Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses

Lubomir Hadjiiski,[a] Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, and Metin Gurcan
Department of Radiology, The University of Michigan, Ann Arbor, Michigan 48109-0904

(Received 22 May 2001; accepted for publication 27 August 2001)

A new classification scheme was developed to classify mammographic masses as malignant and benign by using interval change information. The masses on both the current and the prior mammograms were automatically segmented using an active contour method. From each mass, 20 run length statistics (RLS) texture features, 3 speculation features, and 12 morphological features were extracted. Additionally, 20 difference RLS features were obtained by subtracting the prior RLS features from the corresponding current RLS features. The feature space consisted of the current RLS features, the difference RLS features, the current and prior speculation features, and the current and prior mass sizes. Stepwise feature selection and linear discriminant analysis classification were used to select and merge the most useful features. A leave-one-case-out resampling scheme was used to train and test the classifier using 140 temporal image pairs (85 malignant, 55 benign) obtained from 57 biopsy-proven masses (33 malignant, 24 benign) in 56 patients. An average of 10 features were selected from the 56 training subsets: 4 difference RLS features, 4 RLS features, and 1 speculation feature from the current image, and 1 speculation feature from the prior, were most often chosen. The classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. For comparison, a classifier was trained and tested using features extracted from the 120 current single images. This classifier achieved an average training A_z of 0.90 and a test A_z of 0.82. The information on the prior image significantly ($p=0.015$) improved the accuracy for classification of the masses. [DOI: 10.1118/1.1412242]

Key words: computer-aided diagnosis, interval change, classification, feature analysis, mammography, malignancy

I. INTRODUCTION

Mammography is currently the most effective method for early breast cancer detection.[1,2] Analysis of interval changes is an important method used by radiologists in mammographic interpretation to detect developing malignancy.[3,4] A variety of computer-aided diagnosis (CAD) techniques have been developed to detect abnormalities and to distinguish malignant and benign lesions on mammograms. We are studying the use of CAD techniques to assist radiologists in interval change analysis.

Commonly used lesion classification methods for CAD employ information from a single image. These methods have been shown to perform well in lesion classification problems.[5–12] However, when mammograms from multiple examinations are available, it can be expected that even higher accuracy may be achieved if the computer can utilize the interval change information for classification. New computer vision methods will have to be designed to extract features characterizing temporal changes and to improve the differentiation between benign and malignant masses.

A number of researchers have developed algorithms to register the mass on current and prior mammograms. Sallam *et al.*[13] have proposed a warping technique for mammogram registration based on manually identified control points. A mapping function was calculated for matching each point on the current mammogram to a point on the prior mammogram. Brzakovic *et al.*[14] have investigated a three-step method for comparison of the most recent and the prior mammograms. They first registered two mammograms using the method of principal axis, and partitioned the current mammogram using a hierarchical region-growing technique. Translation, rotation, and scaling were then used for registration of the partitioned regions. Vujovic *et al.*[15] have proposed a multiple-control-point technique for mammogram registration. They first determined several control points independently on the current and prior mammograms based on the intersection points of prominent anatomical structures in the breast. A correspondence between these control points was established based on a search in a local neighborhood around the control point of interest.

The previous techniques depend on the identification of control points. Furthermore, these studies aimed at registration without using the results for interval change analysis.

Gopal *et al.*[16,17] and Hadjiiski *et al.*[18–20] have developed a multistage technique that defines a transformation to locally map the position of the mass on a current mammogram to a search region on the prior mammogram. A local search for the exact mass location is then performed on the prior mammogram. Good *et al.*[21] have developed a technique that defines a transformation to map all points from the current mammogram onto a prior mammogram. The current mammogram is then subtracted from the prior mammogram.

Few studies have been performed so far in the area of automated classification of breast masses based on the interval change information. Gopal *et al.*[22] and Hadjiiski *et al.*[23,24] have carried out a preliminary study of the classification scheme that combines prior and current information automatically extracted from masses on prior and current mammograms, respectively. The classifier using the combined prior and current information performed better than the classifier using current information alone. To our knowledge, no other studies that describe automated classification of malignant and benign breast lesions based on temporal changes of mammographic features have been reported.

The goal of our research is to develop a CAD method for automated analysis of interval changes to be used as an aid to radiologists for detection and classification of malignant and benign lesions on mammograms. In this study, we conducted a preliminary investigation to demonstrate the feasibility of analyzing temporal differences in the texture and morphological features between a mass on the most recent mammogram and a prior mammogram of the same view for the classification task. Additionally, we compared this method with two classification methods, one of which is based on information extracted from the current mammograms alone, the other one is based on information extracted from the prior mammograms alone.

