Medical Physics
The International Journal of Medical Physics Research and Practice

A minimum spanning forest based classification method for dedicated breast CT images

Robert Pike, Ioannis Sechopoulos, and Baowei Fei

Citation: Medical Physics **42**, 6190 (2015); doi: 10.1118/1.4931958
View online: http://dx.doi.org/10.1118/1.4931958
View Table of Contents: http://scitation.aip.org/content/aapm/journal/medphys/42/11?ver=pdfcov
Published by the American Association of Physicists in Medicine

A minimum spanning forest based classification method for dedicated breast CT images

Robert Pike
Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia 30329

Ioannis Sechopoulos
Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia 30329 and Winship Cancer Institute of Emory University, Atlanta, Georgia 30322

Baowei Fei[a)]
Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia 30329; Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, Georgia 30322; Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322; and Winship Cancer Institute of Emory University, Atlanta, Georgia 30322

(Received 14 March 2015; revised 6 July 2015; accepted for publication 11 September 2015; published 8 October 2015)

Purpose: To develop and test an automated algorithm to classify different types of tissue in dedicated breast CT images.
Methods: Images of a single breast of five different patients were acquired with a dedicated breast CT clinical prototype. The breast CT images were processed by a multiscale bilateral filter to reduce noise while keeping edge information and were corrected to overcome cupping artifacts. As skin and glandular tissue have similar CT values on breast CT images, morphologic processing is used to identify the skin based on its position information. A support vector machine (SVM) is trained and the resulting model used to create a pixelwise classification map of fat and glandular tissue. By combining the results of the skin mask with the SVM results, the breast tissue is classified as skin, fat, and glandular tissue. This map is then used to identify markers for a minimum spanning forest that is grown to segment the image using spatial and intensity information. To evaluate the authors' classification method, they use DICE overlap ratios to compare the results of the automated classification to those obtained by manual segmentation on five patient images.
Results: Comparison between the automatic and the manual segmentation shows that the minimum spanning forest based classification method was able to successfully classify dedicated breast CT image with average DICE ratios of 96.9%, 89.8%, and 89.5% for fat, glandular, and skin tissue, respectively.
Conclusions: A 2D minimum spanning forest based classification method was proposed and evaluated for classifying the fat, skin, and glandular tissue in dedicated breast CT images. The classification method can be used for dense breast tissue quantification, radiation dose assessment, and other applications in breast imaging. [http://dx.doi.org/10.1118/1.4931958]

Key words: breast computed tomography (BCT), cupping artifact correction, support vector machine, minimum spanning forest, image classification, breast cancer, segmentation

1. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer in women.[1] The chance of a woman born today in the United States developing breast cancer throughout her lifetime if she makes it to 80 yr old is 1 in 8.[2] It has been well established that earlier detection leads to drastic increases in survival rates.[3] For women diagnosed with nonmetastatic breast cancer, the risk of a distant recurrence is related to the number of axillary nodes and tumor size,[4] both of which increase in later stages of the disease and are closely related.[5] For these reasons, detection methods that can determine the risk of development and can detect the cancer at early stages are needed.

Traditional imaging modalities for breast disease include mammography. Mammography compresses the breast and uses low energy x-rays to obtain a 2D image.[6] The 2D nature of mammography results in tissue superimposition, which leads to inaccuracies by giving the appearance of or masking an abnormality.

Dedicated breast computed tomography (DBCT) is a relatively new imaging modality that can provide 3D images of the breast with a high spatial and contrast resolution. Although conventional computed tomography has been extensively utilized in clinical studies, it faces the challenges of high cost and additional dose through the thoracic cavity when used for breast imaging.[7] DBCT has the additional advantage of imaging the breast without the need for compression, increasing patient comfort.[8]

Classification of breast tissue can provide quantitative assessments of breast tissue composition, density, and

distribution that can be used to evaluate the risk of breast cancer or for dose assessment.[9–11] In addition, changes in the tissue density and distribution could also be monitored over time to determine cancer risk.[12] Chang and Lin have proposed an automatic volumetric segmentation scheme by partitioning a histogram into intervals which are subsequently used as thresholds for the entire image.[13] The use of support vector machines (SVMs) to classify CT images has been studied using different kernel approaches.[14] Anderson *et al.* explored the use of computerized scene segmentation for ductal lesions in the breast CT image.[15] Density based breast classification has also been explored in an effort to create a rating for breast tissue composition.[16] Other methods have focused on cluster detection to locate microcalcification clusters, and these techniques have been used to assess breast cancer with constrained modulus reconstruction.[17,18] Huang and coworkers proposed a radial-geometry edge detection scheme to measure the breast skin thickness on coronal reconstructed breast CT images and they found that the mean breast skin thickness was 1.45 ± 0.30 mm.[19] In another study by Shi and coworkers,[20] a similar algorithm was used to determine the skin thickness and the result on the breast skin thickness was in excellent agreement with Huang *et al.*[19] In this study, we focus on the classification of the tissue within the breast, which includes the glandular and adipose tissue.

In this study, we investigated a minimum spanning forest (MSF) classification approach for breast tissue classification. Minimum spanning forests were introduced as a region based method for image segmentation that is robust to image noise.[21] The reasoning of minimum spanning forests is their ability to incorporate local and global information into the segmentation process by allowing the branches to span the entire image.[22] Minimum spanning forests have been used in color images for facial detection and segmentation[23] and demonstrated the robust nature of the segmentation algorithm even when similarly colored features were in the background of the image. To provide the required markers for the minimum spanning forest algorithms to begin their growth, support vector machines have been used and proven to be applicable for this purpose.[24] Support vector machines have been designed for color image classification on a pixelwise basis.[25] They have also been extensively studied for feature extraction from histograms of images.[26] SVMs in conjunction with other methods have been successfully used for remote sensing data classification.[27–31] Minimum spanning forest based methods have been used for segmentation of other images such as hyperspectral images and have been combined with support vector machines to classify various regions.[32] Other works have incorporated the use of a probabilistic support vector machine to determine highly probable markers for minimum spanning forest based classification.[24] This approach has shown that the minimum spanning forest can improve the pixelwise classification of the support vector machine and is a more robust segmentation algorithm.

The proposed algorithm calls for pixelwise classification to obtain markers to be used as roots of a minimum spanning forest. The effectiveness of this algorithm comes from its ability to effectively incorporate spatial and intensity based information. It also allows for variations in threshold limits and multiple parameters to be tailored specifically to breast CT images. The robustness of the proposed algorithm was demonstrated in patient data. The contribution of this paper includes (1) the use of a minimum spanning forest and support vector machine to accurately classify different tissue types in breast data of human patients and (2) a dynamic dissimilarity measure that better incorporates global information into a local region growing method to account for local noise. Sections 2–5 will describe the classification methods, evaluation results, and the application.

2. MATERIALS AND METHODS

The processing and classification of the breast CT images consist of six major steps: (1) A cupping artifact correction method is used to reduce the cupping artifacts in breast CT images. (2) A multiscale bilateral filter is applied to remove noise but retain edge information. (3) A morphologic method is used to classify the skin of the breast. (4) A support vector machine is used to perform pixelwise classification based upon intensity information. (5) A minimum spanning forest is grown from selected markers to combine spatial and intensity information for image classification. Figure 1 shows the flow chart

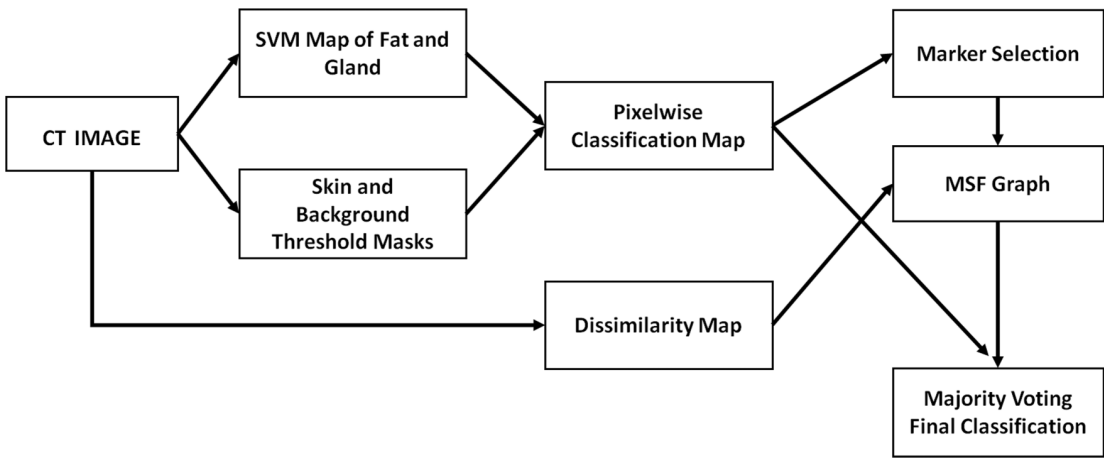

FIG. 1. Outline of the minimal spanning forest algorithm.

that summarizes the proposed method. Details on the methods used for these steps are provided here. Image acquisition and validation methods are also described in this section.

