Published in final edited form as:
Magn Reson Med. 2012 June ; 67(6): 1815–1826. doi:10.1002/mrm.23189.

Multivariate Analysis of Cartilage Degradation using the Support Vector Machine Algorithm

Ping-Chang Lin, **Onyi Irrechukwu**, **Remy Roque**, **Brynne Hancock**, **Kenneth W. Fishbein**, and **Richard G. Spencer**
Magnetic Resonance Imaging and Spectroscopy Section, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224

Abstract

An important limitation in MRI studies of early osteoarthritis is that measured MRI parameters exhibit substantial overlap between different degrees of cartilage degradation. We investigated whether multivariate support vector machine (SVM) analysis would permit improved tissue characterization. Bovine nasal cartilage samples were subjected to pathomimetic degradation and their T_1, T_2, magnetization transfer rate (k_m) and apparent diffusion coefficient (ADC) were measured. SVM analysis performed using certain parameter combinations exhibited particularly favorable classification properties. The areas under the ROC curve for detection of extensive and mild degradation were 1.00 and 0.94, respectively, using the set (T_1, k_m, ADC), compared to 0.97 and 0.60 using T_1, the best univariate classifier. Further, a degradation probability for each sample, derived from the SVM formalism using the parameter set (T_1, k_m, ADC), demonstrated much stronger correlations ($r^2 = 0.77 – 0.88$) with direct measurements of tissue biochemical components than did even the most well-performing individual MRI parameter, T_1 ($r^2 = 0.53 – 0.64$). These results, combined with our previous investigation of Gaussian cluster-based tissue discrimination, indicate that the combinations (T_1, k_m) and (T_1, k_m, ADC) may emerge as particularly useful for multivariate cartilage matrix characterization.

Keywords

classification; osteoarthritis; MRI; discriminant analysis; clustering

Introduction

The development of noninvasive MRI approaches to the detection of early osteoarthritis and to monitoring therapeutic response to interventions has been the subject of intense activity. However, MRI exhibits limited sensitivity to cartilage pathology. As a result, even when a statistically significant difference is observed in the mean values of a given parameter between e.g. control and degraded cartilage, there generally remains a substantial degree of overlap in the parameter values obtained for samples belonging to the two groups (1–3). The situation is even more problematic when attempting to characterize degrees of degradation. This results in a limited ability to classify tissue according to any single MRI parameter (2).

Multivariate discriminant analysis has to date seen limited application to MR studies of tissue (4–7). Previously, we used Gaussian cluster analysis to categorize normal and degraded cartilage in multidimensional MRI-parameter space, as well as to provide a metric of degradation. In the present work, we extend this multiparametric perspective by applying

Address correspondence to: Richard G. Spencer, Magnetic Resonance Imaging and Spectroscopy Section, NIH/National Institute on Aging, GRC 4D-06, Baltimore, MD 21224, spencer@helix.nih.gov, 410-558-8226, FAX: 410-558-8318.

a more flexible type of discriminant analysis, the support vector machine (SVM) formalism, to categorize normal and degraded cartilage as well as to provide a scale of degradation based on assignment probabilities. This approach also allows us to investigate the ability of the SVM to predict macromolecular composition of the tissue.

SVM analysis defines those data points, known as support vectors, that are ultimately correctly classified through the algorithm, but are least characteristic of the category they belong to. These points define the classification decision hypersurface. The SVM is an extension of linear discriminant analysis, and provides an extremely flexible framework for classification, making minimal assumptions regarding the distribution of data points in parameter space. Data points are segregated into groups with reduced parameter overlap through use of a transformation based on a nonlinear kernel function. Further, the SVM can be readily implemented in a space with dimensionality equal to the desired number of MRI parameters to be used for characterization, and also provides a natural metric of degree of degradation based on distance from the classification hypersurface in parameter space (8).

As in our previous work (6), we perform our experiments and analyses on bovine nasal cartilage (BNC). BNC exhibits molecular and tissue characteristics similar to articular cartilage, but is largely homogeneous and isotropic. A trypsin digestion protocol primarily acting to deplete proteoglycan in the extracellular matrix (ECM) of cartilage was used to model the osteoarthritic process, with a lengthy digestion period implemented to ensure sample variation. The parameter spaces used for classification via the SVM were formed from the common MRI outcome measures T_1, T_2, magnetization transfer rate (k_m) and apparent diffusion coefficient (ADC).

ROC analysis is an established technique for evaluating the quality of a binary classification test. We used this approach to define the quality of SVM-based discriminant analysis for the detection of degraded cartilage and, for comparison, to characterize the quality of univariate classifiers. These curves were constructed for both the extensive (24 hr) trypsin degradation protocol implemented for the present study, as well as for samples previously prepared under a much milder (6 hr) trypsin degradation protocol (6).

Finally, one of the goals of noninvasive MR analysis of tissue is to draw inferences regarding macromolecular content; therefore, we quantified sulfated glycosaminoglycan (sGAG), collagen and tissue hydration for each sample before and after extensive degradation in order to investigate the predictive ability of SVM-based degradation probabilities. Assignment probabilities to control or to degraded status were calculated using a sigmoidal distance function specifying the separation between a given data point and the decision hyperplane defined by the SVM algorithm (9). This was compared with a more conventional univariate analysis based on the difference between a given MRI parameter for a particular validation set sample and the mean of that parameter for a training set (2).

Methods

Sample Preparation

8-mm diameter bovine nasal cartilage (BNC) discs were excised from nasal septa of a freshly slaughtered ~6 month-old calf (Green Village Packing, Green Village, NJ). A central 2.5-mm plug was removed from each disc and retained for biochemical tests. The discs with central holes were threaded in groups of 5 or 6 onto a hollow glass tube, four of which were inserted into a well of a susceptibility-matched four-well sample holder containing Dulbecco's phosphate-buffered saline (DPBS; Invitrogen, Grand Island, NY) at pH 7.5 ± 0.1 for initial MRI measurements (Fig. 1A). Following these measurements, cartilage degradation was induced through addition of 1 mg/ml trypsin (Sigma-Aldrich, St. Louis,

MO) into the wells followed by incubation of the entire sample holder at 37 °C in a 5% CO_2 atmosphere for 24 hrs, after which MRI measurements were repeated. A total of 36 BNC discs samples harvested from a young bovine were imaged before and after trypsin depletion.

