Quantification of Atherosclerotic Plaque Components Using In Vivo MRI and Supervised Classifiers

J.M.A. Hofman,1* W.J. Branderhorst,1 H.M.M. ten Eikelder,1 V.C. Cappendijk,2 S. Heeneman,3 M.E. Kooi,2 P.A.J. Hilbers,1 and B.M. ter Haar Romeny1,2

In this work we aimed to study the possibility of using supervised classifiers to quantify the main components of carotid atherosclerotic plaque in vivo on the basis of multisequence MRI data. MRI data consisting of five MR weightings were obtained from 25 symptomatic subjects. Histological micrographs of endarterectomy specimens from the 25 carotids were used as a standard of reference for training and evaluation. The set of subjects was divided in a training set (12 subjects) and an evaluation set (13 subjects). Four different classifiers and two human MRI readers determined the percentages of calcified tissue, fibrous tissue, lipid core, and intraplaque hemorrhage on the subject level for all subjects in the evaluation set. Quantification of the relatively small amounts of calcium could not be done with statistical significance by either the classifiers or the MRI readers. For the other tissues a simple Bayesian classifier (Bayes) performed better than the other classifiers and the MRI readers. All classifiers performed better than the MRI readers in quantifying the sum of hemorrhage and lipid proportions. The MRI readers overestimated the hemorrhage proportions and tended to underestimate the lipid proportions. In conclusion, this pilot study demonstrates the benefits of algorithmic classifiers for quantifying plaque components.

MATERIALS AND METHODS

Methodological Overview

MRI data sets were obtained from 25 subjects with a carotid stenosis of >70% as diagnosed by duplex ultrasonography. All of the subjects had been scheduled for carotid endarterectomy and had experienced transient ischemic attacks or minor strokes within 3 months before surgery. MRI scans of the carotid arteries were performed within 2 weeks before surgery. Histological micrographs were available for these 25 carotid samples. This study was approved by the Institutional Review Committee of Maastricht University Hospital, and all subjects gave written informed consent.

In the present study several supervised classifiers were used to assign pixels in the MRI to tissue types: calcified...
tissue, fibrous tissue, intraplaque hemorrhage, and lipid core. Toward that end the set of MR images obtained from 25 subjects was split into a set that was used for training the classifiers and a set that was used for evaluation. The training set consisted of the MRI data from 12 subjects, and the evaluation set consisted of the data from 13 subjects. Corresponding histological micrographs served as a standard of reference for creating the training data and evaluating the classifiers, as discussed in the following subsections.

The proportions of the four tissue types in the used training set can have a large influence on the performance of the classifiers. For the Bayesian classifier this follows immediately from the Bayes formula, but the effect also exists for other classifiers. Therefore, the proportions of the tissue types should agree with the proportions faced in practice, viz., the actual priors of the tissue types. Hence, it is important to estimate these proportions from a sufficiently large group of representative subjects. In principle the training set (12 subjects) can be used for this purpose. However, additional histological data were available from another group of similar subjects. This allowed us to estimate the proportions on the basis of 37 subjects in total. The group of 37 subjects was separate from the set of 13 evaluation subjects.

In Vivo MRI

The MR scans were performed on a 1.5-T whole-body imager (Intera 8.1.2; Philips Medical Systems, Best, The Netherlands) with a maximum gradient strength of 23 mT/m and a maximum slew rate of 104 mT/m/s. A radiofrequency (RF) surface coil with a diameter of 47 mm was fixed to the skin just above the carotid bifurcation. The subjects were positioned in a head holder to reduce motion artifacts. They were asked to keep their head in a sideways position so that the positioning of the coil was not restricted by the jaw. In our experience, the sideways position of the head enables an optimal positioning of the coil with respect to the carotid bifurcation, and results in an optimal signal-to-noise ratio (SNR).