2.A. Cupping artifact correction for breast CT images

Cupping artifact correction was performed on the breast CT images following the procedure previously reported by us.[12] The main reason for the cupping artifacts is the inclusion of scattered x-rays in the CT projections, as well as beam hardening through the imaged breast. The method uses a coefficient matrix of size $3\times3\times3$ with all values set to 1 (null cupping artifacts). These values are updated iteratively to decrease the entropy of the corrected image. A minimization step is used to modify each coefficient until the entropy value becomes stable. When the entropy becomes stable, the number of coefficients of the matrix is increased using a B-spline interpolation. For this study, the number of bias functions was increased along levels, set to 3, 5, 6, 9, and 11. The number of gray levels $N1$ was equal to 128 and the number of local gray levels $N2$ was equal to 100. The entire process was implemented in MATLAB to preprocess the breast CT images.

2.B. Morphologic operations for skin classification

To segment the skin tissue, the first step is a morphologic operation to obtain a rough classification. The position information is used to perform this procedure.[12] The skin thickness in breast is reported to be 1.45 ± 0.30 mm.[19,20,33] The voxel size of the images is known from the image acquisition. A two-mask method is performed to first exclude the background, and second to obtain the inner mask that contains the tissues beneath the skin. The first mask is created using a threshold technique. The threshold is set at 85% of the intensity of the second peak of the histogram. The first peak is the background, and the second is the start of breast tissue. Thus, by subtracting the voxels below this threshold, we obtain an outer mask containing only breast and skin tissue. A $9\times9\times9$ box is then used to perform erosion operations so that an inner skin mask can be obtained. By subtracting the inner mask from the outer mask, the rough classification of the skin can be obtained. This skin classification coupled with the subsequent pixelwise classification of fat and glandular tissue will create the complete map from which markers can be selected for the following classification step.

2.C. SVM-based classification

The next step in this process involves the preliminary classification of each individual pixel. This provides a framework for the minimum spanning forest to be grown on. Support vector machines are well suited to evaluate images on a pixelwise level in order to provide a classification for each individual pixel. SVMs are able to handle large amounts of training and testing data to provide accurate labeling of pixels. The output of this classification will provide both a classification as well as a probability map of all the pixels. The classification map will simply label each pixel into its respective class, and the probability map will give the probability that the pixel falls into that class. The support vector machine from the LIBSVM library[34] was used in our experiment.

Support vector machines are often used for binary classification methods. Given a training set of instances and labels (x_i,y_i), $i=1,\ldots,L$ where x_i is an element of R^n and y is an element of $\{-1,1\}^L$, the support vector machine solves the following optimization problem to map the training vectors x_{i-} into a higher or infinite dimensional space using the function φ:

$$\min\frac{1}{2}w^Tw+C\sum_{i=1}^{L}\varepsilon_i$$

such that

$$y_i\left(w^T\varphi(x_i)+b\right)\geq1-\varepsilon_i,$$

while

$$\varepsilon_i\geq0,$$

where C is greater than zero and is the penalty parameter for the error term, and ε_i is the slack variable which provides a soft margin and can be used as a measure degree of misclassification of the data.[13] The solution of this optimization problem allows the SVM to find a linear separating hyperplane that has the largest margin between the training data labels. The kernel function $K(x_i,x_j)=\varphi(x_j)^T\varphi(x_j)$ allows for nonlinear classifiers.[35] For this study, the Gaussian radial basis function was used as the kernel function which maps the data into a Hilbert space of infinite dimensions.

The images were first normalized using the skin segmentation previously obtained; this was accomplished with a normalization factor n given by

$$n=\frac{S_v}{\bar{S}},$$

where S_v is an arbitrary skin intensity value which was set at 150 for this study, and S is the set of pixels from which the morphological erosion masked as skin. This normalization was performed on all training and testing data prior to training the SVM model. This method for normalization allows the spatial information to be used to normalize the intensity information of each voxel.

2.D. Connected component labeling and marker selection

There are two primary marker selection techniques that are evaluated in this algorithm. The first technique is a probability based method developed by Tarabalka *et al.*,[24] which uses both spatial and probability information from the SVM. The second technique stems from a previous method[32] that randomly selects a percentage N of the total pixels in each classification to be used as markers. These techniques as well as a simple threshold were evaluated. The simple threshold technique only selects markers above a probability P to be used as markers. This technique proved unstable and led to great overclassification or underclassification depending on small adjustments to P that varied greatly image to image;

for this reason, only the first two techniques discussed were implemented in the experiment. The first technique, using probability data, ensures that both highly probable pixels and large regions are given at least one marker, while the second technique, using random markers, has the advantage of not requiring any probability data and uses a majority voting over multiple segmentations to acquire an accurate classification.

To implement the first technique, connected component labeling must first be performed. In this experiment, a connected component labeling algorithm using a union-find data structure was used. The eight nearest neighbors surrounding each pixel were used to find connected components. This connected component labeling is performed on the SVM classification map and finds connected regions of the same label type. Each of these connected regions is then evaluated individually and separated into two categories based upon the number of pixels M. If a region has M or greater pixels, it is considered a large region, and if it contains fewer than M pixels, it is a small region. The following rules govern the selection of markers based upon region type:

1. For large regions with M or greater pixels, the top N percent of pixels within that region are selected as markers.
2. For small regions ranging from 1 to $M-1$ pixels, only pixels with a probability greater than P will be selected as markers.

This method of marker selection is used with the motivation to ignore the oversegmentation associated when no spatial information is considered. It also ensures that for regions of sufficient size at least one marker will always be present, eliminating one cause of undersegmentation. It is important to note that the markers selected need not be spatially adjacent and are considered independent in our algorithm following their selection. Previous studies[24] have associated all markers of the same type with one single tree root, and the proposed algorithm allows each marker to have its own root and later be classified by majority voting.

The second technique for marker selection implemented in this study was the use of random markers. The motivation of this technique is to avoid the need for a probability classification and to keep a consistent marker selection across all tissue types. This algorithm first uses the classification map to create lists of each label type. A percentage of pixels P are then randomly selected from these lists and are used as markers. This technique does not rely on accurate probabilities for pixels, but instead requires that the segmentation process be run multiple times, and a simple majority voting across all instances determines the final classification map.

2.E. Minimum spanning forest spatial and intensity segmentation

A minimum spanning forest provides the spatial component of this classification algorithm. Given well selected markers, a minimum spanning forest can accurately determine regional boundaries making it well suited for breast CT images. Tissue growth in the body often expands in connected regions that a minimum spanning forest can follow and classify. To grow a minimum spanning forest, edge weightings between each marker must be calculated. The CT images evaluated in this experiment were grayscale images rescaled to values ranging from 0 to 255. Edge weightings between a pixel and its eight nearest neighbors were calculated by taking a simple difference between the pixels given by

$$W_{i,j} = |P_i - P_j|. \tag{1}$$

To construct a minimum spanning forest, we first define the undirected graph G. This graph is constructed from the original grayscale image, where each pixel is considered a vertex V, with edges E connecting a pixel to its surrounding neighbors. For this experiment, we used a varying number of neighbors for both 2D and 3D images. The set of weightings W described in Eq. (1) are used to quantify the edges E of this undirected graph. The graph G is then defined as $G = (V,E,W)$, from which the spanning tree T can be grown.