II. MATERIALS AND METHODS

The new classification technique is based on the design of features that characterize the temporal change in the lesion of interest between two mammographic examinations. The mass to be analyzed can either be identified manually by a radiologist or automatically by a computerized detection program. In this study, the mass on each mammogram was identified by an MQSA certified radiologist. The masses on both the current and the prior mammograms were automatically segmented using an active contour method that has been discussed in detail elsewhere.[25,26] Examples of the segmentation are shown in Figs. 2 and 3 for a malignant and a benign mass, respectively. Features that characterized mammographic masses including texture features, morphological features, and spiculation features were extracted from each mass. Three of the morphological features are related to the mass size. Additionally, difference features were obtained by subtracting a feature of the prior mass from the corresponding feature of the current mass. The current, prior, and difference features formed a multidimensional feature space for the classification task. Stepwise feature selection applied to linear discriminant analysis (LDA) was used to select the most useful features. The selected features were then used as the input predictor variables for the LDA classifier (Fig. 1). The classifier was trained and tested by a leave-one-case-out resampling scheme. A case was considered to contain all regions of interest from a given patient. In each resampling step, the temporal pairs from 55 cases were used for feature selection and formulation of the linear discriminant function, while the temporal pairs from the left-out case were used for testing the trained classifier. A total of 56 training and testing steps were obtained from the 56 cases. The classification results from the 56 test cases were accumulated to evaluate the classifier performance. Since the data set in this study was still small, we chose the feature selection parameters such that the dimensionality of the input feature vector for the LDA classifier was small in order to reduce the possibility of over-training. The feature selection procedure is discussed in Sec. II C.

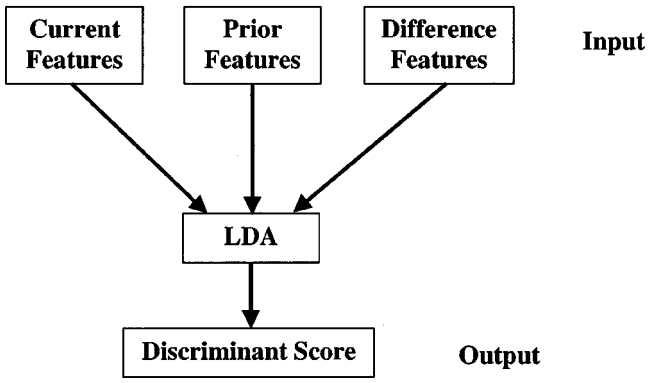

FIG. 1. Block diagram of the classification method.

To evaluate the improvement in the classifier performance designed by using the temporal change information, two additional classifiers were obtained. One of them was trained using the information extracted from the current single images of the temporal pairs. We will refer to these images as current images. The other classifier was trained using the information extracted from the prior single images of the temporal pairs and we will refer to these images as prior images. Comparison of the three classifiers will reveal the effectiveness of interval change analysis for the classification of malignant and benign masses.

A. Data set

A set of 140 temporal pairs of mammograms containing biopsy-proven masses on the current mammograms was used to examine the performance of this approach. The data set consisted of 241 mammograms from 56 patients. The mammograms were digitized with a LUMISCAN 85 laser scanner at a pixel resolution of $50\,\mu m \times 50\,\mu m$ and 4096 gray levels. The digitizer was calibrated so that gray level values were linearly proportional to the optical density (OD) within the range of 0–4 OD units, with a slope of 0.001 OD/pixel value. The digitizer output was linearly converted so that a large pixel value corresponded to a low optical density. The image matrix size was reduced by averaging every 2×2 adjacent pixels and downsampled by a factor of 2, resulting in images with a pixel size of $100\,\mu m \times 100\,\mu m$ for further analysis.

There were 57 biopsy-proven masses (33 malignant and 24 benign) in the 56 cases. The 241 mammograms contained different mammographic views (CC, MLO, and lateral views) and multiple examinations of the masses including the examination when the biopsy decision was made. By matching masses of the same view from two different examinations, a total of 140 temporal pairs were formed, of which

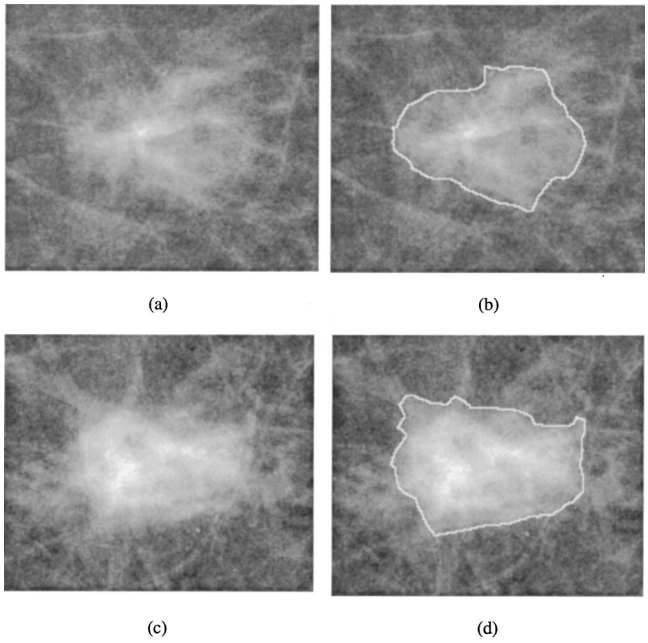

FIG. 2. A malignant mass: (a) the mass in a prior year mammogram (1997), (b) mass outline obtained by active contour segmentation, (c) the mass in a current year mammogram (1998), (d) mass outline obtained by active contour segmentation.