The location of the carotid bifurcation was determined by MR angiography (MRA) using a balanced 3D turbo field echo (TFE) pulse sequence with the following imaging parameters: repetition time (TR)/echo time (TE) = 4.4/2.1 ms, flip angle = 60°, field of view (FOV) = 200 * 200 mm, matrix size = 192 * 192, four signals acquired, section thickness = 0.70 mm over contiguous transverse sections, body coil, and imaging duration = 1 min 27 s. After MRA was performed the surface coil was occasionally repositioned to achieve an optimal SNR depending on the location of the carotid plaque. Afterwards, nine transverse slices of 3 mm thickness were scanned at locations from about 7 mm caudal to 2 cm cranial of the carotid bifurcation using the small surface coil as a receiver coil. This range covered the complete carotid plaque for all patients.

For each of the nine slices the following five MR weightings were acquired:

- (w1) T₁-weighted 2D turbo spin echo (TSE) double-inversion recovery black blood, as described by Edelman et al. (25), with TR/TE/inversion time (TI) = 570/14/255 ms, echo train length (ETL) = 5, two signals acquired, and imaging duration = 7 min 5 s.
- (w2) electrocardiographically (ECG)-gated proton density-weighted turbo SE (TSE), with TR = two heartbeats (R-R), TE = 20 ms, ETL = 5, two signals acquired, and imaging duration = ~3 min. To reduce the signal from flowing blood, superior and inferior spatial presaturation slabs were used.
- (w3) T₁-weighted 3D TFE, with TR/TE/TI = 10.3/4.0/900 ms, flip angle = 15°, 24 shots, shot interval time = 3000 ms, actual imaging percentage = 80%, 163 phase-encoding steps per 24 shots, six signals acquired, and imaging duration = 3 min 37 s. The TI of 900 ms was chosen for optimal blood nulling.
- (w4) ECG-gated partial T₂-weighted TSE, with TR = two heartbeats (R-R), TE = 30 ms, ETL = 9, four signals acquired, superior and inferior spatial presaturation slabs, and imaging duration = ~1.5 min; and
- (w5) ECG-gated T₂-weighted TSE, with TR = two heartbeats (R-R), TE = 50 ms, ETL = 8, two signals acquired, superior and inferior spatial presaturation slabs, and imaging duration = ~3.5 min.

Histological Evaluation

Carotid endarterectomy samples were used as a standard of reference to evaluate the MRI-based classification of the plaque. For this purpose the surgeons removed the carotid endarterectomy specimens in one piece. The tissue was mildly decalcified using the Kristensen protocol. This protocol makes calcium salts soluble (and removes them upon rinsing), does not affect the softer tissues, and facilitates the cutting of the specimens. (De)calcified and/or ossified structures are still detectable with hematoxylin-eosin (H&E) staining.

The specimens were fixed with formalin and then divided into transverse slices of 3 mm thickness. On average, seven slices were obtained per subject, covering the whole plaque region. After these slices were embedded in paraffin, one histological section of 4 µm thickness was taken out of every slice. These histological sections were subjected to H&E and Elastica von Gieson staining. Color prints were made of the histological slices. Two extensively trained investigators examined the histological slices microscopically, and for each slice they segmented the plaque into the major components: fibrous tissue, lipid core, calcified tissue, and intraplaque hemorrhage.

In intraplaque hemorrhage, intact and hemolyzed erythrocytes mixed with sleeves of fibrin had to be present. Additionally there had to be evidence of (early) signs of organization such as ingrowth of (myo)fibroblasts, capillaries, and infiltration of macrophages and/or inflammatory cells. The investigators were unable to reach consensus for approximately 6% of the slices, and for these slices a pathologist determined the final segmentation of the plaque. The investigators and the pathologist were unaware of the classification results obtained on the basis of MRI. For each histological slice the areas of the separate plaque components were measured relative to the area of the vessel wall. These measurements were performed by using a square grid of lines printed on a transparent sheet that was superimposed on the histological color print.
Coregistration of MRI With Histology

The MR slices obtained from the 25 subjects were coregistered with the histological sections by one investigator. Recall that for each subject there were on average seven histological sections and nine MRI slices. Hence, on average two MRI slices per subject were omitted. In all cases the omitted slices were at the farthest cranial or caudal positions. Consequently, on average seven MRI slices were coregistered with histology per subject.