From the undirected connected graph G, a spanning tree $T = (V,E_T)$ can be constructed while E_T is a subset of E. A minimum spanning tree, $T_{\min}$, of the graph G is defined as the spanning tree $T_{\min} = (V, E_{T_{\min}})$ such that the associated edge weightings W of $T_{\min}$ are minimal given by

$$T_{\min} \in \arg\min_{T \in \mathrm{ST}} \left\{ \sum_{e_{i,j} \in E_T} w_{i,j} \right\}, \tag{2}$$

where ST is the set of all possible spanning trees constructed from the graph G.[36]

Similarly, a spanning forest $F = (V,E_F)$ is defined as a nonconnected graph without cycles while E_F is a subset of E, and the minimum spanning forest $F_{\min}$ can be defined by

$$F_{\min} \in \arg\min_{F \in \mathrm{SF}} \left\{ \sum_{e_{i,j} \in E_F} w_{i,j} \right\}, \tag{3}$$

with SF being the set of all constructed spanning forests, with the same roots, of the graph G. To grow a minimum spanning forest on a specific set of M roots, additional vertices r_i, $i = 1,\ldots,M$ are added. These vertices connect the root r_i to a previously determined marker and are used as the basis for the growth of the minimum spanning forest. If an additional root R is added such that R is connected with null weighting to the additional vertices r_i, a minimum spanning tree of the graph G from the selected markers can be obtained. A minimum spanning forest is then created when the vertex R is removed. Alternative minimum spanning tree algorithms[37,38] can be implemented but Prim's algorithm offers an efficient implementation when using a binary heap to store the edge weightings.[39] This algorithm allows for the time complexity of $O(|E| \log |V|)$.[40]

To grow the minimum spanning forest using Prim's algorithm, first the root markers and their associated edges are added to the binary heap, while the vertices associated with these markers are added to the classification map. Subsequently, the edge of minimal weighting, not connected to a currently labeled pixel, is removed from the binary heap

and that vertex is added to the classification map and given the label of its associated marker. This iteration is repeated until all pixels in the classification map have been labeled, producing a segmentation map using a minimum spanning forest.[41]

In this study, we modified previous 2D minimum spanning forest algorithms to classify 3D CT images. This allows the forest to grow along the same channels that the glandular tissue expands along. Instead of segmenting individual 2D slices and combining the results, this algorithm allows for the rooted markers to expand along the third dimension. For this study, the nearest six neighbors (north, south, east, west, up, and down) were used to evaluate the 3D image classification.

This study introduced a new method for calculating dissimilarity while the minimum spanning forest is being grown. The weightings are initially calculated using Eq. (1). Prim's algorithm iteratively adds the reaming edge weightings of each pixel that has been labeled. When these edges are added to the binary heap, our modified algorithm created new edge weightings that reflect the classification of the labeled marker. Two methods can then be used to construct these new weightings. The first method calculated an additional weighting between the connecting pixel and the markers from which that label is to be classified. The second method creates a weighting between the connecting pixel and all the pixels connected to that pixels branch of the minimum spanning forest. These methods create a more robust segmentation process that better distinguishes along noisy gradients. In this study, the first method was used to determine dissimilarity; this method works well with CT images because similarly classified regions have low varying intensity values across the entire image. The second method would be employed when intensity values vary between local regions.

When random markers are used, the entire marker selection and segmentation process are repeated C times and the mode of each pixel's classification is the label given to that pixel.

2.F. Majority voting

The classification of each pixel is given by the label of the marker from which it is connected. Since a minimum spanning forest is an unconnected graph, there will only be one marker associated with each pixel. To account for errors in classification stemming from a misclassified label, we introduce a majority voting rule. Previous methods[24] have used connected components to determine regions and perform majority voting across entire regions. The proposed method instead calls for a majority voting to be performed for each branch of the minimum spanning forest. This is illustrated in Fig. 2. Each marker is given a unique label in the growing of the minimum spanning forest allowing for a greater distinction when performing majority voting. This method allows not only for large regions to be reclassified but also region boundaries to be adjusted more finely, increasing accuracy along boundaries. This method calls first for a classification map to be constructed with each marker being given a unique label. Following each of these branches along the minimum spanning forest gives the pixels associated with each marker; the mode of the SVM classification associated with these pixels gives the label to be assigned to the entire branch.

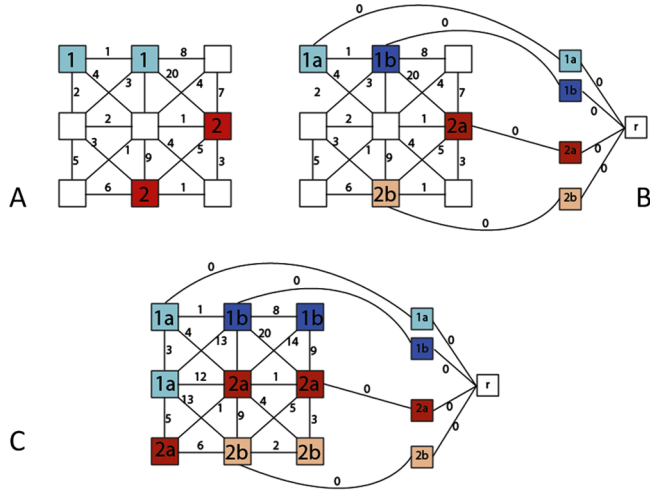

FIG. 2. Visualization of the minimum spanning forest construction from the marker selection (A) to the marker labeling (B) and to the complete construction of the minimum spanning forest (C) that will be classified with majority voting.

2.G. Breast CT image acquisition

The proposed method detailed above was tested on a set of five different human subjects. The CT images were acquired with a breast CT prototype system (Koning Corporation, West Henrietta, NY). A breast CT scan involves acquisitions of 300 projections over a 360° revolution of the X-ray tube and detector around a vertical axis in 10 s. The tube voltage of the system is fixed at 49 kVp giving an x-ray spectrum having a first half value layer of 1.39 mm Al.[42] The maximum allowed tube current is set at 100 mA, with the appropriate tube voltage being selected automatically using two scout projection images. This gave an average glandular dose of approximately 8.5 mGy to an average breast.[42] With a detector pixel size of 388 μm, the reconstructed voxel size of the image is $0.27 \times 0.27 \times 0.27$ mm^3. No scatter correction method was used and the Feldkamp–Davis–Kress (FDK) algorithm was used for reconstruction. The human study was approved by the Institutional Review Board (IRB) of Emory University.

2.H. Classification evaluation

To evaluate the results of the automatic segmentation and classification, manual segmentation, which was conducted by a radiologist with over 10 years of experience in CT imaging, was used as the gold standard. Analyze 10.0 (AnalyzeDirect, Inc., Overland Park, KS) was used to conduct the manual segmentation prior to the computer classification being performed. The manual segmentation was done in seven steps that were performed in Analyze: (1) Smooth 3D images using a five-point 3D median filter. (2) Select three, 2D

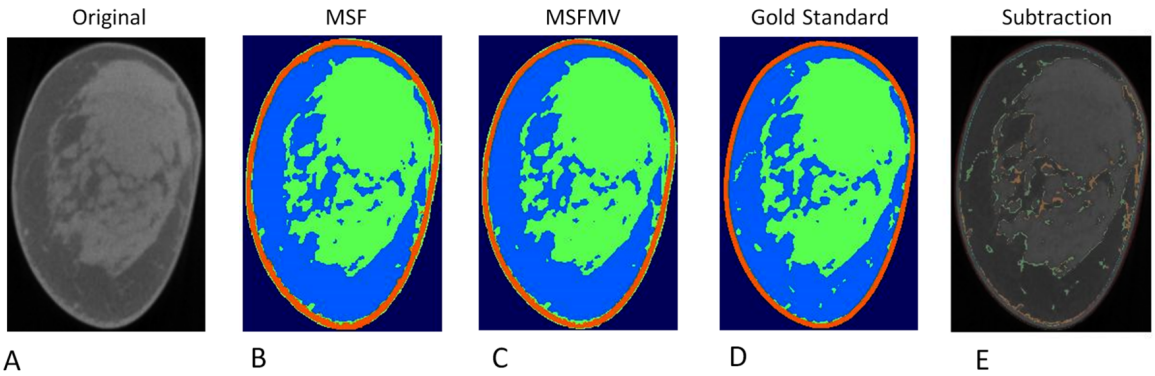

FIG. 3. Classification of the original image (A), the MSF result (B), the MSF with majority voting result (C), the gold standard by manual segmentation (D), and the difference between the MSF result and the gold standard (E).

images from the 3D volume data for segmentation. (3) Obtain the breast mask by setting a threshold for each of the three selected images. (4) Create an inner skin mask by increasing the threshold previously obtained in order to obtain a visualization of the skin, along with manually selecting interior boundary points of the skin. (5) Combine the breast mask and the skin inner mask to segment the skin. (6) Use multiple operator-selected thresholds that are individually applied to small areas of the image to segment glandular tissue. The presence of cupping artifacts restricts the feasibility to separate glandular from adipose voxels over large regions. The glandular tissue segmentation is a full combination of the individual regions. (7) The fat is segmented by subtracting both the skin and the glandular tissue from the breast mask. This entire manual process required often over 1 h/image. Figure 3 shows the results of this final classification method on patient data.

