculated hyperplane. This process is repeated such that for each trial a different case is removed from the total data set for validation. This technique was conducted 94 times, with support vector machines attempting to classify the remaining case after training on the other 93 test cases.

The results for a linear kernel applied to each feature vector for both the 2-class and 3-class problems for leave-one-out cross-validation are provided in table II. The areas under each

receiver operating characteristic curve are provided (under the heading Area). We have also provided the corresponding error ratio (cancer:benign under the heading ErR), sensitivity (sens), specificity (spec), positive predictive value (PPV), negative predictive value (NPV) and overall accuracy (OA) for the point on the ROC curve geometrically closest to 100% sensitivity, 100% specificity. The cancerous cutoff value for the signal enhancement ratio solution is also provided under the ErR heading. For comparative reasons we have provided the results of the SER method in each of the SVM kernel function tables (II–IV) below.

The results for a polynomial kernel are provided in table III. The kernel parameters γ , a and d were varied according to the following equation: $\{\gamma, a, d\} = e^n$, where $n \in \{-7.0, 4.5\}$ in steps of 0.5. The highest areas under the ROC curve found for a polynomial kernel are provided. The table was populated in the same manner as Table II.

The results for a radial basis function kernel are provided in table IV. The kernel parameter γ was varied according to the following equation: $\gamma = e^n$, where $n \in \{-7.0, 4.5\}$ in steps of 0.1. The highest areas under the ROC curve found for a radial basis function kernel are provided. The table was populated in the same manner as Tables II and III.

The highest area under the ROC curve was obtained when using a radial basis function kernel with γ set to 4.06 on the fourth feature vector for the 2-class problem. The area under the curve was 0.74 (corresponding entry highlighted in table IV).

B. Pixel-by-Pixel Randomized Trials

Although leave-one-out validation is a useful tool for evaluating a given approach, we are also interested in the robustness of the algorithms. It is also beneficial to evaluate these approaches in a more clinically viable manner. For the purposes of this study we are not differentiating between benign lesions and normal tissue (both are simply considered non-cancer). All support vector machine results were obtained using the third feature vector (relative signal intensities and their derivatives) and the fourth feature vector (Maximum signal intensity enhancement, time to maximum enhancement, maximum washout). The first and second feature vectors were not considered in this section as they consistently underperformed the third and fourth feature vectors in the previous leave-one-out trials.

We have conducted a series of 100 random trials whereby 75% of the cases (94 in all) from each class were randomly selected and used for training (2544 vectors in all). The calculated hyperplane was then applied to the remaining examinations (validation set) on a pixel-by-pixel basis across the whole parenchyma tissue for one slice as identified by visual inspection. If at least one pixel within a radiologically identified cancerous lesion was diagnosed as cancer the case was counted as a true positive. If any pixel in a radiologically identified benign lesion or normal breast parenchyma was diagnosed as cancer, the case was counted as a false positive. This enumeration scheme was employed as we know the histopathologist's diagnosis of each radiologically identified lesion, but we have no official gold standard diagnosis for each voxel. Noise removal was performed by 5×5 neighbourhood median filtering. The results for the signal enhancement ratio (SER) method are included as well and have been subjected to the same noise filtering and test evaluation metrics as the support vector machine approaches. All classification methods were subjected

to a 50% enhancement threshold, whereby voxels whose maximum enhancement from the pre-contrast volume to any post-contrast volume is less than the given threshold are classified as non-cancerous. This threshold was selected as it is in-line with thresholds quoted in the literature [8, 14] and was determined to be below the minimum enhancement of all cancerous lesions in our data set (where the calculated enhancement is averaged across all the hyperintense voxels in the given lesion).

As was the case in the leave-one-out trials, ROC curve generation was accomplished for both methods (SVM and SER) by the same methods outlined in the sections II.D and II.E. The resultant test parameter means and standard deviations (in brackets) for the test case on the ROC curve whose sensitivity and specificity values are geometrically closest to 100%, 100% are provided in Table V for features 3 and 4. For each SVM test case, any relevant kernel parameter and error ratio settings are provided. In an effort to thoroughly evaluate the classification functions, we have also tracked the number of false positive pixels per cancerous examination (FPP) and the number of true positive pixels per cancerous examination (TPP). Since the SER method does not require training, it was applied on a pixel-by-pixel basis on all 94 examinations.