The carotid bifurcation and/or the narrowest lumen were used as landmarks in the longitudinal direction. The cross-sectional shape of the plaque was used to adjust the cross-sectional orientation of the slices. For the slices included in the evaluation set (13 subjects) the coregistration was repeated by a second investigator. In a few cases the investigators found a different coregistration result. The largest difference found was a relative translation in the longitudinal direction of one slice thickness in the MRI (3 mm). In these cases the images were submitted to a third investigator, who made a decision concerning the best coregistration.

Evaluation by MRI Readers

Two MRI readers examined the MRI slices in the evaluation set independently. They were blinded to the histological and classifier results. The slices (five MR weightings each) were presented to the readers in random order. The MRI readers scored the relative signal intensity (rSI) as very low, low, equal, or high compared to the surrounding muscle tissue of the maximal three most distinguishable (in terms of SI) ROIs per MRI slice, together covering the complete vessel wall. Note that an MRI reader could assign one or more than three ROIs that were not joined together to one ROI.

For classification the MRI readers inspected all five contrast weightings, as described by Cappendijk et al. (22). The segmentation of the artery wall region was performed by drawing subregions on a standard oval grid (on bifurcation level) or a standard circular grid (on all other levels). By counting squares of the grid, the areas covered by the different tissues were determined relative to the total area of the vessel wall.

Preprocessing of MR Images

The MR images for all 25 subjects were preprocessed for the purpose of automatic classification. Preprocessing was performed by using a dedicated program that supports user interactions with graphical data. This program was developed in Mathematica®. All processing steps were completed by agreement of at least two investigators. For each slice all five MRI weightings were processed as follows:

1. Manual selection of subimages, including the internal carotid artery.
2. Automatic coregistration of the five weightings by searching for 2D translations and rotations that optimize the normalized mutual information (26).
3. Manual delineation of two regions of muscle tissue (one located laterally and one located medially from the internal carotid artery). The direction from lateral to medial is called the x-direction.
4. For each pixel the rSI was determined by dividing the pixels' intensity by the local intensity of muscle tissue. This local muscle-tissue intensity was determined by linear interpolation of the average intensities of the two muscle-tissue regions. Interpolation was done in the x-direction only. The linear interpolation is reasonable because both regions of muscle tissue are close to the internal carotid artery, and hence the difference between the actual local muscle tissue intensity and its linear approximation will be small. rSI is introduced in this way to partially correct for differences in coil load and the decay of the sensitivity of the surface coil as a function of distance. Using intensities of two muscle regions visually improved the correction for spatial sensitivity decay.

After the common preprocessing of all available MR images was performed, the images included in the evaluation and training sets were subjected to additional processing steps to obtain evaluation images and training data.

Additional Preprocessing of Evaluation Images

As mentioned above, about seven MRI slices were coregistered with histology per subject. For the 13 evaluation subjects a total of 89 slices were coregistered with histology. Some of the MRI slices in the evaluation set were excluded for the following reasons:

1. The internal carotid artery could not be distinguished entirely, due to noise (N = 7).
2. Motion artifacts were clearly seen in regions that partially overlapped the internal carotid artery (N = 6).
3. Not all five contrast weightings were available. Since the weighting w1 (T1wTSE) required a relatively long scan time, this weighting was sometimes acquired for only five locations to diminish motion artifacts (N = 14).

After this selection was completed the evaluation set consisted of 62 multisequence MR slices obtained from 13 subjects.
For all 62 slices in the evaluation set, the outer boundary of the internal carotid artery and the boundary of the lumen were delineated by the following procedure: The five contrast weightings for a given slice were presented on a horizontal row in the image processing application that was written in Mathematica®. This application enabled the user to draw a colored contour that was shown in all five MR-weighted images identically. The correctness of these contours was always verified by examining the five images with and without superimposed contours. If necessary, the outlines were corrected to ensure that only artery wall tissues were enclosed by the outlines of the lumen and outer boundary (Fig. 1).