MRI Measurements

Imaging was performed using a 9.4T/105-mm Bruker DMX spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany) equipped with a 30-mm proton birdcage resonator with sample temperature maintained at 4.0 ± 0.1 °C through use of cold air from a vortex tube (Exair, Cincinnati, OH) to reduce loss of proteoglycans. Images were acquired from 0.5-mm-thick sagittal slices defined through the center of each of the 4 wells in the sample holder, permitting all samples within a given well to be imaged simultaneously. T_2 data were measured using a 64-echo CPMG pulse sequence with TR/TE = 5 s/12.8 ms. A progressive saturation spin-echo sequence with TE = 12.8 ms and TR varying logarithmically from 100 ms to 15 s in 12 steps was employed to perform T_1 mapping. The sample configuration and a representative T_1 image are shown in Fig. 1. MT data were acquired using the same spin-echo sequence preceded by a 6 kHz off-resonance saturation pulse with amplitude B_1 = 12 μT and duration incremented from 0.1 to 4.6 s in 8 steps. This offset frequency avoided direct saturation of bulk water, while the power level of 12 μT produced a 180° rotation with a pulse length of 1 ms (10). Apparent diffusion coefficient (ADC) was measured using a spin-echo sequence (TR = 5s) incorporating a pair of identical gradient pulses of δ = 5 ms duration placed on either side of the 180° refocusing pulse, with a constant interval of Δ =12.5 ms between the gradient pulse centers. The gradient strength G was increased from 0 to 320 mT/m in 8 steps. Other parameters included NEX = 2, BW = 50 kHz, FOV = 4.0 × 1.5 cm (read × phase direction, corresponding to vertical × horizontal orientation within the vertical-bore magnet), matrix size = 256 × 128 and resolution = 156 × 117 μm. A region of interest (ROI) was selected to cover all of the pixels visualized within a BNC disk for each image, as illustrated by the highlighted area in Fig. 1B. Averaged signal intensities were computed independently within the ROI for every sample and then fit to appropriate three-parameter monoexponential functions to determine T_1, T_2, k_m, and ADC (2,10–12).

Biochemical analysis of BNC samples

sGAG and collagen contents for each BNC sample were quantified prior to degradation by biochemical analysis of the removed central plug, and after degradation by analysis of the remaining disc. Samples were weighed before and after overnight vacuum drying at room temperature to determine water content. Tissue hydration was defined as the ratio of the difference between each sample's wet and dry weights to its wet weight. The vacuum dried samples were then individually immersed in a solution of 1mg/ml proteinase K (Sigma-Aldrich, St. Louis, MO) containing 50 mM Tris hydrochloride and 1 mM $CaCl_2$ at pH 8.0, digested in a 60 °C water bath overnight, and centrifuged at 14000 × g for 5 min. A sample of each digest was removed to determine sGAG concentration using the dimethylmethylene blue (DMMB) dye-binding assay (13). In addition, an aliquot of each digest was hydrolyzed in 6N hydrochloric acid solution at 105 °C overnight and subsequently neutralized with sodium hydroxide. Chloramine −T was then used to oxidize the free hydroxyproline in the neutralized digest, followed by the colorimetric collagen assay to quantify hydroxyproline concentration, from which collagen content was calculated assuming 14% abundance of hydroxyproline in collagen (14).

Classification

Two datasets were analyzed to assess classification by the SVM algorithm. The first was the group of 36 BNC samples prepared for the present investigation with MRI measurements

performed pre- and post- 24-hour trypsin degradation as described above. The second dataset, prepared for a previous investigation, consisted of 40 control samples and 40 samples subjected to a milder 6-hour trypsin degradation protocol. Again, MRI measurements were performed before and after exposure to the enzyme (6). No biochemical measurements of sGAG or collagen were available for this latter dataset.

In constructing the classification model, bias was reduced by randomly dividing the samples into training sets and validation sets through 10-fold cross-validation (9). The averaged training set and validation set results are reported. Two independent classification approaches, through use of mean values and through use of the SVM algorithm, were applied in order to evaluate the quality of the multivariate SVM approach as compared to conventional classification. All analyses were performed using in-house designed scripts written in Matlab 7.4 (The Mathworks Inc., Natick, MA) for the univariate approach and the e1071 package, based on libsvm library, written in the R language for the multivariate SVM approach (15,16).

Assignment probability based on arithmetic means of individual MRI parameters

Degradation probabilities were derived for each sample in the pre- and post- 24-hr trypsin degradation dataset. For each step of the cross-validation, the arithmetic mean of a given MRI parameter s was calculated for a training set consisting of 90% of the samples. Each remaining sample j in the validation set was then assigned a pair of probabilities ($P_{pre,\,j}$, $P_{post,\,j}$), where $P_{pre,\,j} + P_{post,\,j} = 1$, based on the difference, d, between the value of s for sample j and the mean values of parameter s for the pre- and post-degraded groups, ($\bar{s}_{pre}$, $\bar{s}_{post}$), in the training set. For sample values $s_j \in (P_{pre,\,j}, P_{post,\,j})$ the probabilities were calculated according to (17):

$$(P_{pre,j}, P_{post,j}) = \left(1 - \frac{d(s_j, \bar{s}_{pre})}{d(s_j, \bar{s}_{pre}) + d(s_j, \bar{s}_{post})}, 1 - \frac{d(s_j, \bar{s}_{post})}{d(s_j, \bar{s}_{pre}) + d(s_j, \bar{s}_{post})}\right) \quad (1)$$

For sample values outside the range ($P_{pre,\,j}$, $P_{post,\,j}$), the pair of probabilities ($P_{pre,\,j}$, $P_{post,\,j}$) was defined as (1, 0) or (0,1), as appropriate. The same probability calculation was applied to training set samples. These results, based on Eq. (1), were compared to assignment probabilities derived from the SVM approach described below.

Assignment probability based on support vector machine analysis

The SVM algorithm as applied in this work is a generalization of linear discriminant analysis for classification of two groups in a native feature space. In the case in which the groups are linearly separable, the separating hyperplane can be described by the equation

$$X^T \cdot \beta + \beta_0 = 0, \quad (2)$$

where the first term is the inner product of parameter space coordinates X and a vector β, defining the normal to the hyperplane, and where β_0 is an intercept (9). The hyperplane, which is optimal in the sense of maximizing the separation between the two groups, is found through the method of Lagrange multipliers, with constraints $y_i (X_i \cdot \beta + \beta_0) \geq C$, where y_i is the group label of the data point X_i and C defines the margin between the most difficult to classify points, referred to as support vectors, with $2C = 2/\|\beta\|$ (9).

While complete separation in the native feature space is in general not possible, the training set data are transformed by a mapping function $h(X)$ into a higher dimensional feature space in which the maximal margin between the two groups can be determined by solving a linear

convex optimization problem (9). This optimal hyperplane is then transformed to nonlinear boundaries in the original parameter space defined by the solution function

$$f(X)=h(X)^T\beta+\beta_0=\sum_{i=1}^{N}\alpha_i y_i\langle h(X), h(X_i)\rangle+\beta_0 \quad (3)$$

where β, derived from the Lagrangian, is given by $\beta=\sum_{i=1}^{N}\alpha_i y_i h(X_i)$ (9). α_i is a positive-valued Lagrange multiplier satisfying $\alpha_i < C$. C serves as a regularization parameter that represents the tradeoff between minimizing the error of classification (smaller values of C) and maximizing the support vector margin (larger values of C). In accordance with the SVM algorithm as implemented in the e1071 package, data points were transformed from their original space, as defined by the MRI parameters selected to describe the samples, into a higher dimensional feature space in which the optimal hyperplane as described above could be determined by maximizing the Lagrangian dual function (9). This transformation was achieved through use of a kernel function $K(X, X')=\langle h(X), h(X')\rangle$, implicitly defining an inner product within the transformed feature space (18). We selected the widely-used Gaussian radial basis function,

$$K(X, X')=\exp(-\left|\left|X-X'\right|\right|^2/\sigma). \quad (4)$$

which introduces a single adjustable parameter, σ, defining its extent and curvature (19,20). The values for σ and C were defined by implementing an exponentially-spaced grid search to minimize the classification error over the ranges $\sigma\in[2^{-3}, 2^6]$ and $C\in[2^{-2}, 2^4]$ (15). Finally, an estimate of the probability that a data point belongs to one of the two classes defined by the SVM decision hypersurface in the input parameter space (or the hyperplane in the transformed feature space) can be calculated from a sigmoidal function that equals 0.5 for a point on the decision hypersurface and approaches unity for the selected class for a data point progressively more distant from the decision hypersurface (8,9):

$$\Pr(y=1|X)=\frac{1}{1+e^{-\beta_0-\sum_{i=1}^{N}\alpha_i K(X,X_i)}} \quad (5)$$