C. Visual Evaluation

We implemented the plotting technique described in section II.F on the radial basis function kernel classifier that maximized the area under the ROC curve (feature vector 4, $\gamma=4.06$, ROC Area=0.74, see bolded entry in table IV) for the leave-one-out trials (94 vectors). We fixed the error ratio at 3:1 (cancer:benign). The resultant plot is provided in figure 4.

We also implemented this plotting technique on two of our radial basis function classifiers from the randomized trials (2544 vectors) in figure 5 for feature vector 3. Contour lines are provided for SVM solutions with error penalty ratios of both 3:1 and 4:1. The classifier contour line for the SER approach is provided along with the contour line representative of the 50% threshold described in III.B. The test parameters corresponding to the two SVM classifiers and the SER classifier are bolded in Table V. In order to assist in the interpretation of figure 5, we have sampled four points in principal component space (seen on figure 5 as black diamonds at locations (1,0.7), (0.7,0), (0,0.4), and (0,0)) and the corresponding signal intensity time curves are provided in figure 6.

We were interested in comparing a scatter plot showing both invasive and non-invasive cancers to help shed light on the results of our 3-class problem experiments and provided the plot in figure 7.

Finally, we have demonstrated how the results of support vector machine classification can be visualized for radiological analysis in figure 8. Here we have colour coded any pixel that fell on the malignant side of the support vector machine hyperplane in red. The automated support vector machine based diagnosis of a single 256×256 slice takes about 3 seconds on a 3 GHz Intel Pentium 4 workstation.

IV. Discussion

We have evaluated support vector machines as a potential mechanism for the design of a classifier responsible for delineating between malignant and benign breast lesions from DCE-MRI time-series data. There were a number of motivations for selecting SVMs as a classification mechanism. SVMs have been shown to perform well in medical diagnosis applications [16], and have also been shown to perform well when dealing with relatively small training sets [17]. This was particularly appealing given the inherent difficulty in acquiring large amounts of screening data devoted exclusively to training. Support vector machines also perform well and classify reasonably quickly on high dimensional data. We evaluated SVMs by comparing their performance with the signal enhancement ratio (SER) based breast MR computer-aided diagnosis method. We selected the SER method as it is both commercially available (Confirma Inc., Kirkland WA USA) and currently in use by radiologists.

We have provided a visual plot of the classification function with the highest ROC area for leave-one-out trials in figure 4. For feature vector 4, our data is divided into four clusters that reflect the fourth feature's second variable (the time in seconds to peak enhancement). Since there are only four post-contrast volumes acquired there are only four possible values for this time to enhancement term. We can see from the plot in figure 4 that the classification function is simply dividing the data based on its time to enhancement. This was easily visualized with the provided axis projections and demonstrates the benefit of the proposed classifier visualization technique as a method to effectively interpret the behaviour of a classification function.

Although we have provided results for the cases with the highest receiver operating characteristic (ROC) curve areas, support vector machines produced many computer-aided diagnostic design alternatives (examples provided in Table V). It should be noted that the standard deviations in Table V are relatively high. It is believed that this is due to insufficient tuning of the error penalty term to suit the training data for a given trial. High standard deviations could also be attributed to the inseparability of our data due to both the patient population and the low temporal resolution of our MR acquisition protocol. Since our data is not normally distributed we have elected to test for statistical significance using the Wilcoxon-Signed Rank Test. We compared the ROC area distributions for feature 4 (polynomial kernel) and feature 3 (RBF kernel) and the calculated probability that the observed differences (bolded entries in Table V) occurred randomly was 0.0012. This indicates that the observed differences between feature 3 and feature 4 are significant. We also ran the same statistical significance test comparing feature 3 (polynomial kernel) with feature 3 (RBF kernel) and found that the probability that the observed differences occurred randomly was 0.6762, indicating that these two are not statistically significant.

Figure 5 provides a plot of the data vectors used for training in the pixel-by-pixel randomized trials in principal component space. It can be seen from figure 5 and the corresponding entry in table V (bolded with error ratio 3:1) that the given radial basis function solution achieves a high specificity while maintaining reasonable sensitivity values given the inseparability of this data set. High sensitivities were difficult to obtain without

severely compromising the test's specificity; this may be due to the presence of cancerous time-curves which do not show any evidence of a wash-out phase.