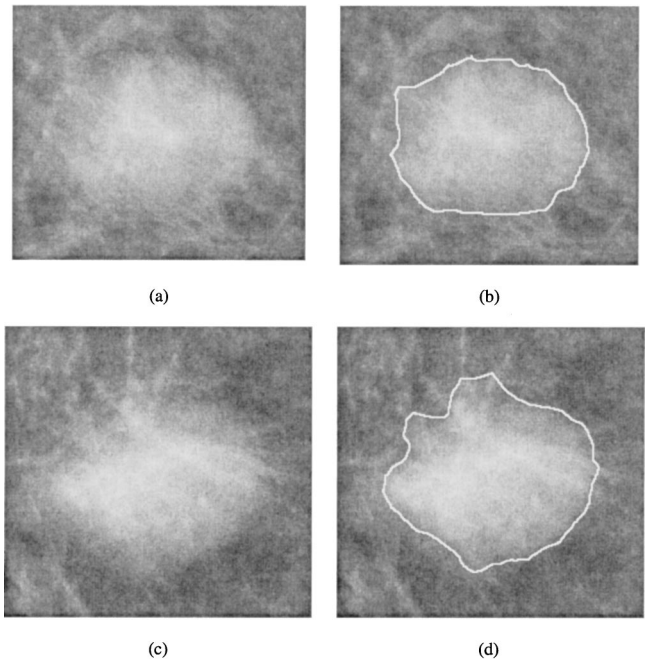

FIG. 3. A benign mass: (a) the mass on a prior year mammogram (1995), (b) mass outline obtained by active contour segmentation, (c) the mass on a current year mammogram (1996), (d) mass outline obtained by active contour segmentation.

85 were malignant and 55 benign. A malignant temporal pair consisted of a biopsy-proven malignant mass or a mass that was initially not recommended for biopsy and later found to be malignant by biopsy in a future year. A similar definition was used for the benign temporal pairs. Within a pair, the current mammogram was defined as the mammogram with the later date, and the prior mammogram was defined as the one with the earlier date. Therefore, in cases with three consecutive exams, more than one temporal pair could be formed and two of the mammograms could be called ''current.'' Among the 140 temporal pairs, we had 120 unique current mammograms. Of the masses in the 120 current mammograms, 70 were malignant and 50 benign.

Since all cases in this data set had undergone biopsy, the benign masses in this set could not be distinguished easily from the malignant ones based on current mammographic criteria. Changes occurred for the benign masses that prompted the radiologists to recommend biopsy. Examples of such cases are shown in Figs. 2 and 3. The malignant mass in Fig. 2 did not increase in size but changed its density. The benign mass (Fig. 3), on the other hand, appeared to have spicules. For the malignant masses in this data set, the average mass size, estimated by the radiologist as the longest dimension of the mass on the mammogram, was 8.2 mm on the prior mammograms and 12.7 mm on the current mammograms. The corresponding sizes were 10.6 and 12.2 mm, respectively, for the benign masses. As discussed in Sec. IV, 25 of the masses on the prior mammograms were too subtle for the radiologist to estimate their sizes. The average sizes given previously were obtained after excluding all temporal pairs that involved these masses.

The radiologist also rated the visibility of the masses on the mammograms relative to those encountered in clinical practice on a 10-point scale, with 1 representing the most obvious and 10 representing the most subtle masses. The visibility of the masses on the current mammogram is plotted against those on the prior mammogram in Fig. 4 for the malignant and benign temporal pairs. Generally the malignant masses were less visible on the prior than on the current mammograms while the visibility of the benign masses was found to be more similar on the current and prior mammograms. The mean difference in the visibility rating between the prior and the current mammograms for the malignant masses is 2.8 compared to 1.2 for the benign masses ($p = 0.0007$ with an unpaired t-test between the malignant and benign masses). The correlation coefficient is 0.02 for malignant masses [Fig. 4(a)] and 0.37 for benign masses [Fig. 4(b)]. In addition, the radiologist also estimated the likelihood of malignancy of the current masses on a 10-point confidence scale (1—definitely benign and 10—definitely malignant) based on the 120 current mammograms alone without comparison with the prior (Fig. 5). The temporal pairs had a time interval of 6–36 months (Fig. 6). More than 70% of the pairs had a time interval of 12 months.

B. Feature extraction

A rectangular region of interest (ROI) was defined to include the radiologist-identified mass with an additional surrounding breast tissue region of at least 40 pixels wide from any point of the mass border. A fully automated method was then used for segmentation of the mass from the breast tissue background within the ROI. The masses on both the current and the prior mammograms were automatically segmented