The automatic classification method has been tested in both simulation and real patient data. The simulation data have the ground truth for the validation of the algorithm. The real patient data were used to test the applicability of the automatic classification method. Simulated images were constructed from the five real human subject images. These simulated images were made using values taken from the manually segmented gold standard data. As seen in Fig. 4, the images were constructed with either high density or low density glandular tissue. These images were created by rooting a random walk inside the simulation image which grew to create the glandular tissue. The high density and low density images were created by controlling the distance of the walks and the density of the roots. The simulation images were created to be an accurate representation of the real patient data by using randomly selected pixels from the real patients to fill the respective classification labels.

To evaluate the accuracy of the automatic segmentation, five different measures were used, the overall accuracy (OA), average accuracy (AA) of the images, sensitivity, specificity, and DICE overlap ratio,

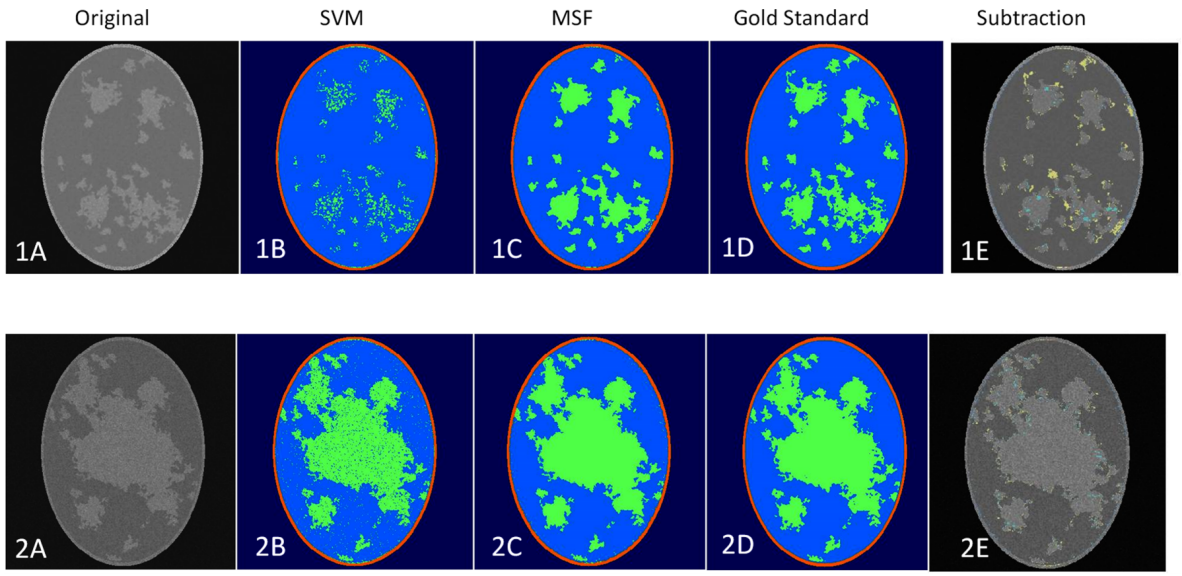

FIG. 4. Classification of the low density simulation image (1A), the SVM result (1B), the MSF result (1C), the gold standard by manual segmentation (1D), and the difference between the MSF result and the gold standard (1E), and classification of the high density simulation image (2A), the SVM result (2B), the MSF result (2C), the gold standard by manual segmentation (2D), and the difference between the MSF result and the gold standard (2E).

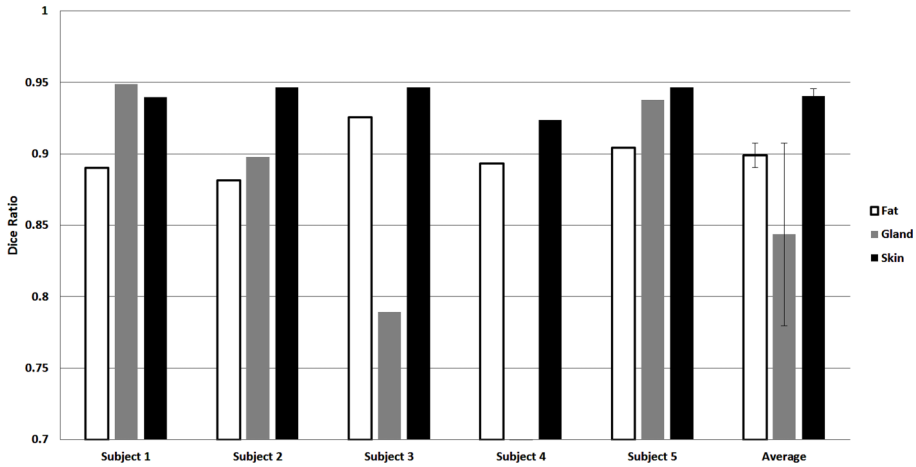

FIG. 5. The DICE overlap ratios for the fat, glandular, and skin tissue when the support vector machine was used for pixelwise classification for the simulation images.

$$\mathrm{OA} = \frac{\text{correctly classified pixels}}{\text{total number of pixels}},$$

$$\mathrm{AA} = \frac{\sum_{i=1}^{N} \frac{S_i \cap G_i}{|G_i|}}{N},$$

where S_i are the voxels classified by the proposed algorithm to a specific class i and G_i are the voxels manually classified to a specific class i, and N is the number of images,

$$\text{Sensitivity} = \frac{S \cap G}{|G|},$$

where S represents the voxels of one type associated with the automatic segmentation and G represents the voxels of the same type associated with the manual gold segmentation,

$$\text{Specificity}_i = \frac{S_j \cap G_j}{|G_j|},$$

where S_j and G_j are the respective voxels that do not belong to the class i in the automatic and manual classifications,

$$\text{Dice}(S,G) = \frac{2|S \cap G|}{|S| + |G|},$$

where S represents the voxels of one type associated with the automatic segmentation and G represents the voxels of the same type associated with the manual gold standard segmentation.

3. RESULTS

Figure 3 shows the result of the automatic method and its comparison with the gold standard for a typical patient, demonstrating the good performance of the automatic classification method in human patient data.

3.A. Results of simulated images

The support vector machine was able to classify the simulated images with average DICE overlap ratios for the fat, glandular, and skin tissues at $89\% \pm 1.7\%$, $84\% \pm 12.8\%$, and $94\% \pm 0.9\%$, respectively, as seen in Fig. 5.

The minimum spanning forest was then used and produced average DICE overlap ratios for the fat, glandular, and skin tissues at $96\% \pm 0.3\%$, $92\% \pm 4.5\%$, and $92\% \pm 1.4\%$, respectively, as shown in Fig. 6. These simulation results demonstrate

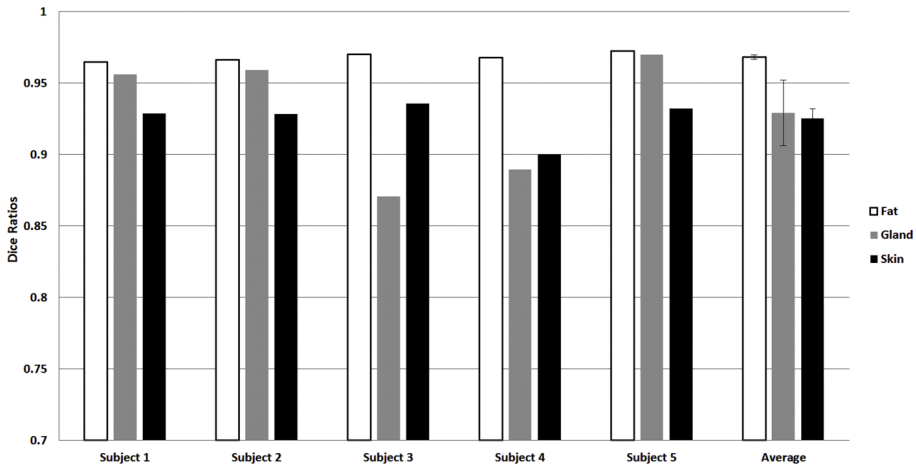

FIG. 6. The DICE overlap ratios for the fat, glandular, and skin tissue when the minimum spanning forest was constructed in 2D using the optimal parameters for the simulation data.