The signal enhancement ratio method is based on a linear equation, consequently the classifying border in principal component space is also linear. It should be noted that the observed nonlinear regions of the signal enhancement ratio boundary function (see figure 5) are in fact caused by the 50% threshold (border provided in black) we applied to all classifiers as mentioned in section III.B. Future research will experiment with varying this threshold over a wide range of values to determine its effect on the operation of classification based computer-aided diagnostic systems. It can be seen in figure 5 that the 50% threshold border divides the inseparable data enforcing a non-cancerous diagnosis of a few cancerous vectors. It is thought that despite efforts to draw the regions of interest around the most enhancing part of each lesion there was still some variance within some of our cancerous tumours.

Although feature 3 (demonstrated in principal component space in figure 5) resulted in the best performance (see Table V), it is compromised by being locked to the MR acquisition protocol. In theory feature 4 should be able to function on any acquisition protocol, and so although it performs somewhat worse than feature 3, it remains an interesting design option. Future work will involve evaluating feature 4 on a variety of MR acquisition protocols.

These examinations were also analyzed by a trained radiologist blinded to the findings of any of the computer-aided diagnosis techniques. The radiologist diagnosed each examination on the BI-RADS scale [27]. The examinations considered in this study were also deemed difficult to diagnose by the radiologist. Of the total 94 examinations, 76 were rated BI-RADS 0 (Needs further work-up). Of the remaining 18 cases there were 5 true positives, 5 false positives, 8 true negatives and 0 false negatives when we consider BI-RADS 4 and 5 as a cancerous diagnosis and BI-RADS 1–3 as benign.

We experimented with dividing our group of cancers into two classes (invasive and non-invasive cancer) and referred to these experiments as a 3-class problem (the third class is benign). This division was made as invasive cancer is characterized by having spread into neighbouring tissues and through angiogenesis, the tumours are supplied with characteristically leaky blood vessels. This increased permeability increases the tumour tissue's uptake and washout rates of our Gadolinium based (Gd-DTPA) contrast agent. Thus we might expect to see different signal-intensity time curves for invasive cancer as compared to non-invasive cases which have not spread into neighbouring tissues. In practice the 3-class problem consistently underperformed the 2-class approach for all kernel functions and as such we elected to forgo the 3-class problem for the randomized trials presented in section III.B. Ultimately the efficacy of solving the problem with two independent classification functions will depend on how distinct the two classes (invasive and non-invasive) are. As can be seen in figure 7, although there is some separation between invasive and non-invasive cancers, the two groups overlap each other significantly. This indicates that perhaps creating two classification functions (one for each invasive and non-invasive cancers) will not assist in the delineation between malignant and benign tissues.

When considering feature vectors, the first (relative signal intensities) was selected as it is the most obvious choice. As mentioned in section II.C features 2 and 3 were selected with biological motivations (the relationship between the derivative of the signal intensity time curve and the permeability of the vasculature in cancerous tumours). It was the third feature vector (both relative signal intensities and their derivatives) that performed best in the more clinically realistic pixel-by-pixel trials. The additional information in the third feature vector (relative SI and their derivative) is present in the original data set (feature #1 – relative signal intensities) in the form of the relationship between adjacent data points. However, it is believed that including these derivatives explicitly in the feature vector facilitates the exploitation of these differences by support vector machines.

Ideally, we would like to convert the signal intensity time curves into concentration time curves since the concentration does not depend on the MR imaging sequence used. This, however, requires a pre-contrast map of T1 values which is not currently available. Our current classification algorithm uses the relative signal intensity values from the screening MR images as 3 to 8 dimensional “feature vectors”. If we were to apply feature vectors 1–3 with a higher temporal resolution, the size of the feature vector would increase, however it is known that the stability of a classification algorithm generally decreases as the number of dimensions increases [28], so caution is warranted. One way of overcoming this problem is to parameterize the signal intensity time curves using pharmacokinetic modeling [29]. Ideally, any future computer aided diagnostic system will have the flexibility to accommodate a variety of magnetic resonance imaging protocols without compromising the separability of the data. In addition to the dynamic information, spatial patterns have also been shown to provide considerable discrimination between malignant and benign breast MR lesions [18]. Future work will examine techniques to extract spatial features in addition to the dynamic information already considered in this study. Additionally, in order to form firm conclusions about the efficacy of the discussed methods, validation will need to be performed on independent data sets.