Collection of Training Data

The MR images in the training set were examined by two investigators. For each type of tissue (fibrous tissue, lipid core, calcified tissue, or intraplaque hemorrhage) they selected a large number of pixels taken from homogeneous regions of interest (ROIs) in the MR images. They obtained training ROIs by first looking for relatively large patches of tissues in the histology images, while restricting themselves to patches that could easily be matched to homogeneous regions in the corresponding MRI images. For any tissue patch, a training ROI was selected as a smaller region near the center of the patch in order to avoid partial volume effects.

Each pixel in a ROI yields five rSI values and its corresponding tissue type. Parameters were estimated by the maximum-likelihood method.

2. k-Nearest neighbor (k-NN) classifier. The input data for this classifier were only the rSI values of the three MR weightings w1, w3, and w5. This choice will be explained in the next subsection.

3. A feed-forward neural network (NN) with one hidden layer and four output neurons (one for every tissue type).

4. Bayes2, a combination of the Bayes classifier with a Parzen classifier (24) that takes into account the spatial context of each pixel. The objective of this modification was to reduce the effect of noise in the MR images and thus produce more homogeneous regions in the classified images with fewer misclassified pixels.

Note that types 1–3 are pixel-based classifiers, i.e., they determine the tissue type of a pixel purely on the basis of the rSI values of that pixel. The principles of these classifiers are described in standard textbooks on pattern classification (24). The last classifier, Bayes2, is not pixel based. Bayes2 also uses information about neighboring pixels to find the type of the current pixel. This algorithm consists of the following steps:

1. Use the Bayes classifier to compute the four probabilities (posteriors) that a given pixel belongs to any of the four tissue types.

Classification Algorithms

In our study the following types of supervised classifiers were used:

1. Bayes, a standard Bayesian classifier that assumes a multivariate normal distribution for each class (tissue type). Parameters were estimated by the maximum-likelihood method.

2. k-Nearest neighbor (k-NN) classifier. The input data for this classifier were only the rSI values of the three MR weightings w1, w3, and w5. This choice will be explained in the next subsection.

3. A feed-forward neural network (NN) with one hidden layer and four output neurons (one for every tissue type).

4. Bayes2, a combination of the Bayes classifier with a Parzen classifier (24) that takes into account the spatial context of each pixel. The objective of this modification was to reduce the effect of noise in the MR images and thus produce more homogeneous regions in the classified images with fewer misclassified pixels.

Note that types 1–3 are pixel-based classifiers, i.e., they determine the tissue type of a pixel purely on the basis of the rSI values of that pixel. The principles of these classifiers are described in standard textbooks on pattern classification (24). The last classifier, Bayes2, is not pixel based. Bayes2 also uses information about neighboring pixels to find the type of the current pixel. This algorithm consists of the following steps:

1. Use the Bayes classifier to compute the four probabilities (posteriors) that a given pixel belongs to any of the four tissue types.
2. Assign each pixel to the tissue type with the highest posterior. This assignment results into four binary images, one for each tissue type. Pixels that belong to a given tissue type have pixel values equal to 1 (otherwise 0).

3. Calculate alternative posteriors by using a Parzen classifier operating in the spatial domain. This is done by convolving each of the four binary images with a Gaussian kernel. The width of the Gaussian kernel is one pixel.

4. Compute combined posteriors by using the posteriors determined in steps 1 and 3. The combination is done by application of the sum rule (27).

5. Repeat steps 2 (with the combined posteriors as input), 3, and 4 until the tissue types assigned to all pixels do not change any more.

In this way the classification on the basis of rSI values only (the spectral information incorporated by step 1) is modified by using the spatial context.

Training the Classifiers

The classifiers were implemented in Matlab® and Mathematica® codes that were specifically dedicated to the classification of image data. The parameter $k$ of the k-NN classifier was determined by using the leave-one-out method. For a training set of size $N$, this means that one pixel was taken out of the training set and the classifier based on the remaining $N-1$ pixels was used to predict the class of the unseen pixel. This procedure was repeated $N$ times by leaving out each pixel in the training set once. The total number of left-out pixels that were wrongly classified in this way was used as an error measure to determine the optimum value of $k$. The same method was used to determine the optimal number $h$ of neurons in the hidden layer of the NN. In this case the NN was trained using the remaining $N-1$ pixels with the Levenberg-Marquardt method. Moreover, a random part of the set of $N-1$ pixels was used as validation set to prevent overtraining (24). With the values of $k$ and $h$ found in this way, the final k-NN and NN classifiers, respectively, were constructed, now using all pixels in the training set.