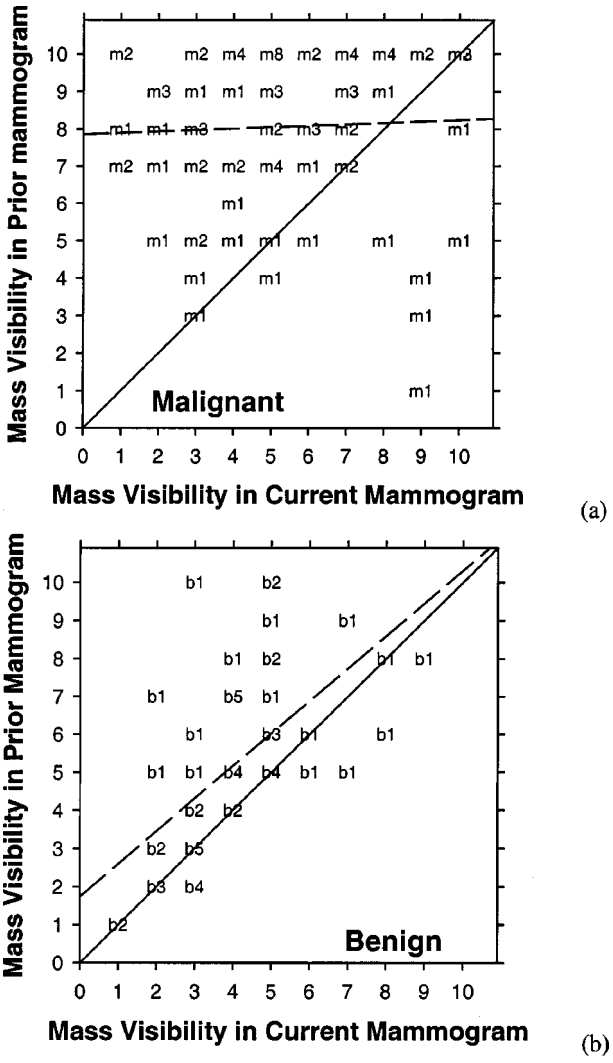

FIG. 4. Visibility of the masses on the current mammogram plotted against those on the prior mammogram for (a) malignant and (b) benign temporal pairs. The visibility was rated on a 10-point discrete scale (1 = most obvious, 10 = most subtle). Because many of the data points overlap, we indicate the number of points with the same rating by a number next to the symbol (m or b). The diagonal line on the graph represents the cases when the current and the prior mass sizes are identical. The dashed lines are the linear regression lines for the data defined by $y = 0.038x + 7.86$ for (a) and by $y = 0.857x + 1.742$ for (b). The correlation coefficient for malignant masses is 0.02 and for benign masses is 0.37.

within the ROI using a two-dimensional active contour method that was initialized by K-mean clustering.[25,26]

The texture features used in this study were calculated from run-length statistics (RLS) matrices.[27] The RLS matrices were computed from the images obtained by the rubber band straightening transform (RBST).[6] The RBST maps a band of pixels surrounding the mass onto the Cartesian plane (a rectangular region). In the transformed image, the mass border appears approximately as a horizontal edge, and spiculations appear approximately as vertical lines. A complete description of the RBST can be found in the literature.[6]

RLS texture features were extracted from the vertical and horizontal gradient magnitude images, which were obtained by filtering the RBST image with horizontally or vertically

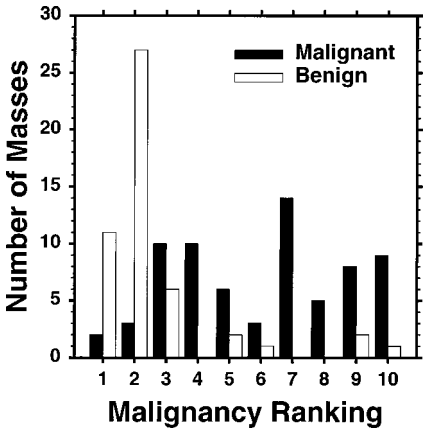

FIG. 5. The distribution of the malignancy ranking of the masses in the 120 current mammograms. The rating was performed by an experienced MQSA radiologist (1: definitely benign, 10: definitely malignant).

oriented Sobel filters and computing the absolute gradient values of the filtered image.[6] Five texture measures, namely, short run emphasis (SRE), long run emphasis (LRE), gray level nonuniformity (GLN), run length nonuniformity (RLN), and run percentage (RP) were extracted from the vertical and horizontal gradient images in two directions, $\theta = 0°$, and $\theta = 90°$. Therefore, a total of 20 RLS features were calculated for each ROI. The definition of the RLS feature measures can be found in the Appendix and in the literature.[27]

Morphological features were extracted from the automatically segmented mass shape. Five of the morphological features were based on the normalized radial length (NRL), defined as the Euclidean distance from the object's centroid to each of its edge pixels, i.e., the radial length, and normalized relative to the maximum radial length for the object.[11] The following five NRL features were extracted: mean (NRLAVG), standard deviation (NRLSD), entropy (NRLENT), area ratio (NRLAREAR), zero crossing count (NRLZCC). In addition, the perimeter (PERIM), area (AREA), circularity (CIRC), rectangularity (SQR), contrast (CONT), perimeter-to-area ratio (CRR), and Fourier descriptor (FF)

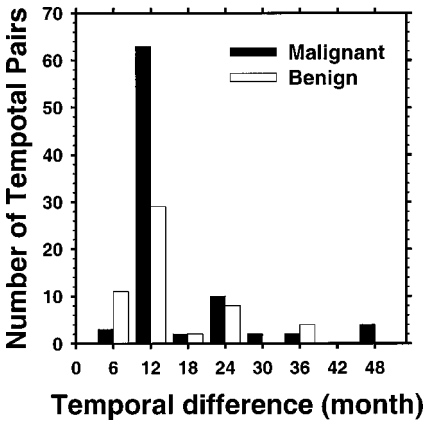

FIG. 6. Temporal interval between the current and the prior mammograms for the 140 temporal pairs in our data set.

features were extracted. The definitions of the morphological features can be found in the literature.[26,28] Three of the morphological features (perimeter, area, and perimeter-to-area ratio) are related to the mass size and thus are feature descriptors of the mass size.