Construction of ROC curves

Given assignment probabilities resulting from either univariate arithmetic means or multivariate SVM analysis, the true positive rate (TP; sensitivity) and false positive rate (FP; 1 - specificity) depend upon a selected decision threshold. In our case, these were based on assignment probabilities P. For a value P ranging between zero and unity, samples were classified as degraded if their assignment probability to the degraded group was greater than P. This permits construction of ROC curves defining false positives and negatives as a function of the decision threshold. ROC curves were constructed in this manner for T_1, the best univariate classifier, as well as for the 11 available multivariate combinations of the measured MRI parameters. The quality of a particular classifier was determined by the area under its ROC curve (21).

SVM-based Matrix Composition Predictions

The ability to predict biochemical composition of degraded cartilage from individual MRI outcome parameters was assessed by forming correlations between the components of matrix composition (sGAG, collagen and hydration) and T_1, T_2, k_m and ADC. Similarly, regression was used to construct correlations between biochemical outcomes and SVM-

based degradation probabilities in order to determine whether this multiparametric approach yielded more predictive relationships.

Statistical Analysis

Values of MRI parameters and biochemical content corresponding to the pre-and post-degraded samples are reported as mean ± SD. The Anderson-Darling test was used to assess normality of the data distribution for each parameter (22). The two-tailed paired Student's t-test was employed to determine the significance of the difference between the pre- and post-degraded parameter values for normally distributed parameters, while significance for the one parameter that did not satisfy the normality assumption, k_m, was tested using the Wilcoxon signed rank test (23).

Results

Table 1 shows average biochemical and MRI parameter results. Treatment with trypsin resulted in a statistically significant decrease in sGAG content per wet weight and a concomitant increase in tissue hydration and in collagen concentration per wet weight. The increase in collagen/ww follows from the extensive and preferential loss of PG, resulting also in net loss of water. The substantial relative increase in collagen/ww follows from this loss of PG and water, rather than from any absolute increase in collagen macromolecules. Substantial and statistically significant increases in T_1 and ADC, and a decrease in k_m, accompanied these biochemical changes. There was also a statistically significant, though small, increase in T_2 with degradation.

We explored the correlation between the parameters that changed significantly between the pre- and post-degraded data sets, and found that the correlations were not necessarily large. For example, while the correlation coefficient for the relationship between T_1 and k_m was $r^2 = 0.61$, a smaller value of $r^2 = 0.42$ was found for the correlation between T_1 and ADC, while k_m and ADC were essentially uncorrelated, with $r^2 = 0.10$. Together, these results indicate the lack of redundancy between these measurements, and are consistent with the classification improvement that results from multivariate analysis.

An example of SVM analysis is shown in Fig. 2, in which BNC samples before and after exposure to trypsin for 24 hours are represented in (T_1, k_m, ADC) space. A single iteration of the 10-fold cross-validation procedure is illustrated, with a decision hypersurface constructed from the SVM algorithm using 90% of the samples and with classification performed on the remaining 10%. The decision hypersurface is defined as the locus of points on which samples would have an equal probability of being assigned to the pre-degradation or trypsin-degraded groups. In this example, samples with a representation in parameter space lying outside the hypersurface are assigned with a greater than 50% probability to pre-degradation status, while samples within the hypersurface are assigned to the trypsin-degraded class. In the example shown, there are no misclassifications.

An ROC curve permits evaluation of test characteristics over a range of binary classification decision thresholds, with the area under the curve providing a measure of the ability of a variable to discriminate between groups (21). The ROC curves presented in Fig. 3A for training sets (with a portion expanded in Fig. 3B) and in Fig. 3C for validation sets (expanded portion in Fig. 3D) indicate that classification according to particular MRI parameter sets through multivariate SVM analysis generally resulted in an improvement over classification by the best univariate classifiers, T_1 and ADC. Quantitative results are shown in Table 2, columns 2 and 3, in which areas under the ROC curves are presented. The best parameter pair for classification was (T_1, k_m), while the best triplet was (T_1, k_m, ADC). These were also the best or near-best parameter combinations for classification using

Gaussian clustering (6). Similarly, T_1 was found in both the present study as well as in our previous work to be the best univariate classifier (6). We note that (T_2, k_m), the worst bivariate classifier with the SVM approach, performed less well than either T_1 and ADC in both the training and validation sets,. This was also the case for Gaussian clustering (6).

We further evaluated the performance of SVM analysis on mildly degraded cartilage through use of data previously obtained on samples subjected to a substantially lesser degree of trypsin digestion (6). These results are shown in Fig. 4 and Table 2, columns 4 and 5. While, as expected, the areas under the ROC curves were generally smaller than for more extensively degraded tissue, the best multivariate parameter sets, (T_1, k_m) and (T_1, k_m, ADC), maintained excellent performance in both cases. In addition, they were markedly superior to classification according to T_1, which again emerged as the best uniparametric classifier although exhibiting inferior performance in classifying mildly degraded as compared to more extensively degraded samples.

A direct comparison of SVM and Gaussian-based clustering is shown in Table 3 for the data set consisting of control and 6-hr trypsin treated BNC samples (6). Results are again based on 100 realizations of random simple splits into training and validation sets. The results using the SVM approach support the ROC analysis shown in Table 2. Improved classification accuracy in the validation sets across various multivariate classifiers using the SVM analysis is generally observed as compared to those using the Gaussian-based approach for parameter spaces of the same dimension. However, an important observation is that the Gaussian clustering approach exhibits little improvement with the addition of a third classification parameter, while classification with the SVM exhibits substantial improvement. This is consistent with the smaller decline in specificity and sensitivity between training and validation sets in the SVM analysis than in the Gaussian-based clustering analysis. Together, these points reflect the smaller degree of overfitting with the SVM analysis.