All five MRI weightings were used to train the classifiers Bayes, Bayes2, and NN. The optimum number of neurons in the hidden layer of the NN was $h = 5$. Note that our training set consisted of 1811 points. This was large enough for our NN, which with five neurons in the hidden layer has 54 parameters.
The parameters of the Bayesian classifier are the means and coveriances of the data corresponding to the four tissue types. The estimates given by Kalayeh and Landgrebe (28) also show that for this classifier our training set was sufficiently large. Moreover, computing the means and coveriances from the original 6286 data points yielded almost identical results.

For the k-NN classifier the situation is different, since it may be difficult to use this type of classifier with only 1811 training points in a 5D space. This difficulty can be overcome by reducing the dimensionality. To reduce the dimensionality, we plotted the average rSI values of the four tissue types in the training set for each MR weighting (Fig. 2). For a given tissue type and contrast weighting the rSI values show large variations, as indicated by the SD (vertical bars). For all weightings there is a large overlap between the rSI values for hemorrhage and lipid core. On the basis of Fig. 2, we decided to train the k-NN classifier by using only the weightings w1, w3, and w5. With these inputs the optimal value of k in the k-NN classifier was k = 3.

Evaluation of Classifiers and MRI Readers

Each of the aforementioned classifiers was evaluated as follows: The preprocessed MRI slices included in the evaluation set were used as input for the classifiers. For each slice only the pixels in the artery wall region were classified. These pixels were assigned to one of the classes: calcified, fibrous, hemorrhage, and lipid core. For each slice the relative area covered by the pixels of a given class was found by the NN. The proportions of the tissue types were determined on the subject level, as explained in the Materials and Meth-
ods section. These proportions (expressed as percentages of the artery wall) were plotted against the histology-based results for all classifiers used. As an example, the results for the Bayes classifier are shown in Fig. 4 for calcified (a), fibrous (b), hemorrhage (c), and lipid core (d). The results for these tissues are clearly positively correlated.

The performance of the classifiers (and MRI readers) was measured by performing linear regression $y = ax + b$ of the histology (y) against the classifier (x). The results including the correlation coefficient $r$ are given in Tables 2–4. Note that in the ideal case $a = 1$, $b = 0\%$, and $r = 1$.

The inter-rater agreement of the MRI readers was good for fibrous tissue and hemorrhage, but it was moderate for calcium and poor for lipid core (Table 5). Therefore, in Tables 2–4 the coefficients of regression and correlation are given for the MRI readers separately, in addition to the performance measures for the averaged readers results.

The correlations found for calcified tissue were small for all classifiers, as well as for the MRI readers (Table 2). This may be due to the fact that histology-based calcium proportions were less than 5% for 11 of 13 subjects, and thus small random errors would have a relatively large effect on correlation. Calcium proportions were relatively large (about 30%) for two subjects, but the classifiers and the MRI readers were unable to find these high proportions (cf., the results for the Bayes classifier shown in Fig. 4a).

For fibrous tissue, the classifiers Bayes, Bayes2, and k-NN performed reasonably, while the NN performed moderately and the MRI readers performed poorly (Table 3). The performance measures for the proportions of high-risk component (hemorrhage plus lipid proportions) are given in Table 4. The classifiers Bayes, k-NN, and Bayes2 showed a good correlation with histology, while the NN performed moderately and the MRI readers showed only a poor performance. The correlations obtained for the high-risk component are significant on the $P = 0.05$ level for all classifiers as well as for the MRI readers. For lipid core alone, significance on the $P = 0.05$ level was reached only for the classifiers Bayes and Bayes2, as shown in Table 3.