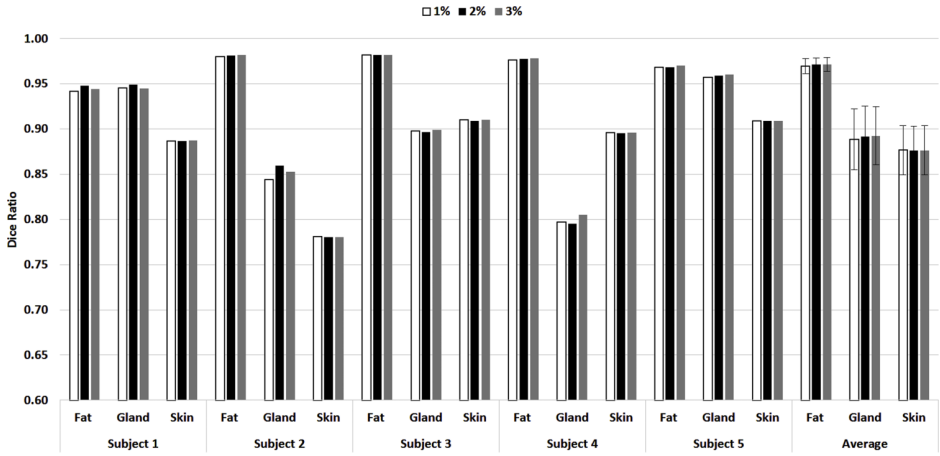

FIG. 7. The DICE overlap ratios for the support vector machine classification when 1%, 2%, and 3% of randomly selected gold standard data was used for training.

the ability of the algorithm to accurately classify the CT images with a high accuracy.

3.B. Results of support vector machine classification for human data

The SVM algorithm classified the images on a pixelwise basis successfully. This pixelwise classification with the lack of any spatial information often leads to over classification. There was hardly any undersegmentation noticed in any of the classification images. Figure 7 shows the results of the SVM classification, which resulted in an average DICE overlap ratio for the fat, glandular, and skin tissue of 97% ± 1.2%, 88% ± 6.5%, and 87% ± 5.4%, respectively. The classification was completed for five patients with three slices per patient. The high accuracy for the fat tissue was a promising start for the classification but the large standard deviations in the glandular and skin tissue required further classification refinement. This provided a good framework for accurate marker selection from which the MSF can be rooted.

3.C. Results of 2D minimum spanning forest with probabilistic marker selection

The 2D minimum spanning forest grown from probabilistic marker selection was evaluated and shown to produce accurate classification. The 2D results were evaluated over a range of parameters, including the percentage P of pixels to be selected as markers from large regions. The results from this classification experiment are shown in Fig. 8. It is observed that the optimal number of pixels to be chosen as markers which provides the higher overall and average accuracy is 20% of the large region.

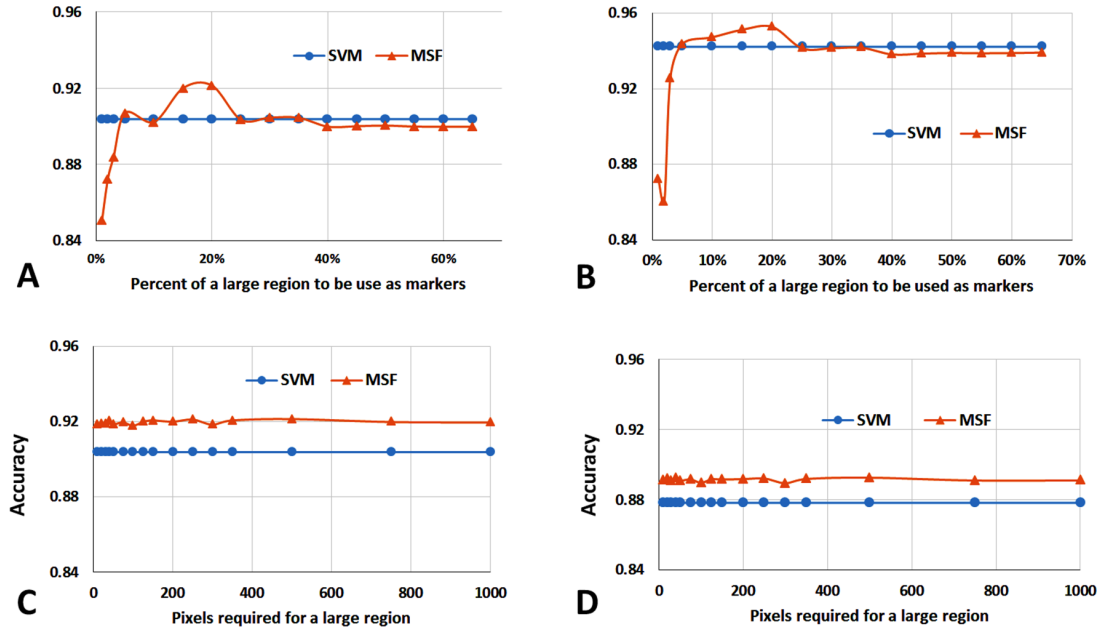

FIG. 8. Accuracy of the minimal spanning forest algorithm with different parameters. The overall accuracy (A) and average accuracy (B) as a function of markers selected from large regions. The overall accuracy (C) and average accuracy (D) shown as a function of pixel numbers required for a region to be classified as a large region.

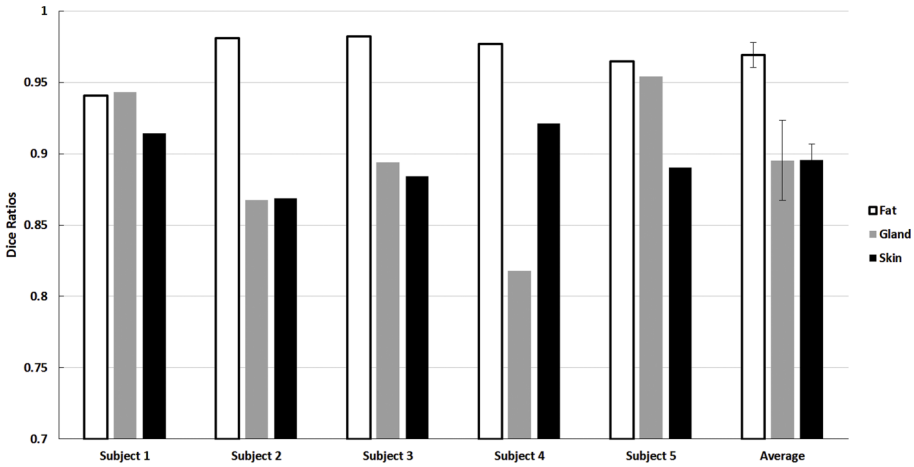

FIG. 9. The DICE overlap ratios for the fat, glandular, and skin tissue when the minimum spanning forest was constructed in 2D using the optimal parameters for the five subjects.

This was the number that was then used in subsequent testing to evaluate the DICE overlap ratios as seen in Fig. 9 for the skin, glandular tissue, and fat tissue. As is seen, an average DICE overlap ratio of 97% ± 1.1%, 90% ± 4.5%, and 90% ± 2.1% was observed while using these optimal parameters. The high accuracy of these results encouraged further evaluation of additional marker selection that could be independent of SVM probabilities and expansion into 3D.

3.D. 2D minimum spanning forest with random marker selection

To reduce the number of parameters and eliminate the need for probability measurements, the random marker selection algorithm was tested and compared to the probabilistic marker selection. The results of this comparison are shown in Figs. 10 and 11.

The overall and average accuracy of the random marker selection shown in Fig. 10 falls short of the accuracy previously obtained by the probabilistic approach as seen in Fig. 8 but is still a viable alternative if probability data are unreliable. The DICE overlap ratios of the random and probabilistic marker selection were compared and shown in Fig. 11. The random marker selection outperformed the probabilistic marker selection for the fat and gland tissue, but did not perform as well as the probabilistic method for the skin tissue. The probabilistic method demonstrates higher sensitivity, specificity, and DICE overlap ratios than the random marker selection with multiple parameter variations to determine the optimal values.

3.E. 2D minimum spanning forest with four nearest neighbors

The 3D CT data increase the number of vertices and edges greatly. Therefore, using pixels or voxels of fewer than the nearest 8 or 26 neighbors was tested. The 4 neighbor approach yielded very similar numbers to using the 8 neighbor approach, suggesting that a 3D minimum spanning forest with less than the equivalent 26 neighbors could be viable. The results of the four nearest neighbors in comparison to the eight nearest neighbors are shown in Fig. 12. The DICE overlap ratios are very close for both methods.