Changes in MRI parameters are seen with enzymatic digestion; as noted, a central question is the extent to which these parameter changes predict matrix characteristics. Fig. 5 provides examples of relations between MRI parameters and tissue biochemical properties in the pre- and post- 24-hr trypsin degradation group. In Figs. 5A – B, each sample's sGAG/ww and hydration values are plotted against its T_1 value. T_1, the best univariate classifier, was used for this analysis since it also yielded the largest correlation coefficients for the illustrated relationships. Relationships between each sample's collagen/ww and T_1 value were also established and showed a significant positive linear correlation ($r^2 = 0.59$). While not artifactual, this highly counterintuitive result of a positive correlation illustrates the importance of interpreting MR results in the context of the full set of biochemical changes in tissue; because of the preferential and substantial loss of sGAG molecules in the samples, degradation resulted in an increase in the relative measure of collagen, that is, collagen/ww, although degradation would not result in an increase in the absolute amount of collagen in the samples. This effect is highlighted by the strong negative ($r^2 = 0.91$) correlation between sGAG/ww and collagen/ww. Strong correlations were also obtained for ADC, the second-best univariate classifier of training sets, as a predictive variable (data not shown). However, consistent with previous findings, within-cluster correlations between T_1 and biochemical outcomes were not observed (24,25). This may be due to the more limited dynamic range of these correlations, or to the limitations of univariate analysis in the setting of complex biochemical changes, as noted above. As expected, k_m, which did not perform well as a univariate classifier, showed markedly weaker correlations with sGAG and hydration (Fig. 5C – D). Although T_2 changed significantly with degradation, it correlated relatively weakly with tissue biochemistry (Fig. 5E – F). This suggests that cartilage T_2 values are strongly affected not only by overall biochemical content, but also by macromolecular structure (26).

We constructed similar correlations using classification probabilities based on the SVM formalism instead of individual MRI outcome measures. Probability was defined with respect to degradation status, with a value of unity assigned to a sample with 100% probability of belonging to the 24-hr-trypsin-degraded group, and a value of zero to a sample with 100% probability of belonging to the pre-digestion group. Intermediate probabilities were derived as described above, permitting assignment of a degree of degradation to each sample. Fig. 6A – B shows the correlations between these assignment probabilities as derived from SVM analysis of the best bivariate classifier, (T_1, k_m), and biochemical characteristics of the corresponding samples. As seen, the r^2 values for these correlations were on the order of 0.8 or greater, representing a marked improvement over the results shown in Fig. 5A – B. Fig. 6C – D shows the corresponding results for the set (T_1, k_m, ADC), the best triplet classifier; improvement over the results for (T_1, k_m) were at best minimal. This is consistent with the marginal improvement in binary classification provided by this triplet as compared to (T_1, k_m), as seen in Table 2. Overall, assignment probabilities obtained through multivariate MRI classifiers, with the exception of (T_2, k_m), exhibited higher correlations with biochemical characteristics than did the best univariate classifier, T_1.

Discussion

Our study was designed to evaluate the ability of SVM analysis to improve upon the conventional uniparametric approach to cartilage characterization. We found that in fact the sensitivity and specificity of binary classification was markedly improved by the SVM approach, as was the ability to predict changes in matrix composition from observed changes in MRI parameters.

In the univariate analysis performed in the present work, we found that T_1 performed well in terms of sensitivity and specificity with respect to cartilage degradation status. This was consistent with our previous results (2,3,6). This sensitivity may be based at least in part to the fact that we performed these studies at a field strength substantially greater than that used in clinical studies (2,27), and T_1, demonstrates a known increase with field strength. However, an actual decrease in the dynamic range of T_1 has been reported in a study of brain (28). To our knowledge, there have been no previous studies of the classification performance of MRI parameters at any field strength, so that classification dependence on field strength remains speculative. This indicates a need for such formal assessment at clinically-relevant field strengths. Finally, we note that the sensitivity of T_1 to degradation status may be due primarily to its sensitivity to tissue hydration rather than to changes in the content of particular macromolecular species. Distinguishing between these two effects would require a different experimental design from that employed in the present work.

Another departure of the present methodology from the clinical research setting is the fact that these experiments were performed at 4° C in order to minimize sample degradation during imaging. Because of the temperature dependence of relaxation times and ADC, numerical values for data collected at, say, body temperature would differ from what we have reported. Indeed, the manner and degree to which sample temperature affects our specific results also remains speculative. However, while a change in the specifics of, for example, the SVM-determined hypersurface is to be expected, the improvement in classification accuracy through use of multivariate as compared to univariate analysis is likely to be retained. These considerations further indicate the need for assessment of multivariate methods in settings of different magnetic fields, temperatures, and pulse sequences, as well as in application to other cartilage models such as osteoarthritis and engineered tissue.

Classification through multivariate discriminant analysis

Assessment of cartilage status through use of the arithmetic means of single MRI parameters, which is, in effect, the conventional approach (29,30), demonstrates limited sensitivity and specificity due to the substantial degree of overlap in MRI parameters between groups (2,3,6). An important alternative is multivariate discriminant analysis, which we have implemented in the present work through the SVM algorithm. The most salient advantage of this is that it incorporates simultaneous knowledge of two or more parameters. This fact, even in the original parameter space defined by the specified MRI parameters, can provide a marked reduction in data overlap between groups. In addition, the SVM incorporates a transformation of the original parameter space into a higher-dimensional feature space in which class members can be more readily distinguished. We implemented the SVM as a supervised learning approach, in which a discriminating model was established based upon the known class memberships of training samples.

In SVM analysis, the transformation into the higher dimensional feature space in which classification occurs permits a great deal of flexibility in constructing an optimal separating hyperplane. To render the optimization problem more tractable, a kernel function is selected that permits evaluation of inner products in feature space through evaluation of this kernel in the original parameter space. This results in the computation of a linear decision hypersurface in the transformed feature space which maps to a highly nonlinear surface in the original parameter space (Fig. 2). We chose the commonly-used radial basis function as a kernel function, with a value that depends only upon the normalized squared distance between points. For a given support vector X, the individual weight of a point X is proportional to $e^{-\|X-X'\|/\sigma}$, with σ representing the width of the Gaussian function. A larger value of σ results in a greater spread of non-negligible kernel values from the support vector X', exhibiting a smoother, and hence more generalizable, but possibly underfit decision hypersurface defined by the training set (31,32). In contrast, a smaller value of σ results in a more highly curved decision hypersurface with localized Gaussian densities centered around support vectors, surrounded by regions of negligible values for the kernel function, exhibiting a greater potential for overfitting (31,32). Similarly, a large value for the regularization parameter C can lead to a larger number of support vectors and hence overfitting, while a small C can result in underfitting (33). Fig. 2 shows an example of a relatively large σ value, $\sigma = 2$, yielding a smooth decision hypersurface, but with the potential for underfitting balanced by a large value of $C = 16$.

The accuracy of binary classification using the univariate procedure or the SVM procedure was examined by constructing ROC curves, in which sensitivity and specificity are displayed as a function of decision threshold. It is evident that classification performance was greatly enhanced by the SVM as compared to the univariate approach in analyses of both the extensively-degraded and the mildly-degraded samples. The poor classification performance of any single MRI outcome measure is a direct consequence of the high degree of overlap of parameter values for control and degraded tissue. The same consideration applies to univariate fuzzy c-means clustering, which can also assign degrees of degradation in the present context, in spite of this approach being less sensitive to extremes of data values (2).