The extension of the simple Bayesian classifier to the classifier Bayes2, which also uses information of neighboring pixels, did not lead to significantly better results (Tables 2–4). As an example, Fig. 5 shows the artery wall region of one MR section in which the classified pixels were obtained by the Bayes (a) and Bayes2 (b) classifiers. Here colors indicate the calcified tissue (green), fibrous tissue (blue), hemorrhage (red), and lipid core (yellow). The classification found by the Bayes2 classifier is somewhat less noisy than the original pixel-based Bayes classifier. As noted above, this did not lead to a larger correlation on the subject level.

Table 4

<table>
<thead>
<tr>
<th>Classifier</th>
<th>High-risk component</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (SD)</td>
<td>b/% (SD/%)</td>
</tr>
<tr>
<td>Bayes</td>
<td>0.8 (0.2)</td>
</tr>
<tr>
<td>k-NN</td>
<td>0.8 (0.2)</td>
</tr>
<tr>
<td>Neural network</td>
<td>0.8 (0.3)</td>
</tr>
<tr>
<td>Bayes2</td>
<td>0.8 (0.2)</td>
</tr>
<tr>
<td>Reader 1</td>
<td>0.5 (0.2)</td>
</tr>
<tr>
<td>Reader 2</td>
<td>0.5 (0.2)</td>
</tr>
<tr>
<td>Readers (average)</td>
<td>0.5 (0.2)</td>
</tr>
</tbody>
</table>

*Results are given as in Table 2.

Evaluation of Classifiers and MRI Readers Using Bland-Altman Plots

Tissue proportions as determined by the Bayes classifier were not biased with respect to histology, as can be seen in the Bland-Altman plots in Fig. 6, where the subfigures (a–d) correspond to calcified tissue, fibrous tissue, intraplaque hemorrhage, and lipid core, respectively. The differences between the results found by the Bayes classifier and histology (vertical axes) are quite large compared to the means of these results (horizontal axes). To measure these differences the limits of the 95% confidence interval

Table 5

<table>
<thead>
<tr>
<th></th>
<th>a (SD)</th>
<th>b/% (SD/%)</th>
<th>r (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>0.7 (0.3)</td>
<td>1 (1)</td>
<td>0.532 (0.061)</td>
</tr>
<tr>
<td>Fibrous</td>
<td>0.8 (0.2)</td>
<td>20 (10)</td>
<td>0.777 (0.002)</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>1.0 (0.2)</td>
<td>−6 (7)</td>
<td>0.876 (0.000)</td>
</tr>
<tr>
<td>Lipid core</td>
<td>0.2 (0.1)</td>
<td>0.2 (0.8)</td>
<td>0.491 (0.088)</td>
</tr>
<tr>
<td>High-risk component</td>
<td>0.9 (0.2)</td>
<td>−6 (8)</td>
<td>0.824 (0.001)</td>
</tr>
</tbody>
</table>

*Results are on the subject level.
are shown (upper and lower horizontal lines). For the results shown in Fig. 6 the bias (viz., the average value of the differences (middle horizontal lines)) is not significantly different from zero because zero is within the 95% CI for the bias values (not shown in Fig. 6). All classifiers were evaluated in this way. The results of these evaluations are summarized by Table 6 (for calcified and fibrous tissue) and Table 7 (for hemorrhage, lipid core, and the high-risk component). The results found by the MRI readers on the basis of the same 13 subjects are also given in these tables.

In Tables 6 and 7 biased results are shown in boldface. For calcified tissue no significant bias was found for all evaluation methods used (classifiers and MRI readers), as shown in Table 6. Moreover, random deviations with respect to histology are almost independent from the evaluation method (limits of ±95% CI are approximately ±18%).

One of the MRI readers (reader 2) slightly underestimated the fibrous proportions, and both MRI readers showed random deviations that were larger than those found for the classifiers. Both MRI readers overestimated the hemorrhage proportions (Table 7), while the NN underestimated the hemorrhage proportions. For hemorrhage, the random deviations with respect to histology are almost independent from the evaluation method (limits of ±95% CI are approximately ±26%).