V. Conclusion

In this paper we studied the use of support vector machines as a classification mechanism for delineating malignant and benign lesions from dynamic contrast-enhanced magnetic resonance images (DCE-MRI) of the breast. We demonstrated the support vector machine approach as offering significant flexibility in the design of a computer aided-diagnostic system for DCE-MRI of the breast. We have also demonstrated the efficacy of a new classification visualization approach to assist in the evaluation of design alternatives. More research is needed to fully implement the support vector machine technique in a clinically acceptable manner that is not constrained with respect to the magnetic resonance acquisition protocol. This needs to be followed by extensive clinical validation against known pathology across a broad range of disease entities to draw firm conclusions about the significance of this technique. However, results from this study suggest that support vector machines are a useful method for computer aided diagnosis of DCE-MRI of the breast with application to hereditary breast cancer surveillance.

Acknowledgments

The authors would like to thank Elizabeth Ramsay for her assistance in image acquisition. The authors would also like to thank Mike Froh for his assistance in image registration.

This work was supported in part by the Canadian Breast Cancer Research Alliance and the Canadian Institute of Health Research.

References

1. Ford S, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *American Journal of Human Genetics*. 1998; 62:676–689. [PubMed: 9497246]
2. Curry, SJ. Fulfilling the potential of cancer prevention and early detection. Washington, DC: National Academies Press; 2003.
3. Warner E, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *Journal of the American Medical Association*. 2004; 292:1317–1325. [PubMed: 15367553]
4. Lucht R, Delorme S, Brix G. Neural network-based segmentation of dynamic MR mammographic images. *Magnetic Resonance Imaging*. 2002; 20:147–154. [PubMed: 12034335]
5. Degenhard A, et al. Comparison between radiological and artificial neural network diagnosis in clinical screening. *Physiological Measurement*. 2002; 23:727–739. [PubMed: 12450272]
6. Lucht RE, Knopp MV, Brix G. Classification of signal-time curves from dynamic MR mammography by neural networks. *Magnetic Resonance Imaging*. 2001; 19:51–57. [PubMed: 11295347]
7. Fischer H, Hennig J. Neural network-based analysis of MR time series. *Magnetic Resonance in Medicine*. 1999; 41:124–131. [PubMed: 10025619]
8. Tzacheva AA, Najarian K, Brockway JP. Breast cancer detection in gadolinium-enhanced MR images by static region descriptors and neural networks. *Journal of Magnetic Resonance Imaging*. 2003; 17:337–342. [PubMed: 12594724]
9. Szabo BK, Aspelin P, Wiberg MK. Neural network approach to the segmentation and classification of dynamic magnetic resonance images of the breast: comparison with empiric and quantitative kinetic parameters. *Academic Radiology*. 2004; 11:1344–1354. [PubMed: 15596372]
10. Vomweg TW, et al. Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Medical Physics*. 2003; 30(9):2350–2359. [PubMed: 14528957]
11. Jacobs MA, et al. Benign and malignant breast lesions: diagnosis with multiparametric MR imaging. *Radiology*. 2003; 229:225–232. [PubMed: 14519877]
12. Gilhuijs KGA, Giger ML. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics*. 1998; 25:1647–1654. [PubMed: 9775369]
13. Chen W, Giger ML, Lan L, Bick U. Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics. *Medical Physics*. 2004; 31(5):1076–1082. [PubMed: 15191295]
14. Niemeyer, T., Wood, C., Stegbauer, K., Smith, J. Comparison of automatic time curve selection methods for breast MR CAD. *Proceedings Medical Imaging 2004: Image Processing*; San Diego, CA, USA. February 2004; p. 785-790.
15. Partridge SC, et al. Accuracy of MR imaging for revealing residual breast cancer in patients who have undergone neoadjuvant chemotherapy. *American Journal of Roentgenology*. 179:1193–1199.
16. Chang, Ruey-Feng, Wu, Wen-Jie, Moon, Woo Kyung, Chou, Yi-Hong, Chen, Dar-Ren. Support vector machine for diagnosis of breast tumors on US images. *Academic Radiology*. 2003; 10:189–197. [PubMed: 12583571]
17. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A Support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*. 2002; 21:1552–1563. [PubMed: 12588039]