We used BNC in the present study, since it is much more spatially homogeneous and isotropic than is articular cartilage while exhibiting essentially the same ultrastructural characteristics (34). This permits sensitivity and specificity analysis to be performed on individual samples while minimizing the complexity of variations of parameter values and biochemical composition within them. Nevertheless, biologic variation, including the fact that biochemical degradation also lacks complete consistency, results in a large dynamic range in tissue properties and MRI parameter values. This, along with the imperfect

correlation of MR outcomes with tissue biochemistry, also results in substantial overlap in measured parameters between the intact and degraded sample groups (2). Thus, BNC is an excellent model for establishing our formalism. However, there remains a great deal of interest in assessment of cartilage, including degradation status, in the clinical research setting. Further studies with articular cartilage will rely upon a pixel-by-pixel multivariate analysis, resulting in a novel means of imaging cartilage degradation (35).

The greatly increased potential to distinguish between data points in different groups through use of multivariate analysis indicates that the MRI parameters used as outcome measures are sensitive to different aspects of tissue properties. In the present study, the SVM approach applied to even the mild enzymatic degradation dataset of Ref. (6) exhibited markedly improved classification accuracy. The emergence of (T_1, k_m) and (T_1, k_m, ADC) as particularly robust classifiers is consistent with the results of our previous investigation of multiparametric classification using Gaussian clusters (6). The consistency of this finding indicates that these parameters may provide particularly non-redundant, complementary, information about the biophysical status of cartilage matrix, regardless of analytic approach. Comparing Table 2 of the present study with Tables 4 and 5 of Ref. (6), it is evident that the SVM approach tended to result in a greater improvement over uniparametric classification than did the Gaussian clustering approach.

As expected, and as indicated in Figs. 3 and 4, the areas under the ROC curves for training set data were larger than those for validation set data, for any MRI parameter or multivariate parameter combination. This was also the case for the Gaussian-model approach previously described (Tables 4 and 5 of Ref. (6)). Although using a sampling method such as cross-validation tends to minimize the problem of overfitting, this problem is also dependent upon the distribution of data points in feature space and the analytic approach. In fact, as shown in Table 3, SVM analysis exhibited a substantial advantage over Gaussian clustering (6). Although training set classification accuracy for mildly-degraded samples was comparable for the SVM and Gaussian clustering, the former exhibited better performance in the validation set, indicating less model overfitting.

Relationship between measured MRI parameters and cartilage composition

Binary classification is clearly inadequate to describe the graded nature of cartilage degradation. In addition, early detection of minimally-degraded tissue is of potential importance in the development of therapeutic interventions which may be more efficacious in early disease (36). To address these considerations, several previous studies have correlated extracellular matrix components of cartilage with MRI parameter values (37–39). Certain trends have been established, although with limited success.

As shown in Fig. 5, although significant changes were observed in each of the MRI parameters measured (Table 1), the correlations with matrix components were highly variable. The correlations with T_1, the best univariate classifier, were very good, while correlations with k_m and T_2 were poor. The correlations with T_1 were largely due to the structure of the data, which exhibited two fairly distinct clusters of points due to the extensive degradation protocol implemented.

Across the entire data set, T_1 correlated well with both sGAG ($r^2 = 0.64$) and hydration ($r^2 = 0.53$); this is consistent with the relatively high correlation between sGAG and tissue hydration ($r^2 = 0.68$). It cannot be determined from these data whether T_1 is specifically dependent upon sGAG content or hydration; however, previous analysis indicated that T_1 may be more highly dependent upon hydration (40). Alternatively, the negative correlation of $r^2 = 0.53$ between T_1 and the overall macromolecular content, defined as 1 – hydration, follows from Fig. 5B.

As noted, the correlation between T_1, as well as the other univariate outcome measures, and macromolecular content was much less robust within the rather restricted dynamic range represented by the distinct clusters evident in each panel of Fig. 5. For example, T_1 and sGAG/ww were essentially uncorrelated, with $r^2 = 0.10$ and $r^2 < 0.01$, within the pre-and post-degraded clusters in Fig. 5A, respectively. A similar result was obtained in the study reported by Stikov *et al.*, in which T_1 demonstrated correlations with sGAG/ww only when both upper and lower portions of plug samples from articular cartilage were analyzed together (24). The work of Potter *et al.* showed strong correlation between T_1 and sGAG, while that of Irrechukwu *et al.* did not find such a correlation (37,41). These variable results can be attributed to the fact that both the MRI and the biochemical measurements necessarily exhibit measurement error, to the necessity for multivariate analysis, or to an intrinsic lack of correspondence between these outcomes. The latter could reflect, for example, the role of factors besides absolute macromolecular content in determining values for MRI outcome measures. Such factors may include macromolecular structure and crosslinking and interactions between macromolecular species.

The ability of MRI parameter combinations to predict changes in cartilage matrix caused by pathomimetic enzymatic degradation has not been extensively studied. We addressed this issue previously through fuzzy assignment to Gaussian clusters (6). In the present work, using the more flexible SVM formalism, we defined a degradation probability for each sample based on its sigmoidal distance from the decision hypersurface. Correlations of these probabilities with tissue biochemistry are shown in Fig. 6. These multiparametric MRI-derived probabilities were much more highly correlated with biochemical outcomes than were the values of any individual MRI parameter. The improved reliability of these probability values, as compared to individual MRI parameters, for predicting biochemical status is directly reflected in the consolidation of the clusters representing non-degraded versus degraded samples. It is of particular interest that the parameter combinations that best reflected matrix component concentrations were the same as those that exhibited the best binary classification accuracies.

In conclusion, conventional uniparametric MRI assessment of matrix status in degenerative cartilage is limited by the overlap in parameter values between varying degrees of degradation. This is based on the intrinsic physicochemical properties of the tissue and the sensitivity of the measurement technique to these properties, and cannot be altered through any statistical means. However, we have shown that multiparametric analysis using two or more standard MRI outcome measures can result in greatly improved characterization of cartilage matrix. The multivariate SVM-based discriminant analysis implemented in the present work yielded classification accuracy in the validation sets similar to the accuracy in the training sets, representing improved performance as compared with model-based clustering (6). In addition, degradation probabilities derived from the SVM procedure demonstrated markedly stronger correlations with biochemical measurements than did individual MRI parameters. These results indicate the ability of multivariate analysis to greatly augment MRI assessment of cartilage matrix status in basic science studies. This approach, based entirely on conventional MRI protocols, may be of potential use in the clinical research and clinical settings as well.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute on Aging.