The MRI readers slightly underestimated the lipid core proportions, whereas the classifiers were not significantly biased (Table 7). For lipid core, the random deviations were smallest for the Bayes and Bayes2 classifiers.

As shown in Table 7, the proportions of the high-risk component were unbiased for the classifiers Bayes, k-NN, and NN, while Bayes2 was slightly biased and MRI reader 2 showed a somewhat larger bias. The proportions of high-risk component found by the MRI readers showed quite large random deviations from the histology-based proportions as compared to the random deviations scored by the classifiers.

**DISCUSSION**

The purpose of this study was to evaluate the efficacy of supervised classifiers for quantifying the main plaque tissue types in carotid arteries on the basis of multisquence MRI. To our knowledge, no previous work has discussed
the combination of supervised classifiers with in vivo MRI of atherosclerotic plaque. The performance of several of supervised classifiers was studied using histology as a standard of reference. All subjects involved in the present study were symptomatic, and the relative numbers of samples (priors) of the four tissue types in the training set were determined from a group of 37 symptomatic patients. In principle, the classifiers could also be used for plaque quantification of asymptomatic subjects, as in epidemiological studies and clinical trials. However, in such cases it is better to use the priors based on statistics of asymptomatic subjects, since the priors may affect the performance of the classifiers.

Our overall conclusion is that the best classifier performed somewhat better than the MRI readers. For calcified tissue both the MRI readers and the various classifiers could not obtain a significant correlation with histology. This may be due to the small calcium proportions and consequently the large relative errors obtained. For fibrous tissue all classifiers yielded significantly better results than the MRI readers (see Table 2). Although the MRI readers showed a good correlation for hemorrhage, they overestimated hemorrhage significantly, as can be seen from the Bland-Altman results. Moreover, the NN significantly underestimated hemorrhage, while Bayes and Bayes2 performed reasonably for hemorrhage. For lipid core the classifiers Bayes and Bayes2 performed better than the MRI readers, who underestimated the lipid core the classifiers Bayes and Bayes2 performed reasonably for hemorrhage. For lipid core the random deviations from histology were smaller for the classifiers than for the MRI readers, although this was found only for fibrous tissue and the high-risk component. For calcified tissue, hemorrhage, and lipid core the random deviations measured for the classifiers were comparable to those obtained by the MRI readers.

Concerning the training process, it is noted that the classifiers were trained on data obtained by manual selection of pixels within small homogeneous regions taken from MR images of multiple subjects. These regions were chosen on the basis of corresponding histological color prints. In addition, the training data set was tuned by using the priors known from histology of 37 subjects. It is unlikely that humans could be trained with this kind of detailed information, and thus one would expect the classifiers to outperform the readers. On the other hand, all classifiers were based on training data consisting of separate pixels. The only classifier that uses spatial context is Bayes2, but only as a postprocessing step that is performed when this classifier is evaluated on previously unseen images. None of the classifiers is trained by using spatial context information. Human MRI readers are capable of remembering and applying spatial context in images on different levels of complexity, e.g., they can mentally form a model of the carotid wall, including a lipid core adjacent to a fibrous cap.

### Table 6
Summary of Bland-Altman Plots for Calcium and Fibrous Tissue as Found by Supervised Classifiers and MRI Readers*