3.F. 3D minimum spanning forest with six nearest neighbors and varying surrounding slices

The minimum spanning forest was applied to 3D images constructed around the slices that were manually segmented. This process was performed for three, five, and seven total slices to reduce computation time while still providing some

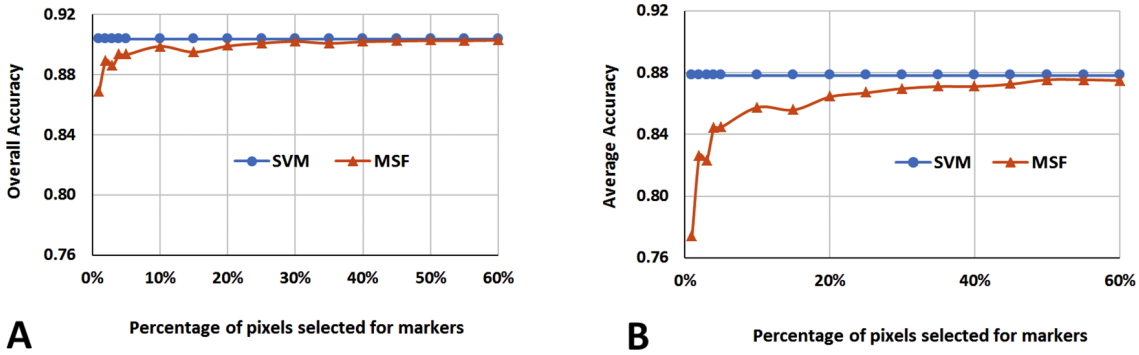

FIG. 10. Overall (A) and average (B) accuracies as a function of percentage of randomly selected pixels to be used as markers in the minimum spanning forest.

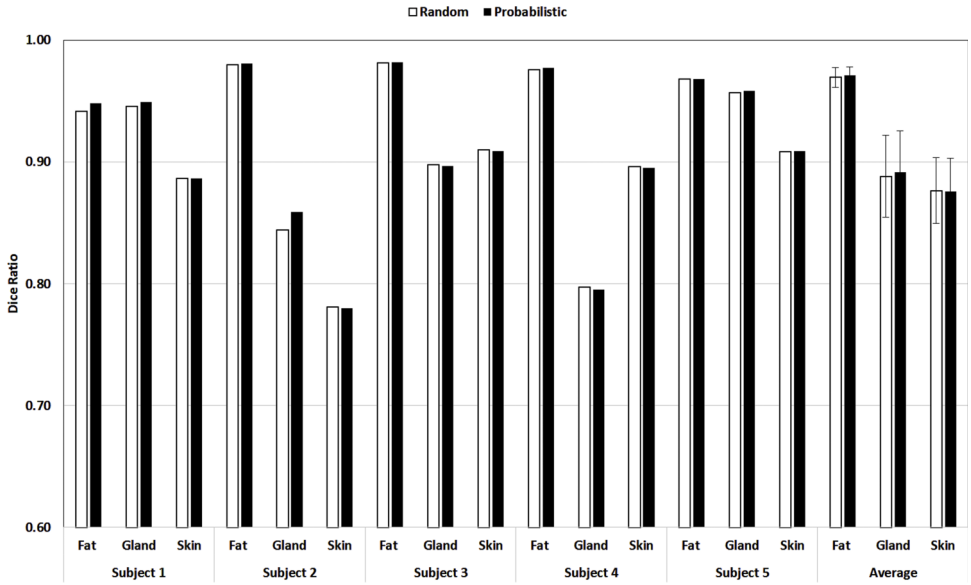

FIG. 11. Comparison of DICE overlap ratios when the two methods are used for marker selection, i.e., randomly selected pixels as markers vs probabilistic marker selection.

3D information. Table I shows the complete results for multiple slices surrounding the slice of interest varying from three to seven slices total. The data provided highly varying results and demonstrated that the limited number of slices was not sufficient to overcome the noise of the images. The additional information from the third dimension did not provide sufficient data for the minimum spanning forest to gain accuracy. A larger amount of slices may be necessary to obtain increased accuracy. However, the addition of more slices is computationally costly.

After comparing different approaches and different parameters, we used the optimal parameters for the classification of the five patient data. As shown in Fig. 13, the average DICE overlap ratios are 96.9%, 89.8%, and 89.5% for the fat, glandular, and skin tissue, respectively, using the 2D image classification which produced the optimal results.

3.G. Classification results compared with previous classification results

The newly developed minimum spanning forest method in 2D produced DICE overlap ratios for the fat, glandular, and skin tissue are 96.9%, 89.8%, and 89.5%, respectively. These results were compared to the previous results given by our laboratory,[12] which produced DICE overlap ratios of 97.4%, 87.7%, and 89.4% for the fat, glandular, and skin tissues, respectively. The improved accuracy for the glandular tissue shows the ability of the minimum spanning forest that grows in a similar fashion as the tissue. The minimum

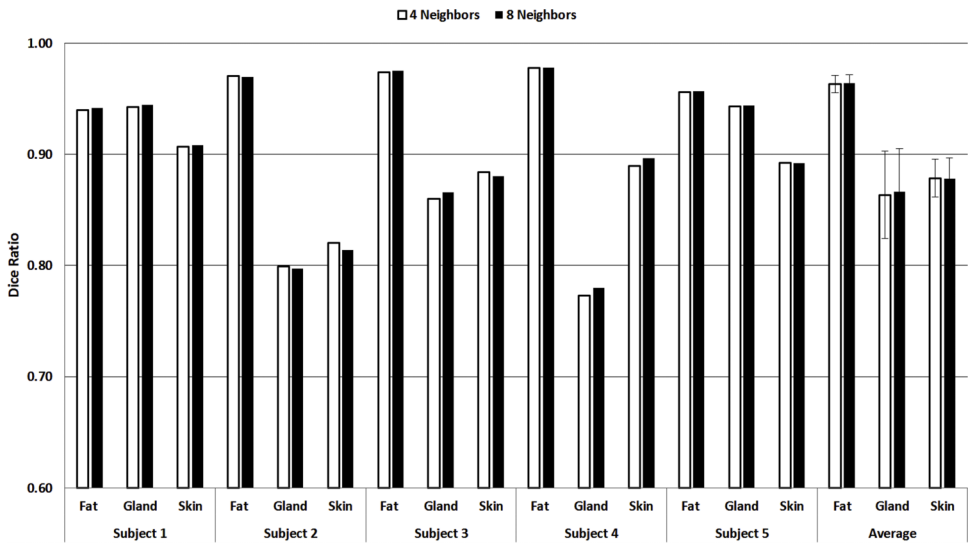

FIG. 12. Comparison between two methods that use the nearest four or eight nearest neighbors to construct the edges for growing the minimum spanning forest.

TABLE I. Dice overlap ratio, sensitivity, and specificity of the 3D minimum spanning forest method when three, five, and seven slices are taken around the gold standard slice.

		Three slices			Five slices			Seven slices		
		Dice	Sensitivity	Specificity	Dice	Sensitivity	Specificity	Dice	Sensitivity	Specificity
Subject 1	Fat	0.928 ± 0.015	0.905 ± 0.031	0.953 ± 0.007	0.930 ± 0.015	0.928 ± 0.037	0.942 ± 0.010	0.940 ± 0.026	0.933 ± 0.042	0.924 ± 0.012
	Gland	0.920 ± 0.014	0.955 ± 0.005	0.912 ± 0.034	0.929 ± 0.011	0.942 ± 0.013	0.931 ± 0.033	0.943 ± 0.006	0.953 ± 0.014	0.909 ± 0.035
	Skin	0.902 ± 0.033	0.947 ± 0.043	0.929 ± 0.014	0.900 ± 0.010	0.944 ± 0.005	0.935 ± 0.014	0.848 ± 0.002	0.777 ± 0.000	0.944 ± 0.017
Subject 2	Fat	0.973 ± 0.001	0.976 ± 0.002	0.765 ± 0.016	0.966 ± 0.001	0.982 ± 0.005	0.665 ± 0.030	0.973 ± 0.001	0.984 ± 0.003	0.717 ± 0.015
	Gland	0.815 ± 0.014	0.723 ± 0.022	0.972 ± 0.002	0.744 ± 0.015	0.602 ± 0.040	0.976 ± 0.005	0.800 ± 0.013	0.686 ± 0.024	0.974 ± 0.004
	Skin	0.764 ± 0.007	0.901 ± 0.001	0.936 ± 0.002	0.741 ± 0.002	0.866 ± 0.002	0.923 ± 0.001	0.744 ± 0.016	0.815 ± 0.015	0.937 ± 0.001
Subject 3	Fat	0.964 ± 0.001	0.979 ± 0.006	0.827 ± 0.013	0.979 ± 0.001	0.993 ± 0.004	0.850 ± 0.020	0.976 ± 0.001	0.984 ± 0.003	0.836 ± 0.016
	Gland	0.830 ± 0.005	0.799 ± 0.019	0.975 ± 0.005	0.854 ± 0.009	0.831 ± 0.028	0.988 ± 0.003	0.862 ± 0.007	0.830 ± 0.022	0.976 ± 0.003
	Skin	0.899 ± 0.001	0.915 ± 0.002	0.947 ± 0.002	0.903 ± 0.000	0.909 ± 0.003	0.964 ± 0.002	0.907 ± 0.002	0.853 ± 0.004	0.957 ± 0.001
Subject 4	Fat	0.977 ± 0.001	0.992 ± 0.003	0.758 ± 0.012	0.973 ± 0.001	0.990 ± 0.002	0.734 ± 0.015	0.978 ± 0.001	0.993 ± 0.003	0.720 ± 0.007
	Gland	0.798 ± 0.011	0.712 ± 0.017	0.979 ± 0.003	0.643 ± 0.015	0.580 ± 0.025	0.990 ± 0.001	0.754 ± 0.025	0.635 ± 0.015	0.983 ± 0.003
	Skin	0.890 ± 0.002	0.835 ± 0.002	0.956 ± 0.001	0.899 ± 0.002	0.988 ± 0.003	0.938 ± 0.001	0.850 ± 0.005	0.860 ± 0.006	0.948 ± 0.003
Subject 5	Fat	0.966 ± 0.001	0.970 ± 0.005	0.936 ± 0.005	0.965 ± 0.004	0.977 ± 0.004	0.940 ± 0.006	0.963 ± 0.002	0.967 ± 0.007	0.935 ± 0.003
	Gland	0.955 ± 0.000	0.955 ± 0.006	0.950 ± 0.004	0.856 ± 0.005	0.941 ± 0.007	0.971 ± 0.003	0.952 ± 0.025	0.948 ± 0.004	0.953 ± 0.006
	Skin	0.895 ± 0.002	0.821 ± 0.004	0.963 ± 0.001	0.908 ± 0.003	0.930 ± 0.002	0.960 ± 0.004	0.905 ± 0.002	0.859 ± 0.004	0.958 ± 0.002