A spiculation measure was defined for each pixel on the mass border by using the statistics of the image gradient direction relative to the normal direction to the mass border. The statistics was determined in a 90° sector centered about the normal at the border pixel and outside of the mass border.[25,26] The spiculation measure for each border pixel was normalized to be between 0 and $\pi/2$, with a value of $\pi/4$ indicating a random orientation of image gradients, and larger values indicating a higher likelihood of spiculation. Three features were extracted from the spiculation measure. The first feature (AVG) was the average of the spiculation measure for all pixels on the mass boundary. The second feature (PERC_ABV) was the percentage of border pixels with a spiculation measure larger than $\pi/4$, and the third feature (AVE_ABV) was the average of the spiculation measure for those pixels with a spiculation measure larger than $\pi/4$.

A total of 35 features (20 RLS, 12 morphological, and 3 spiculation) were therefore extracted from each ROI. Additionally, difference features were obtained by subtracting a prior feature from the corresponding current feature. Therefore, 35 difference features were derived from the 20 RLS, 12 morphological, and 3 spiculation features.

C. Feature selection

In order to reduce the number of the features and to obtain the best feature subset to design an effective classifier, feature selection with stepwise linear discriminant analysis[29] was applied. At each step of the stepwise selection procedure one feature is entered or removed from the feature pool by analyzing its effect on the selection criterion. In this study, the Wilks' lambda (the ratio of within-group sum of squares to the total sum of squares[30]) was used as a selection criterion. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for measuring feature correlation with the other features. In a feature entry step, the features not yet selected are entered into the selected feature pool one at a time, the significance of the change in the Wilks' lambda caused by this feature is estimated based on F statistics. The feature with the highest significance is entered into the feature pool if its significance is higher than F_{in} and its correlation value with the rest of the features in the pool is below T. In a feature removal step, the features that have already been entered in the selected feature pool are removed one at a time and the significance of the change in the Wilks' lambda is estimated. The feature with the least significance is removed from the selected feature pool if the significance is less than F_{out}. Since the appropriate values of F_{in}, F_{out} and T are not known *a priori*, we examined a range of F_{in}, F_{out}, and T values using an automated simplex optimization method.[31,32] The appropriate thresholds were chosen in such a way that a minimum number of features were selected to achieve a high accuracy of classification by LDA. More details about the stepwise linear discriminant analysis and its application to CAD can be found elsewhere.[5,6]

TABLE I. Classification results for the classifier based on the temporal change information, the classifier based on current single image information, and the classifier based on prior single image information.

Classification	Avg. No. of selected features	Training A_z	Test A_z	Test partial $A_z^{(0.9)}$
Temporal pairs	10	0.92	0.88 ± 0.03	0.37 ± 0.10
Current images	11	0.90	0.82 ± 0.04	0.32 ± 0.08
Prior images	4	0.78	0.76 ± 0.04	0.24 ± 0.08

The feature selection in this study was performed by applying the stepwise feature selection to the entire feature space (combination of texture, spiculation, and morphological features altogether) as well as subspaces obtained by different combinations of the three feature subspaces: texture, spiculation, and morphological features. The stepwise feature selection uses a sequential forward inclusion and backward elimination approach. The procedure does not exhaustively evaluate all possible combinations of individual features. It is therefore not optimal, especially when the feature space is large and the training sample is small. By limiting the input to the feature subspaces, the dimensionality was reduced compared to the entire feature space. We found that better feature subsets could be selected by the stepwise feature selection in the subspaces than in the entire feature space.

D. Evaluation methods

To evaluate the classifier performance, the training and test discriminant scores were analyzed using receiver operating characteristic (ROC) methodology.[33] The discriminant scores of the malignant and benign masses were used as decision variables in the LABROC1 program,[34] which fits a binormal ROC curve based on maximum likelihood estimation. The classification accuracy was evaluated as the area under the ROC curve, A_z. The performances of the classifiers were also assessed by estimating the partial area index ($A_z^{(0.9)}$). The partial area index ($A_z^{(0.9)}$) is defined as the area that lies under the ROC curve but above a sensitivity threshold of 0.9 ($\text{TPF}_0=0.9$) normalized to the total area above TPF_0, $(1-\text{TPF}_0)$. The partial $A_z^{(0.9)}$ indicates the performance of the classifier in the high sensitivity (low false negative) region which is most important for a cancer detection task.

III. RESULTS

The performances of the classifiers based on the temporal pairs, the current images, and the prior images are summarized in Table I. The classifiers that achieved the highest test A_z values with a small average number of features were presented here. Table II is a summary of the features selected for each classifier. For the 56 training subsets of temporal pairs used in this study, an average of 10 features were selected for

TABLE II. Selected features for classifiers based on temporal pairs, current images, and prior images. The letter "H" or "V" at the beginning of the texture feature labels indicates that the features were extracted from the horizontal or vertical gradient magnitude images, respectively. The number (0 or 90) at the end of the texture feature labels shows the direction at which the features were extracted.