<table>
<thead>
<tr>
<th>Tissue type</th>
<th>Calcium</th>
<th></th>
<th>Fibrous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (95% CI)</td>
<td>Limits of 95% CI</td>
<td>Bias (95% CI)</td>
<td>Limits of 95% CI</td>
</tr>
<tr>
<td>Bayes</td>
<td>-1% (-7%, 4%)</td>
<td>±19%</td>
<td>-4% (-14%, 5%)</td>
<td>±30%</td>
</tr>
<tr>
<td>k-NN</td>
<td>0% (-6%, 6%)</td>
<td>±19%</td>
<td>1% (-7%, 9%)</td>
<td>±28%</td>
</tr>
<tr>
<td>NN</td>
<td>0% (-6%, 6%)</td>
<td>±19%</td>
<td>3% (-7%, 13%)</td>
<td>±31%</td>
</tr>
<tr>
<td>Bayes2</td>
<td>-1% (-7%, 5%)</td>
<td>±18%</td>
<td>-7% (-15%, 2%)</td>
<td>±28%</td>
</tr>
<tr>
<td>Reader 1</td>
<td>-5% (-10%, 1%)</td>
<td>±18%</td>
<td>0% (-13%, 12%)</td>
<td>±40%</td>
</tr>
<tr>
<td>Reader 2</td>
<td>-4% (-10%, 1%)</td>
<td>±18%</td>
<td>-12% (-25%, 0%)</td>
<td>±40%</td>
</tr>
<tr>
<td>Readers (average)</td>
<td>-5% (-10%, 1%)</td>
<td>±18%</td>
<td>-6% (-18%, 5%)</td>
<td>±37%</td>
</tr>
</tbody>
</table>

*Results are based on MRI data obtained from 13 subjects. Histology is used as the standard of reference to determine the bias (and its 95% CI), and the limits of the 95% CI for the differences. All stated 95% confidence intervals are two-tailed. A significantly positive or negative bias is typed in boldface.

### Table 7
Evaluation of Results for Hemorrhage, Lipid Core, and High-Risk Component*

<table>
<thead>
<tr>
<th>Tissue type</th>
<th>Hemorrhage</th>
<th>Lipid core</th>
<th>High-risk component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (95% CI)</td>
<td>Limits of 95% CI</td>
<td>Bias (95% CI)</td>
</tr>
<tr>
<td>Bayes</td>
<td>3% (-5%, 11%)</td>
<td>±26%</td>
<td>3% (-3%, 10%)</td>
</tr>
<tr>
<td>k-NN</td>
<td>-6% (-14%, 2%)</td>
<td>±25%</td>
<td>5% (-4%, 14%)</td>
</tr>
<tr>
<td>NN</td>
<td>-11% (-19%, -2%)</td>
<td>±29%</td>
<td>6% (-5%, 16%)</td>
</tr>
<tr>
<td>Bayes2</td>
<td>3% (-5%, 11%)</td>
<td>±25%</td>
<td>5% (-2%, 11%)</td>
</tr>
<tr>
<td>Reader 1</td>
<td>15% (7%, 23%)</td>
<td>±25%</td>
<td>-10% (-20%, -1%)</td>
</tr>
<tr>
<td>Reader 2</td>
<td>22% (13%, 31%)</td>
<td>±29%</td>
<td>-8% (-17%, 1%)</td>
</tr>
<tr>
<td>Readers (average)</td>
<td>18% (11%, 26%)</td>
<td>±24%</td>
<td>-9% (-18%, 0%)</td>
</tr>
</tbody>
</table>

*Statistics are as in Table 6.
In this work Bayes2 was the first attempt to include spatial context of pixels in a classifier. Although the method appears to yield more homogeneous classification results (Fig. 5), this did not lead to better results for the various percentages on the subject level. Possible improvements may be obtained by using context information from neighboring MRI slices, or by modeling the topology of the whole carotid artery. In this study no information from neighboring slices was used in the classification performed by the classifiers and the MRI readers.

Of the classifiers used in the present study, the Bayes classifier is the most primitive one because it models the data of each tissue type only by means of multivariate normal distributions. Strikingly, for the present data this classifier performed at least as well as the more-complex classifiers (i.e., the NN and k-NN classifiers). Therefore, in our view, it is not useful to use more advanced classifiers for the current data.

The evaluation presented in this article shows that supervised classifiers can facilitate the quantification of plaque components, particularly the high-risk component. Larger prospective studies are warranted to confirm these encouraging results.

ACKNOWLEDGMENTS

We thank Mark Lobbes, Suzanne Gerretsen (Maastricht University Hospital), and Martin Knýt (Eindhoven University of Technology) for their help in preprocessing the MRI data. In addition we thank Jos M.A. van Engelschoven (MRI reader) and Robert J. van Suylen (histology reader), both of whom are affiliated with Maastricht University Hospital.

REFERENCES