spanning forest takes approximately 3 min/slice to complete the segmentation process, which is faster than the previous method. The increased accuracy of the minimum spanning forest can be attributed to its ability to incorporate global and local information when performing the classification. The global information of average intensity values for specific tissue types is combined with the local information by directly comparing differences in intensity between spatially adjacent nodes. This allows the algorithm to expand and grow in a way that mimics the growth of the actually tissue.

4. DISCUSSION

An automatic image classification method based on minimum spanning forests was proposed and evaluated for breast CT images. The breast image was classified into three distinct tissue types. Comparison of the automatic to manual segmentation shows that the MSF-based classification method was able to successfully classify dedicated breast CT images with an accuracy of 90% for skin and glandular tissue and 97% for fat tissue. This method is able to accurately determine location of each tissue type, as well as quantify tissue types and measure density within the breast.

The proposed algorithm modifies a minimum spanning forest approach to be used with morphological skin segmentation and a support vector machine classification. This approach allowed for normalization of images prior to the support vector machine classification for robust classification. The minimum spanning forest could then grow in 2D or 3D for accurate image classification. The minimal spanning forest method uses pixelwise classification to obtain markers to be used as roots of a minimum spanning forest. The effectiveness of the proposed algorithm comes from its ability to effectively incorporate spatial and intensity based information. The classification method may be applied to other breast images such as tomosynthesis images or other imaging modalities such as optical images, as demonstrated in our preliminary study.[43] Compared with other classification methods, the proposed method allows for variations in threshold limits and

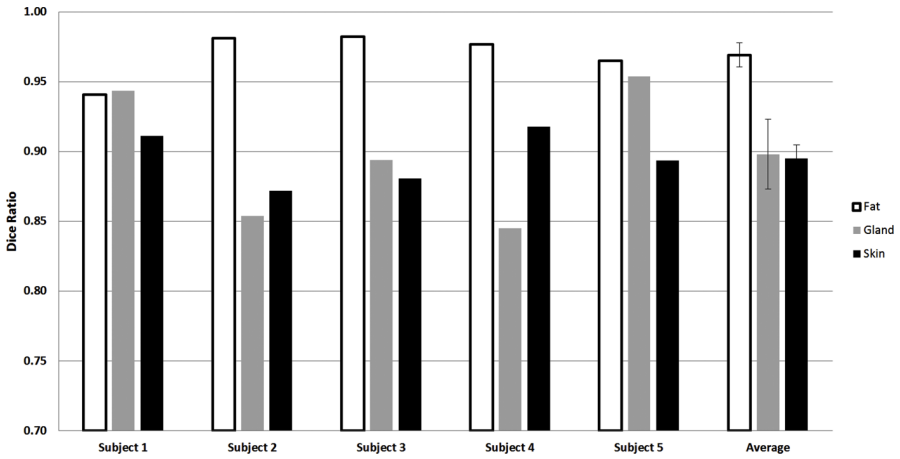

FIG. 13. The final DICE overlap ratios for each individual subject using optimal parameters specific to each patient.

multiple parameters to be tailored specifically to breast CT images.

The 2D minimum spanning forest classification method proved accurate for segmenting fat, skin, and glandular tissue. The classification results relied upon the support vector machine for both classification and marker selection for the minimum spanning forest. This was the largest source of errors between subjects; if the support vector machine misclassified or undersegmented smaller regions, then the minimum spanning forest could not fully correct those errors. Despite these errors, this method was still able to provide an accurate classification of the breast CT images.

An investigation into the difference between the automatic and manual segmentation provides insight into the areas of weakness and strength of the algorithm. The skin tissue is of similar intensity to the glandular tissue and thus was a source of error when glandular tissue lay close to the skin. Manual segmentation is very time consuming, which would ideally be performed over a much broader range of patients to gather more slices of interest. With greater numbers of slices manually segmented in each subject, further investigation of 3D minimum spanning forests could be achieved. In this study, we used manual segmentation as the gold standard because we used the data from human patients. An experienced radiologist manually segmented the tissue from breast CT images. As an alternative for evaluation, a breast phantom could be created using 3D printing technology for accurate evaluation of tissue classification in the future. In the current study, we used simulation data to validate the proposed classification algorithm. This simulation study allows us to not only measure the volume of the tissue but also measure the accuracy of the classification. One limitation of this study was the small sample size of human subjects. Different acquisition systems could be used to test this algorithm on multiple imaging devices and the normalization method could be evaluated across multiple platforms.

The use of a 3D minimum spanning forest was explored and provided mixed results. 3D classification proved computationally expensive and the resulting classification could not make full use of additional information in the third plane. Further testing that includes additional 3D data could prove valuable. The ability to incorporate the third dimension provides valuable information for detecting thin fingerlike growths that would otherwise be considered noise in the other planes. To detect these growths however, a large number of slices must be used. The addition of these slices increases the number of vertices and edges used in the minimum spanning forest construction. This causes a drastic increase in computation time which is the main prohibiting factor of this method. The algorithm was implemented in an Intel Xeon 3.0 GHz processor in a Dell T7400 with 16 GB of memory. Classification time was approximately 3 min/slice. We used MATLAB for the implementation of the algorithm. The code can be further optimized in order to reduce the time. Furthermore, the minimal spanning forest algorithm can be significantly speeded up by using parallel processing, graphic processing unit (GPU), and c++ language.

Future studies that can take advantage of the unique growing process of the minimum spanning forest would be to investigate larger 3D data sets that would allow the algorithm to follow different types of tissue in multiple dimensions. Additional dissimilarity information, including but not limited to, larger scaled intensity information could also lead to more accurate classification. Future studies would include coupling multiple dissimilarity measures between voxels in order to better distinguish tissue type. Registered images from multiple modalities could also be used to improve classification.

5. CONCLUSION

We developed an automatic image classification method for dedicated breast CT images, which couples spatial and intensity information from the images. By normalizing the intensity of images with information gathered in the spatial domain, we were able to train and make use of support vector machines to provide a rough classification of the image. From this SVM classification, root markers were automatically selected and a minimum spanning forest was constructed to follow the various tissue types and complete the classification. Majority voting within the SVM classification of these spanning trees then decided the class of each voxel. The method was shown that it could be a quantification tool to determine breast tissue compositions from dedicated breast CT images.

ACKNOWLEDGMENTS

This research was supported in part by NIH grants (Nos. R01CA156775, R01CA163746, and R21CA176684), Georgia Cancer Coalition Distinguished Clinicians and Scientists Award, Emory Molecular and Translational Imaging Center (No. NIH P50CA128301), SPORE in Head and Neck Cancer (No. NIH P50CA128613), and Atlanta Clinical and Translational Science Institute (ACTSI) that is supported by the PHS Grant No. UL1 RR025008 from the Clinical and Translational Science Award program.