			Temporal pairs			Current images	Prior images
Feature type	Group	Features	Curr	Pr	Diff	Curr	Pr
Texture	SRE	H_SRE_0			×		
		H_SRE_90	×		×		
		V_SRE_0	×		×	×	×
		V_SRE_90				×	
	LRE	V_LRE_0			×		×
		H_LRE_0				×	
	RLN	V_RLN_0	×			×	
	RP	H_RP_0	×				×
Spiculation		PERC_ABV	×			×	
		AVG		×			
		AVG_ABV					×
Morphological		CRR				×	
		NRLZCC				×	
		PERIM				×	
		NRLAVG				×	
		SQR				×	
		CONT				×	

the classification task. The most frequently selected features included 4 difference RLS features (3 SRE and 1 LRE), 4 RLS features (2 SRE, 1 RLN and 1 RP), 1 spiculation feature from the current image, and 1 spiculation feature from the prior image (Table II). The LDA classifier achieved an average training A_z of 0.92 and a test A_z of 0.88. The test partial $A_z^{(0.9)}$ was 0.37.

For classification of malignant and benign masses using the current single images (the current images of the temporal pairs), the LDA classifier selected an average of 11 features for the 56 training subsets. The most frequently selected features were 4 RLS features (2 SRE, 1 LRE and 1 RLN), 1 spiculation feature, and 6 morphological features (Table II). The classifier achieved an average training A_z of 0.90, a test A_z of 0.82, and a test partial $A_z^{(0.9)}$ of 0.32.

For the classification of masses based on the prior single images alone, an average of 4 features were selected for the 56 training subsets. The most frequently selected features were 3 RLS features (1 SRE, 1 LRE, and 1 RP) and 1 spiculation feature. The LDA classifier achieved an average training A_z of 0.78, test A_z of 0.76, and test partial $A_z^{(0.9)}$ of 0.24.

The test ROC curves for the three classifiers are compared in Fig. 7. The difference in the test A_z between the classifier based on the temporal pairs and that based on the current images alone is statistically significant ($p=0.015$). The difference in the test A_z between the classifier based on the temporal pairs and that based on the prior images alone is also statistically significant ($p=0.001$). The partial area index for the classifier based on the temporal pairs is also improved compared to the classifiers based on the current or the prior images alone, although the differences did not achieve statistical significance.

IV. DISCUSSION

Texture and spiculation features were important for malignant and benign classification of mammographic masses for all three types of classifiers: the classifier based on temporal pair information, the classifier based on current image information, and the classifier based on prior image information. One or more of the spiculation features were always selected in all training partitions for all three classifiers. The most frequently selected texture features were the short run emphasis (SRE) features. They comprised more than 50% of the texture features selected for the three classifiers (Table II).

The temporal-information-based classifier showed improved performance compared to the classifiers based on cur-

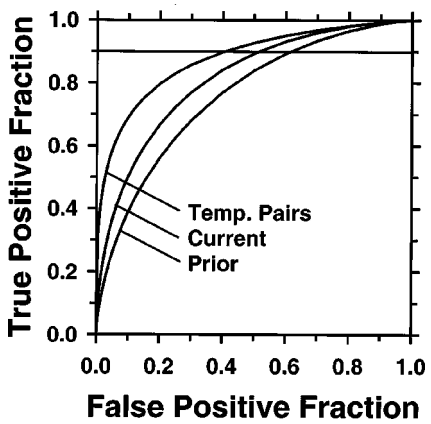

FIG. 7. The test ROC curves for the classifiers based on temporal pair information, current image information, and prior image information.

rent or prior image information alone. The input feature space to the temporal-information-based classifiers included the current, prior, and difference features. This allows the classifier to choose the individual features or the difference features. Using the stepwise feature selection procedure and the linear discriminant classifier, it was found that the texture and the spiculation features contained useful temporal information to perform malignant and benign mass classification. Texture features appeared to provide the best information by the difference features obtained from subtracting the prior from the corresponding current features (SRE and LRE difference features). On the other hand, the best use of the spiculation features appeared to be a direct combination of current and prior features in the input feature vector by the LDA since the individual features were chosen.

We found that better feature subsets could be selected by the stepwise feature selection in the subspaces than in the entire feature space. For example, for the temporal-information-based classifier, a better feature subset with a higher test A_z at 0.88 was found when the input feature space included only the texture and spiculation subspaces. The addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.84. Similarly, in the case of the classifier based on prior image information, a better feature subset was obtained when the texture and spiculation feature subspaces were used in the input feature space for stepwise feature selection. Again the addition of the morphological feature subspace to the input feature space reduced the highest test A_z to 0.72. The classifier based on current image information was the only one, among the three, that obtained a better result, as shown in Table I, when the morphological feature subspace was included in the input feature space.

One reason for the poor performance of the morphological features may be due to the fact that the masses were more subtle in the prior images. In fact, the experienced MQSA mammographer was not confident in seeing 25 of the ''masses'' on the prior images and could not provide a mass size estimation for them. Although the active contour model would stop the iteration based on the preset criteria and found an ''outline'' of the masses on the prior mammograms, generally these mass outlines were less reliable than those on the current masses in providing morphological characteristics of the masses. Texture features did not depend as strongly on the precise mass boundary as morphological features. Three out of the four features selected for classification of the malignant and benign masses on the prior images were RLS texture features. A spiculation feature was also found to be a good discriminator.