References

1. Laurent D, Wasvary J, Rudin M, O'Byrne E, Pellas T. In vivo assessment of macromolecular content in articular cartilage of the goat knee. Magn Reson Med. 2003; 49(6):1037–1046. [PubMed: 12768582]
2. Lin P-C, Reiter DA, Spencer RG. Sensitivity and specificity of univariate MRI analysis of experimentally degraded cartilage. Magn Reson Med. 2009; 62(5):1311–1318. [PubMed: 19705467]
3. Chen CT, Fishbein KW, Torzilli PA, Hilger A, Spencer RG, Horton WE Jr. Matrix fixed-charge density as determined by magnetic resonance microscopy of bioreactor-derived hyaline cartilage correlates with biochemical and biomechanical properties. Arthritis Rheum. 2003; 48(4):1047–1056. [PubMed: 12687548]
4. Valerio M, Panebianco V, Sciarra A, Osimani M, Salsiccia S, Casciani L, Giuliani A, Bizzarri M, Di Silverio F, Passariello R, Conti F. Classification of prostatic diseases by means of multivariate analysis on in vivo proton MRSI and DCE-MRI data. NMR In Biomedicine. 2009; 22(10):1036–1046. [PubMed: 19579258]
5. Asllani I, Habeck C, Scarmeas N, Borogovac A, Brown TR, Stern Y. Multivariate and univariate analysis of continuous arterial spin labeling perfusion MRI in Alzheimer's disease. Journal Of Cerebral Blood Flow And Metabolism. 2008; 28(4):725–736. [PubMed: 17960142]
6. Lin P-C, Reiter DA, Spencer RG. Classification of degraded cartilage through multiparametric MRI analysis. Journal Of Magnetic Resonance. 2009; 201(1):61–71. [PubMed: 19762258]
7. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, Bryan RN, Davatzikos C. Computer-Assisted Segmentation of White Matter Lesions in 3D MR Images Using Support Vector Machine. Academic Radiology. 2008; 15(3):300–313. [PubMed: 18280928]
8. Hastie T, Rosset S, Tibshirani R, Zhu J. The entire regularization path for the support vector machine. Journal Of Machine Learning Research. 2004; 5:1391–1415.
9. Hastie, T. The Elements of Statistical Learning. Springer; 2001.
10. Gray ML, Burstein D, Lesperance LM, Gehrke L. Magnetization transfer in cartilage and its constituent macromolecules. Magn Reson Med. 1995; 34(3):319–325. [PubMed: 7500869]
11. Stejskal EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. Journal of Chemical Physics. 1965; 42(1):288–292.
12. Hajnal JV, Baudouin CJ, Oatridge A, Young IR, Bydder GM. Design and implementation of magnetization transfer pulse sequences for clinical use. Journal of Computer Assisted Tomography. 1992; 16(1):7–18. [PubMed: 1729310]
13. Farndale RW, Buttle DJ, Barrett AJ. Improved Quantitation And Discrimination Of Sulfated Glycosaminoglycans By Use Of Dimethylmethylene Blue. Biochimica Et Biophysica Acta. 1986; 883(2):173–177. [PubMed: 3091074]
14. Reddy GK, Enwemeka CS. A simplified method for the analysis of hydroxyproline in biological tissues. Clinical Biochemistry. 1996; 29(3):225–229. [PubMed: 8740508]
15. Chang, C-C.; Lin, C-J. LIBSVM: a library for support vector machines. 2001.
16. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A. e1071: Misc Functions of the Department of Statistics. 2009.
17. Suckling J, Sigmundsson T, Greenwood K, Bullmore ET. A modified fuzzy clustering algorithm for operator independent brain tissue classification of dual echo MR images. Magnetic Resonance Imaging. 1999; 17(7):1065–1076. [PubMed: 10463658]
18. Hofmann T, Scholkopf B, Smola AJ. Kernel methods in machine learning. Annals of Statistics. 2008; 36(3):1171–1220.
19. Scholkopf B, Sung KK, Burges CJC, Girosi F, Niyogi P, Poggio T, Vapnik V. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Transactions On Signal Processing. 1997; 45(11):2758–2765.
20. Lin K-M, Lin C-J. A study on reduced support vector machines. Neural Networks, IEEE Transactions on. 2003; 14(6):1449.
21. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997; 30(7):1145–1159.

22. Thode, HC, Jr. Testing for normality. New York: Marcel Dekker; 2002.
23. Hollander, M.; Wolfe, DA. Nonparametric statistical methods. New York: John Wiley & Sons; 1973.
24. Stikov N, Keenan KE, Pauly JM, Smith RL, Dougherty RF, Gold GE. Cross-relaxation imaging of human articular cartilage. Magn Reson Med. 201110.1002/mrm.22865
25. Bashir A, Gray ML, Hartke J, Burstein D. Nondestructive imaging of human cartilage glycosaminoglycan concentration by MRI. Magn Reson Med. 1999; 41(5):857–865. [PubMed: 10332865]
26. Gründer W, Wagner M, Werner A. MR-microscopic visualization of anisotropic internal cartilage structures using the magic angle technique. Magn Reson Med. 1998; 39(3):376–382. [PubMed: 9498593]
27. Regatte RR, Schweitzer ME. Ultra-high-field MRI of the musculoskeletal system at 7. 0T. Journal of Magnetic Resonance Imaging. 2007; 25(2):262–269. [PubMed: 17260399]
28. de Graaf RA, Brown PB, McIntyre S, Nixon TW, Behar KL, Rothman DL. High magnetic field water and metabolite proton T1 and T2 relaxation in rat brain in vivo. Magn Reson Med. 2006; 56(2):386–394. [PubMed: 16767752]
29. Mosher TJ, Dardzinski BJ, Smith MB. Human articular cartilage: influence of aging and early symptomatic degeneration on the spatial variation of T2--preliminary findings at 3 T. Radiology. 2000; 214(1):259–266. [PubMed: 10644134]
30. Dunn TC, Lu Y, Jin H, Ries MD, Majumdar S. T2 relaxation time of cartilage at MR imaging: Comparison with severity of knee osteoarthritis. Radiology. 2004; 232(2):592–598. [PubMed: 15215540]
31. Burges CJC. A tutorial on Support Vector Machines for pattern recognition. Data Mining And Knowledge Discovery. 1998; 2(2):121–167.
32. Keerthi SS, Lin CJ. Asymptotic behaviors of support vector machines with Gaussian kernel. Neural Computation. 2003; 15(7):1667–1689. [PubMed: 12816571]
33. Alpaydin, E. Introduction to machine learning. Cambridge, MA: The MIT Press; 2004.
34. Reiter DA, Lin PC, Fishbein KW, Spencer RG. Multicomponent T-2 Relaxation Analysis in Cartilage. Magn Reson Med. 2009; 61(4):803–809. [PubMed: 19189393]
35. Lin, P-C.; Irrechukwu, O.; Spencer, RG. Mapping Cartilage Degradation through Support Vector Machine Probabilistic Classification. Proceedings 19th ISMRM Annual Meeting and Exhibition; Montréal, Ouébec, Canada. 2011. p. 1107
36. Gurkan I, Ranganathan A, Yang X, Horton WE Jr, Todman M, Huckle J, Pleshko N, Spencer RG. Modification of osteoarthritis in the guinea pig with pulsed low-intensity ultrasound treatment. Osteoarthritis and Cartilage. 2010; 18(5):724–733. [PubMed: 20175971]
37. Potter K, Butler JJ, Horton WE, Spencer RGS. Response of engineered cartilage tissue to biochemical agents as studied by proton magnetic resonance microscopy. Arthritis And Rheumatism. 2000; 43(7):1580–1590. [PubMed: 10902763]
38. Gray ML, Burstein D, Xia Y. Biochemical (and functional) imaging of articular cartilage. Semin Musculoskelet Radiol. 2001; 5(4):329–343. [PubMed: 11745049]
39. Nissi MJ, Toyras J, Laasanen MS, Rieppo J, Saarakkala S, Lappalainen R, Jurvelin JS, Nieminen MT. Proteoglycan and collagen sensitive MRI evaluation of normal and degenerated articular cartilage. J Orthop Res. 2004; 22(3):557–564. [PubMed: 15099635]
40. Berberat JE, Nissi MJ, Jurvelin JS, Nieminen MT. Assessment of interstitial water content of articular cartilage with T1 relaxation. Magnetic Resonance Imaging. 2009; 27(5):727–732. [PubMed: 19056195]
41. Irrechukwu ON, Lin P-C, Fritton K, Doty S, Pleshko N, Spencer RG. Magnetic Resonance Studies of Macromolecular Content in Engineered Cartilage Treated with Pulsed Low-Intensity Ultrasound. Tissue Engineering Part A. 2011; 17(3–4):407–415. [PubMed: 20807015]