18. Nattkemper TW, et al. Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artificial Intelligence in Medicine*. 2005; 34:129–139. [PubMed: 15894177]
19. Levman, J., Plewes, DB., Martel, AL. Validation of SVM based classification of DCE-MRI breast lesions. *Proceedings MICCAI 2006 Workshop on Medical Image Processing: The Challenges in Clinical Oncology*; Copenhagen, DK. October 2006;
20. Nattkemper TW, Wismuller A. Tumor feature visualization with unsupervised learning. *Medical Image Analysis*. 2005; 9:344–351. [PubMed: 15907392]
21. Komura D, Nakamura H, Tsutsumi S, Aburatani H, Ihara S. Multidimensional support vector machines for visualization of gene expression data. *Bioinformatics*. 2005; 21(4):439–444. [PubMed: 15608050]
22. Somorjai, R., Dolenko, B. Class-preserving mapping of high-dimensional biomedical data: visualization, classification, clustering. *Proceedings Interface and Classification Society of North America*; St. Louis, MO, USA. June 2005;
23. Greenman RL, Lenkinski RE, Schnall MD. Bilateral imaging using separate interleaved 3D volumes and dynamically switched multiple receive coil arrays. *Magn Reson Med*. 1998; 39:108–115. [PubMed: 9438444]
24. Froh, MS., Barber, DC., Brock, KK., Plewes, DB., Martel, AL. Piecewise-quadrilateral registration by optical flow – applications in contrast-enhanced MRI of the breast. *Proceedings MICCAI 2006*; Copenhagen, DK. October 2006; p. 686-693.
25. Vapnik, VN. *The Nature of Statistical Learning Theory*. New York, NY: Springer; 2000.
26. Chang C-C, Lin C-J. Accessed Sept. 10 2006 LIBSVM – A library for support vector machines [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
27. Morris EA. Illustrated breast MR lexicon. *Semin Roentgenol*. 2001; 36:108–115. [PubMed: 11329653]
28. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1991; 13(3): 252–264.
29. Srikanchana, R., Thomasson, D., Choyke, P., Dwyer, A. A comparison of pharmacokinetic models of dynamic contrast enhanced MRI. *Proceedings IEEE Symposium on Computer-Based Medical Systems*; Bethesda, MD, USA. 2004.

Fig. 1.
Subtraction images of an invasive ductal carcinoma (left) and a fibroadenoma (right).

Fig. 2.
A malignant (solid) and benign (dashed) signal intensity time curve.

Fig. 3.

An example plot of two-class data in principal component space (A), the resultant binary image (B) and the final high dimensional classifier visualized plot in principal component space (C).

Fig. 4.

Contour line (blue) of a radial basis function kernel classifier ($\gamma = 4.06$, error ratio 3:1) in principal component space (94% of total data variance displayed from feature vector 4).

Fig. 5.

Contour lines (blue) of a radial basis function kernel classifier ($\gamma = 0.22$, error ratio 3:1 – dashed line, error ratio 4:1 – dotted line) and the SER based classifier (cancer cut off 1.1, solid magenta line) in principal components space (88% of data variance displayed from feature vector 3).

Fig. 6.

Signal intensity time curves of points sampled in principal component space (black diamonds from figure 6).

Fig. 7.

Invasive and non-invasive cancers plotted in principal component space (91% of total data variance displayed from feature vector 3).

Fig. 8. A ductal carcinoma in situ (left) correctly diagnosed as cancer by a radial basis function ($\gamma = 0.22$, error ratio 3:1) support vector machine based classifier (right).

TABLE I

Quantity and Pathological Diagnosis of Breast Lesions

Cancer/Benign	Pathological Diagnosis	Quantity
Cancer	Invasive Ductal Carcinoma	8
Cancer	Ductal Carcinoma in Situ (DCIS)	12
Cancer	DCIS/Microinvasion	3
Cancer	Invasive Lobular Carcinoma	1
Benign	Fibroadenoma	4
Benign	Fibrocystic Disease	3
Benign	Normal Parenchyma	8
Benign	Papilloma	2
Benign	Stromal Fibrosis	2
Benign	Ductal Hyperplasia	3
Benign	Focal Fibrosis	2
Benign	Fibrocystic Changes	3
Benign	Benign by Assumption	43

TABLE II

Results of Leave-One-Out Trials – Linear Kernel

Class	Test		Area	Sens. %	Spec. %	PPV %	NPV %	OA %
	Feature	ErR						
2	1	4:1	0.59	66.7	58.6	35.6	83.7	60.6
3	1	4:1	0.48	37.5	90.0	56.3	80.8	76.6
2	2	3:1	0.60	58.3	64.3	35.9	81.8	62.8
3	2	6:1	0.58	79.2	37.1	30.2	83.3	47.3
2	3	4:1	0.69	75.0	52.9	35.3	86.0	58.5
3	3	5:1	0.60	45.8	75.7	39.3	80.3	68.1
2	4	4:1	0.68	62.5	71.4	42.9	84.7	69.1
3	4	5:1	0.65	50.0	78.6	44.4	82.1	71.3
SER	-	0.9	0.69	62.5	74.3	45.5	85.2	71.3