We also performed ROC analysis of the malignancy confidence ratings provided by the experienced MQSA radiologist for the current image data set (120 images). The distribution of the malignancy ratings is shown in Fig. 5, which resulted in an A_z value of 0.80 ± 0.04. This indicates that the masses in the current mammograms cannot be easily distinguished as malignant or benign even by an experienced radiologist, consistent with the fact that all lesions had indeed undergone biopsy. The classifier based on the current image information has an A_z value of 0.82 ± 0.04, similar to the accuracy of the radiologist for this data set.

In this study, the locations of the masses were identified manually on both the current and the prior mammograms by a radiologist. This simulated the situation when a radiologist finds a mass either in a diagnostic or a screening setting and call upon the CAD algorithm to seek a second opinion on the likelihood of malignancy of the mass based on the interval change information. We are developing an automated regional registration technique that can automatically locate the mass on the prior mammogram based on its location on the current mammogram. The location of the mass on the current mammogram can be identified by a radiologist or by an automated mass detection algorithm. In the latter case, the process of mass detection, current and prior mass registration, and classification can be fully automated. The analysis of interval change can be incorporated as one of the functions provided by a CAD system for interpretation of mammograms.

In this study, we employed a simple measure of temporal change by taking the difference between the feature from the current mass and the corresponding feature from the prior mass. We observed improvement in classification with this simple temporal information. It will be important to evaluate other similarity measures that can characterize small difference in image features of the object of interest. It can be expected that a more sensitive similarity measure will provide a better measurement of dissimilarity, or difference, between the current and prior masses and further improve the utilization of the temporal change information on mammograms.

V. CONCLUSION

We performed a preliminary study to evaluate the effectiveness of interval change analysis for classification of malignant and benign masses on mammograms. It was found that the difference RLS texture features and spiculation features were useful for identification of malignancy in temporal pairs of mammograms. The information on the prior image was important for characterization of the masses; 5 out of the 10 selected features contained prior information. We found that the mass size descriptors were not discriminatory features for these difficult cases because many of the benign masses also grew over time. In comparison with the classification based on image information from the current images alone, the temporal change information significantly ($p = 0.015$) improved the accuracy for classification of the masses in terms of the total area under the ROC curve (A_z). The partial area under the ROC curve for the classifier based on the temporal pairs ($A_z^{(0.9)} = 0.37$) is also improved compared to the classifier based only on the current images ($A_z^{(0.9)} = 0.32$), although the difference did not achieve statistical significance. Further studies are under way to improve this temporal change classification technique and to evaluate its performance on a larger data set.

ACKNOWLEDGMENTS

This work is supported by a Career Development Award from the USAMRMC (No. DAMD 17-98-1-8211) (L.H.), USPHS Grant No. CA 48129, and a USAMRMC grant (No. DAMD 17-96-1-6254). The content of this publication does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned in the publication should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

APPENDIX: RUN LENGTH STATISTICS TEXTURE FEATURES

A gray level run length is a set of consecutive collinear pixels all having the same gray level value. The length of the run is the number of pixels in the run. For a given image it is possible to compute a gray level run length matrix for runs in any given direction. In this study, two directions are used: $\theta=0°$, and $\theta=90°$. Let $p(i,j)$ be the number of times there is a run of length j that has a gray level i. Let N_g be the number of gray levels and N_r be the number of runs. The short run emphasis is defined as

$$\mathrm{SRE}=\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{p(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}.$$

This feature divides the frequency of each run length by the length of the run squared. This tends to emphasize short runs. The denominator is the total number of runs in the image and serves as a normalizing factor. The long run emphasis is defined as

$$\mathrm{LRE}=\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2p(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}.$$

This feature multiplies the frequency of each run length by the length of the run squared. This tends to emphasize long runs.

The gray level nonuniformity is defined as

$$\mathrm{GLN}=\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_g}p(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}.$$

This feature squares the number of run lengths for each gray level. This measures the gray level nonuniformity of the image. If the runs are equally distributed over all gray levels, the feature takes on its lowest values. A larger run length contributes more to the feature value.

Run length nonuniformity is defined as

$$\mathrm{RLN}=\frac{\sum_{j=1}^{N_r}(\sum_{i=1}^{N_r}p(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}.$$

This feature measures the nonuniformity of the run lengths. If the runs are equally distributed over all lengths, the feature will have a low value. A larger run contour contributes more to the feature value.

Run percentage is defined as

$$\mathrm{RP}=\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j)}{P}.$$

This feature is a ratio of the total number of runs to the total number of possible runs (P) if all runs have a length of one.

The above-given definitions are based on Galloway[27] and more details can be found in this reference.