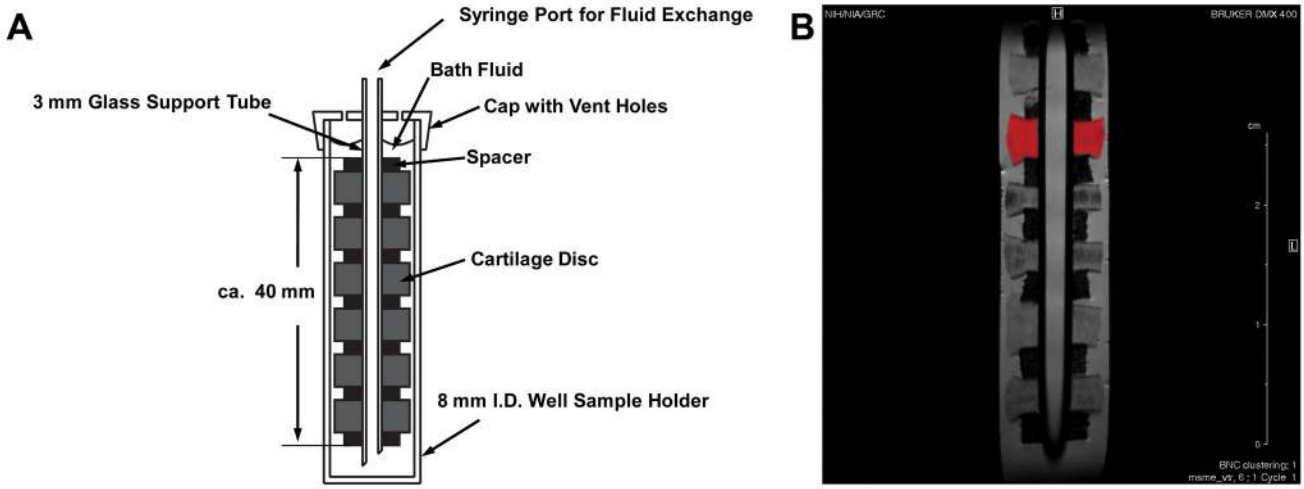

Figure 1.

(A) Cross-sectional diagram of a single well of the 4-well sample holder for BNC samples. The samples were spaced to permit contact with bath fluid on all external surfaces. (B) A T_1 image of BNC samples used for delineation of ROIs, as indicated, within the 4-well sample holder following 24-hr trypsin digestion.

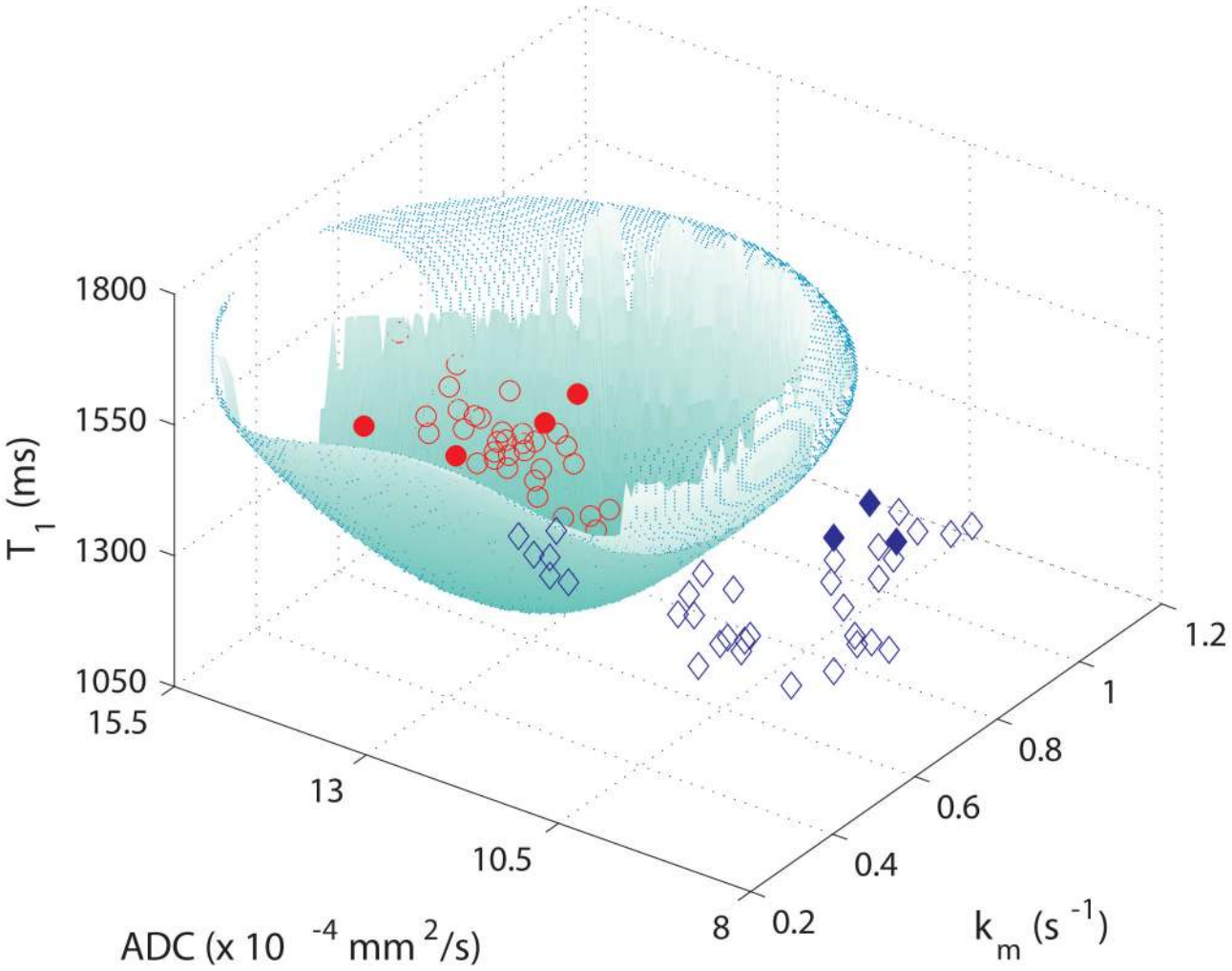

Figure 2.