TABLE III

Results of Leave-One-Out Trials – Polynomial Kernel

Class	Feature	Test γ	a	d	ErR	Area	Sens. %	Spec. %	PPV %	NPV %	OA %
2	1	2.72	2.72	1.65	4:1	0.66	75.0	55.7	36.7	86.7	60.6
3	1	0.05	0.37	2.72	4:1	0.65	33.3	97.1	80.0	81.0	80.9
2	2	0.37	2.72	2.72	3:1	0.65	66.7	65.7	40.0	85.2	66.0
3	2	1.00	2.72	2.72	5:1	0.67	70.8	64.3	40.5	86.5	66.0
2	3	2.72	1.65	2.72	3:1	0.70	66.7	65.7	40.0	85.2	66.0
3	3	0.37	0.001	1.00	5:1	0.66	41.7	90.0	58.8	81.8	77.7
2	4	0.22	0.14	2.72	3:1	0.71	62.5	78.6	50.0	85.9	74.5
3	4	0.03	2.72	2.72	5:1	0.71	62.5	78.6	50.0	85.9	74.5
SER	-	-	-	-	0.9	0.69	62.5	74.3	45.5	85.2	71.3

TABLE IV

Results of Leave-One-Out Trials – Radial Basis Function Kernel

Class	Test Feature	γ	ErR	Area	Sens. %	Spec. %	PPV %	NPV %	OA %
2	1	0.30	4:1	0.63	75.0	37.1	29.0	81.3	46.8
3	1	0.07	5:1	0.63	33.3	91.4	57.1	80.0	76.6
2	2	0.15	3:1	0.64	45.8	81.4	45.8	81.4	72.3
3	2	4.06	5:1	0.60	54.2	64.3	34.2	80.4	61.7
2	3	2.46	3:1	0.63	62.5	71.4	42.9	84.7	69.1
3	3	0.20	5:1	0.64	37.5	88.6	52.9	80.5	75.5
2	4	4.06	3:1	0.74	62.5	78.6	50.0	85.9	74.5
3	4	0.09	5:1	0.71	62.5	78.6	50.0	85.9	74.5
SER	-	-	0.9	0.69	62.5	74.3	45.5	85.2	71.3

TABLE V

Results of Pixel-by-Pixel Randomized Trials

Feature	Test		γ, a, d	ErR	Area	Sens. %	Spec. %	PPV %	NPV %	OA %	FPP	TPP
	Kernel											
3	Linear	-	-	4:1	0.54 (0.13)	55.5 (21.7)	48.3 (24.1)	27.2 (12.9)	73.6 (14.8)	50.1 (15.8)	16.3 (35.5)	29.6 (21.9)
4	Linear	-	-	3:1	0.40 (0.09)	52.8 (32.4)	35.8 (35.9)	16.8 (9.8)	60.7 (24.6)	40.1 (20.1)	35.8 (41.2)	26.2 (24.3)
3	Poly	0.37,0.61,2.72	-	4:1	0.62 (0.11)	44.5 (22.6)	73.4 (25.0)	43.1 (22.6)	76.8 (15.4)	66.2 (16.2)	17.7 (59.7)	26.9 (27.5)
4	Poly	2.72,1.0,2.72	-	2:1	0.59 (0.11)	41.7 (19.6)	76.3 (18.2)	41.6 (21.3)	79.6 (5.5)	67.7 (12.1)	5.6 (10.0)	25.6 (24.1)
3	RBF	0.22	-	3:1	0.62 (0.11)	39.8 (18.3)	85.1 (10.8)	48.9 (23.2)	81.1 (4.7)	73.7 (8.2)	5.1 (10.9)	20.5 (20.5)
3	RBF	0.22	-	4:1	0.62 (0.11)	55.0 (21.8)	60.1 (28.0)	36.6 (17.3)	76.2 (16.6)	58.8 (18.8)	18.4 (44.7)	35 (30.1)
4	RBF	1.65	-	2:1	0.57 (0.11)	42.2 (19.7)	73.2 (19.5)	39.8 (22.4)	78.7 (7.3)	65.5 (12.8)	6.3 (11.2)	25.3 (22.9)
SER	-	-	-	1:1	0.49	58.3	50.0	28.6	77.8	52.1	34.5	20.5