[a]Electronic mail: lhadjisk@umich.edu

[1] H. C. Zuckerman, "The role of mammography in the diagnosis of breast cancer," in *Breast Cancer, Diagnosis and Treatment*, edited by I. M. Ariel and J. B. Cleary (McGraw–Hill, New York, 1987).
[2] L. Tabar and P. B. Dean, "The control of breast cancer through mammography screening," Radiol. Clin. North Am. **25**, 993–1005 (1987).
[3] L. W. Bassett, B. Shayestehfar, and I. Hirbawi, "Obtaining previous mammograms for comparison: Usefulness and costs," AJR, Am. J. Roentgenol. **163**, 1083–1086 (1994).
[4] E. A. Sickles, "Periodic mammographic follow-up of probably benign lesions: Results in 3183 consecutive cases," Radiology **179**, 463–468 (1991).
[5] H. P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space," Phys. Med. Biol. **40**, 857–876 (1995).
[6] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," Med. Phys. **25**, 516–526 (1998).
[7] H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, K. L. Leung, D. D. Adler, and M. M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network," Phys. Med. Biol. **42**, 549–567 (1997).
[8] L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, and M. A. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach," IEEE Trans. Med. Imaging **18**, 1178–1187 (1999).
[9] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer," Radiology **187**, 81–87 (1993).
[10] V. Goldberg, A. Manduca, D. L. Evert, J. J. Gisvold, and J. F. Greenleaf, "Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence," Med. Phys. **19**, 1475–1481 (1992).
[11] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computer-aided image analysis," IEEE Trans. Med. Imaging **12**, 664–669 (1993).
[12] Z. M. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," Acad. Radiol. **5**, 155–168 (1998).
[13] M. Sallam and K. Bowyer, "Detecting abnormal densities in mammograms by comparison with previous screenings," in *Digital Mammography 96*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996).
[14] D. Brzakovic, N. Vujovic, M. Neskovic, P. Brzakovic, and K. Fogarty, "Mammogram analysis by comparison with previous screenings," in *Digital Mammography*, edited by A. G. Gale, S. M. Astley, D. R. Dance, and A. Y. Cairns (Elsevier, Amsterdam, 1994).
[15] N. Vujovic and D. Brzakovic, "Establishing the correspondence between control points in pairs of mammographic images," IEEE Trans. Med. Imaging **6**, 1388–1399 (1997).
[16] S. S. Gopal, H.-P. Chan, N. Petrick, T. E. Wilson, B. Sahiner, M. A. Helvie, and M. Goodsitt, "A regional registration technique for automated analysis of interval changes of breast lesions," Proc. SPIE **3338**, 118–131 (1998).
[17] S. S. Gopal, H.-P. Chan, T. E. Wilson, M. A. Helvie, N. Petrick, and B.

Sahiner, ''A regional registration technique for automated interval change analysis of breast lesions on mammograms,'' Med. Phys. **26**, 2669–2679 (1999).
[18] L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, and S. Sanjay-Gopal, ''Automated identification of breast lesions in temporal pairs of mammograms for interval change analysis,'' Radiology **213(P)**, 229–230 (1999).
[19] L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, M. A. Helvie, S. Paquerault, and C. Zhou, ''Interval change analysis in temporal pairs of mammograms using a local affine transformation,'' Proc. SPIE **3979**, 847–853 (2000).
[20] L. M. Hadjiiski, H. P. Chan, B. Sahiner, N. Petrick, and M. A. Helvie, ''Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis—Local affine transformation for improved localization,'' Med. Phys. **28**, 1070–1079 (2001).
[21] W. F. Good, B. Zheng, Y. H. Chang, Z. H. Wang, and G. S. Maitz, ''Generalized procrustean image deformation for substraction of mammograms,'' Proc. SPIE **3661**, 1562–1573 (1999).
[22] S. S. Gopal, H.-P. Chan, B. Sahiner, N. Petrick, T. E. Wilson, and M. A. Helvie, ''Evaluation of interval change in mammographic features for computerized classification of malignant and benign masses,'' Radiology **205(P)**, 216 (1997).
[23] L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, ''Computer-aided classification of malignant and benign breast masses by analysis of interval change of features in temporal pairs of mammograms,'' Radiology **217(P)**, 435 (2000).
[24] L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. Gurcan, ''Analysis of temporal change of mammographic features for computer-aided characterization of malignant and benign masses,'' Proc. SPIE **4322**, 661–666 (2001).
[25] B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, M. A. Helvie, and S. Paquerault, ''Active contour models for segmentation and characterization of mammographic masses,'' The Fifth International Workshop on Digital Mammography Proceedings, 2000, IWDM-2000, pp. 357–362.
[26] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski, ''Improvement of mammographic mass characterization using spiculation measures and morphological features,'' Med. Phys. **28**, 1455–1465 (2001).
[27] M. M. Galloway, ''Texture classification using gray level run lengths,'' Comput. Graph. Image Process. **4**, 172–179 (1975).
[28] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, ''Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms,'' Med. Phys. **26**, 1642–1654 (1999).
[29] M. J. Norusis, *SPSS for Windows Release 6 Professional Statistics* (SPSS, Chicago, IL, 1993).
[30] M. M. Tatsuoka, *Multivariate Analysis, Techniques for Educational and Psychological Research*, 2nd ed. (Macmillan, New York, 1988).
[31] S. S. Rao, *Optimization: Theory and Applications* (Wiley, New York, 1979).
[32] F. A. Lootsma, *Numerical Methods for Non-linear Optimization* (Academic, New York, 1972).
[33] C. E. Metz, ''ROC methodology in radiologic imaging,'' Invest. Radiol. **21**, 720–733 (1986).
[34] C. E. Metz, B. A. Herman, and J. H. Shen, ''Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data,'' Stat. Med. **17**, 1033–1053 (1998).