[a]Author to whom correspondence should be addressed. Electronic mail: bfei@emory.edu; Telephone: 404-712-5649; Fax: 404-712-5689; http://www.feilab.org.

[1]A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," Ca-Cancer J. Clin. **61**, 69–90 (2011).

[2]E. Amir, O. C. Freedman, B. Seruga, and D. G. Evans, "Assessing women at high risk of breast cancer: A review of risk assessment models," JNCI, J. Natl. Cancer Inst. **102**, 680–691 (2010).

[3]J. A. Tice and K. Kerlikowske, "Screening and prevention of breast cancer in primary care," Primary Care **36**, 533–558 (2009).

[4]M. Baum and P. M. Ravdin, "Decision-making in early breast cancer: Guidelines and decision tools," Eur. J. Cancer **38**, 745–749 (2002).

[5]M. J. Silverstein, K. A. Skinner, and T. J. Lomis, "Predicting axillary nodal positivity in 2282 patients with breast carcinoma," World J. Surg. **25**, 767–772 (2001).

[6]M. Kalager, M. Zelen, F. Langmark, and H.-O. Adami, "Effect of screening mammography on breast-cancer mortality in Norway," N. Engl. J. Med. **363**, 1203–1210 (2010).

[7]C. H. Chang, D. E. Nesbit, D. R. Fisher, S. L. Fritz, S. J. Dwyer III, A. W. Templeton, F. Lin, and W. R. Jewell, "Computed tomographic

mammography using a conventional body scanner," AJR, Am. J. Roentgenol. **138**, 553–558 (1982).
[8]A. O'Connell, D. L. Conover, Y. Zhang, P. Seifert, W. Logan-Young, C.-F. L. Lin, L. Sahler, and R. Ning, "Cone-beam CT for breast imaging: Radiation dose, breast coverage, and image quality," Am. J. Roentgenol. **195**, 496–509 (2010).
[9]M. J. Yaffe, J. M. Boone, N. Packard, O. Alonzo-Proulx, S. Y. Huang, C. L. Peressotti, A. Al-Mayah, and K. Brock, "The myth of the 50-50 breast," Med. Phys. **36**, 5437–5443 (2009).
[10]S. Vedantham, L. Shi, A. Karellas, and A. M. O'Connell, "Dedicated breast CT: Fibroglandular volume measurements in a diagnostic population," Med. Phys. **39**, 7317–7328 (2012).
[11]H. C. Kuo, M. L. Giger, I. Reiser, J. M. Boone, K. K. Lindfors, K. Yang, and A. Edwards, "Level set segmentation of breast masses in contrast-enhanced dedicated breast CT and evaluation of stopping criteria," J. Digital Imaging **27**, 237–247 (2014).
[12]X. Yang, S. Wu, I. Sechopoulos, and B. Fei, "Cupping artifact correction and automated classification for high-resolution dedicated breast CT images," Med. Phys. **39**, 6397–6406 (2012).
[13]C. C. Chang and C. J. Lin, "Training ν-support vector classifiers: theory and algorithms," Neural Comput. **13**, 2119–2147 (2001).
[14]N. T. Renukadevi and P. Thangaraj, "Optimizing of SVM for CT classification," J. Theor. Appl. Inf. Technol. **55**, 203–208 (2013).
[15]N. H. Anderson, P. W. Hamilton, J. M. Sloan, P. H. Bartels, D. Thompson, and R. Montironi, "Computerized scene segmentation for the discrimination of architectural features in ductal proliferative lesions of the breast," J. Pathol. **181**, 374–380 (1997).
[16]Y. H. Chang, X. H. Wang, L. A. Hardesty, T. S. Chang, W. R. Poller, W. F. Good, and D. Gur, "Computerized assessment of tissue composition on digitized mammograms," Acad. Radiol. **9**, 899–905 (2002).
[17]B. Sahiner, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, J. Wei, C. Zhou, and Y. Lu, "Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: A 3D approach," Med. Phys. **39**, 28–39 (2012).
[18]A. Samani, J. Bishop, and D. B. Plewes, "A constrained modulus reconstruction technique for breast cancer assessment," IEEE Trans. Med. Imaging **20**, 877–885 (2001).
[19]S. Y. Huang, J. M. Boone, K. Yang, A. L. Kwan, and N. J. Packard, "The effect of skin thickness determined using breast CT on mammographic dosimetry," Med. Phys. **35**, 1199–1206 (2008).
[20]L. Shi, S. Vedantham, A. Karellas, and A. M. O'Connell, "Technical Note: Skin thickness measurements using high-resolution flat-panel cone-beam dedicated breast CT," Med. Phys. **40**, 031913 (6pp.) (2013).
[21]F. Meyer, "Minimum spanning forests for morphological segmentation," in *Mathematical Morphology Its Applications to Image Processing*, edited by J. Serra and P. Soille (Springer, Netherlands, 1994), Vol. 2, pp. 77–84.
[22]Y. Xu and E. C. Uberbacher, "2D image segmentation using minimum spanning trees," Image Vision Comput. **15**, 47–57 (1997).
[23]M. Abdel-Mottaleb and A. Elgammal, "Face detection in complex environments from color images," in *Proceedings 1999 International Conference on Image Processing, Vol. 3* (IEEE, Washington, DC, 1999), Vol. 623, pp. 622–626.
[24]Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," IEEE Trans. Syst., Man, Cybern., Part B (Cybern.) **40**, 1267–1279 (2010).
[25]X.-Y. Wang, T. Wang, and J. Bu, "Color image segmentation using pixel wise support vector machine classification," Pattern Recognit. **44**, 777–787 (2011).
[26]O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," IEEE Trans. Neural Networks **10**, 1055–1064 (1999).
[27]A. A. Farag, R. M. Mohamed, and A. El-Baz, "A unified framework for MAP estimation in remote sensing image segmentation," IEEE Trans. Geosci. Remote Sens. **43**, 1617–1634 (2005).
[28]M. Fauvel, "Spectral and spatial methods for the classification of urban remote sensing data," Thèse de Doctorat, Institut Technologique de Grenoble–Université d'Islande, 2007.
[29]M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," IEEE Trans. Geosci. Remote Sens. **46**, 3804–3814 (2008).
[30]R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," IEEE Trans. Geosci. Remote Sens. **14**, 19–26 (1976).
[31]R. Gaetano, G. Moser, G. Poggi, G. Scarpa, and S. B. Serpico, "Region-based classification of multisensor optical-SAR images," *IGARSS 2008–2008 IEEE International Geoscience and Remote Sensing Symposium* (IEEE, Washington, DC, 2008), pp. IV-81–IV-84.
[32]K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," IEEE Trans. Image Process. **21**, 2008–2021 (2012).
[33]S. A. Willson, E. J. Adam, and A. K. Tucker, "Patterns of breast skin thickness in normal mammograms," Clin. Radiol. **33**, 691–693 (2001).
[34]C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol. **2**, 1–27 (2011).
[35]S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," Neural Networks **12**, 783–789 (1999).
[36]J. Stawiaski, "Mathematical morphology and graphs: Application to interactive medical image segmentation," Ph.D. dissertation, Paris School Mines, Paris, France, 2008.
[37]J. B. Kruskal, Jr., "On the shortest spanning subtree of a graph and the traveling salesman problem," Proc. Am. Math. Soc. **7**, 48–50 (1956).
[38]H. Gabow, Z. Galil, T. Spencer, and R. Tarjan, "Efficient algorithms for finding minimum spanning trees in undirected and directed graphs," Combinatorica **6**, 109–122 (1986).
[39]R. C. Prim, "Shortest connection networks and some generalizations," Bell Syst. Tech. J. **36**, 1389–1401 (1957).
[40]B. E. Moret and H. Shapiro, "An empirical analysis of algorithms for constructing a minimum spanning tree," in *Algorithms and Data Structures*, edited by F. Dehne, J.-R. Sack, and N. Santoro (Springer, Berlin, Heidelberg, 1991), Vol. 519, pp. 400–411.
[41]T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3rd ed. (The MIT Press, Cambridge, MA, 2009).
[42]I. Sechopoulos, S. S. Feng, and C. J. D'Orsi, "Dosimetric characterization of a dedicated breast computed tomography clinical prototype," Med. Phys. **37**, 4110–4120 (2010).
[43]R. Pike, S. K. Patton, G. Lu, L. V. Halig, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest based hyperspectral image classification method for cancerous tissue detection," Proc. SPIE **9034**, 90341W (2014).