Scatter plots of MRI measurements collected on pre- and post-24-hr enzymatically digested BNC samples. A single trial of the 10-fold cross-validation is shown, with the division of the full set of data points into a training set (empty symbols) and a validation set (filled symbols) for classification. Each validation set data point was classified into the pre-degraded class (◇ or ◆) or the post-degraded class (○ or ●) through the SVM model established using the training set. Results were obtained with a Gaussian kernel, $K(X, X')$= e x−‖pX−(X'‖^2/σ), with parameter values C = 16 and σ = 2 (see text) in (T_1, k_m, ADC) space. The contour surface shown indicates the decision hypersurface on which the assignment probabilities to pre- and post-24-hr degraded classes are the same. Note that all data points from the pre-degraded group are located outside the bowl-shaped hypersurface, while all the post-degraded class points are inside the hypersurface; therefore, no classification errors occur for this particular trial.

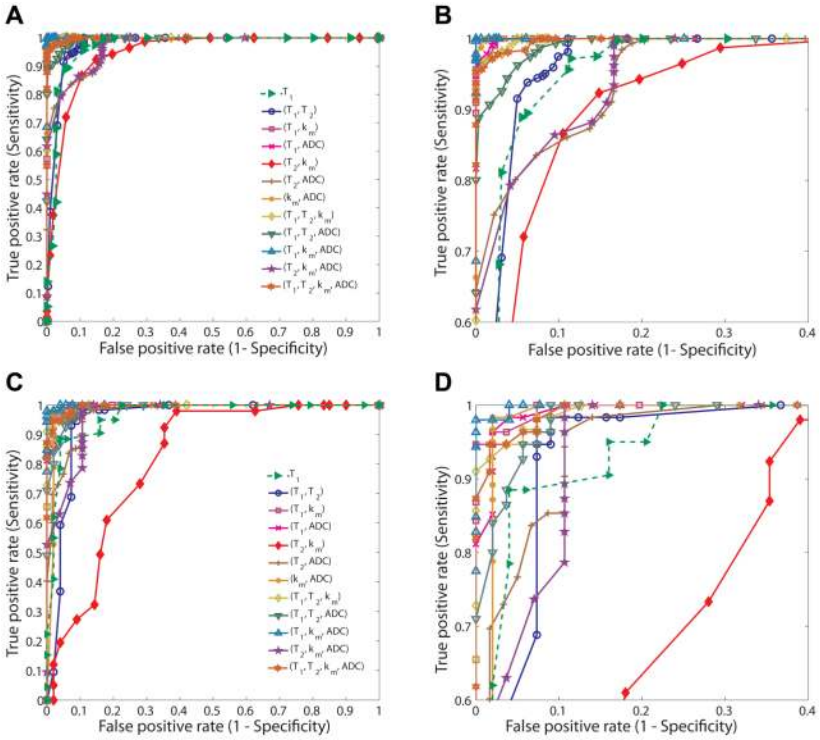

Figure 3.

ROC curves for evaluating the quality of classification of BNC samples into pre- and post-24-hr degradation groups, with the latter representing a high degree of degradation. Classification was based on the SVM algorithm for the multivariate MRI classifiers and on arithmetic means for the univariate MRI classifier. Each point on the ROC curves was calculated from the averaged classification results for the training sets (Panels A and B) or the validation sets (Panels C and D) in the 10-fold cross-validation. Panels B and D show enlargements of the regions between 0 – 0.4 on the abscissa and between 0.6 – 1 on the ordinate axes in Panels A and C, respectively. Twelve ROC curves corresponding to the eleven multivariate MRI classifiers and the best univariate T_1 classifier are illustrated in each panel. The areas under the ROC curves for the validation sets are shown in Table 2.

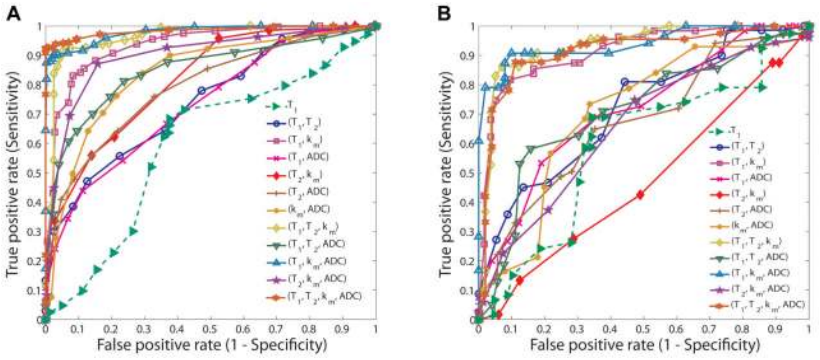

Figure 4.

ROC curves for evaluating the quality of classification of BNC samples into pre- and post-mild trypsin degradation groups, using data from Ref. (6). The procedures for classification and ROC curve construction were as described in Fig. 2, with areas under the ROC curves shown in Table 2 as well.

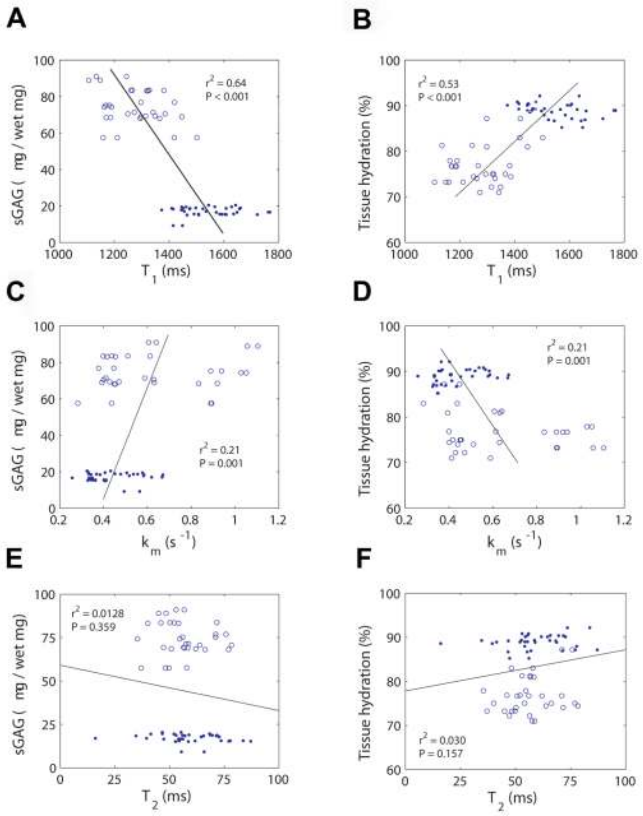

Figure 5.

Relationships between the biochemical content of BNC and corresponding MRI parameters for control samples and samples following 24-hr trypsin degradation; n=36, with each sample evaluated before (open circles) and after (filled circles) degradation. Panels A – B: T_1 vs. sGAG concentration per wet weight and tissue hydration. Panels C – D: k_m vs. sGAG concentration per wet weight and tissue hydration. Panels E – F: T_2 vs. sGAG concentration per wet weight and tissue hydration.

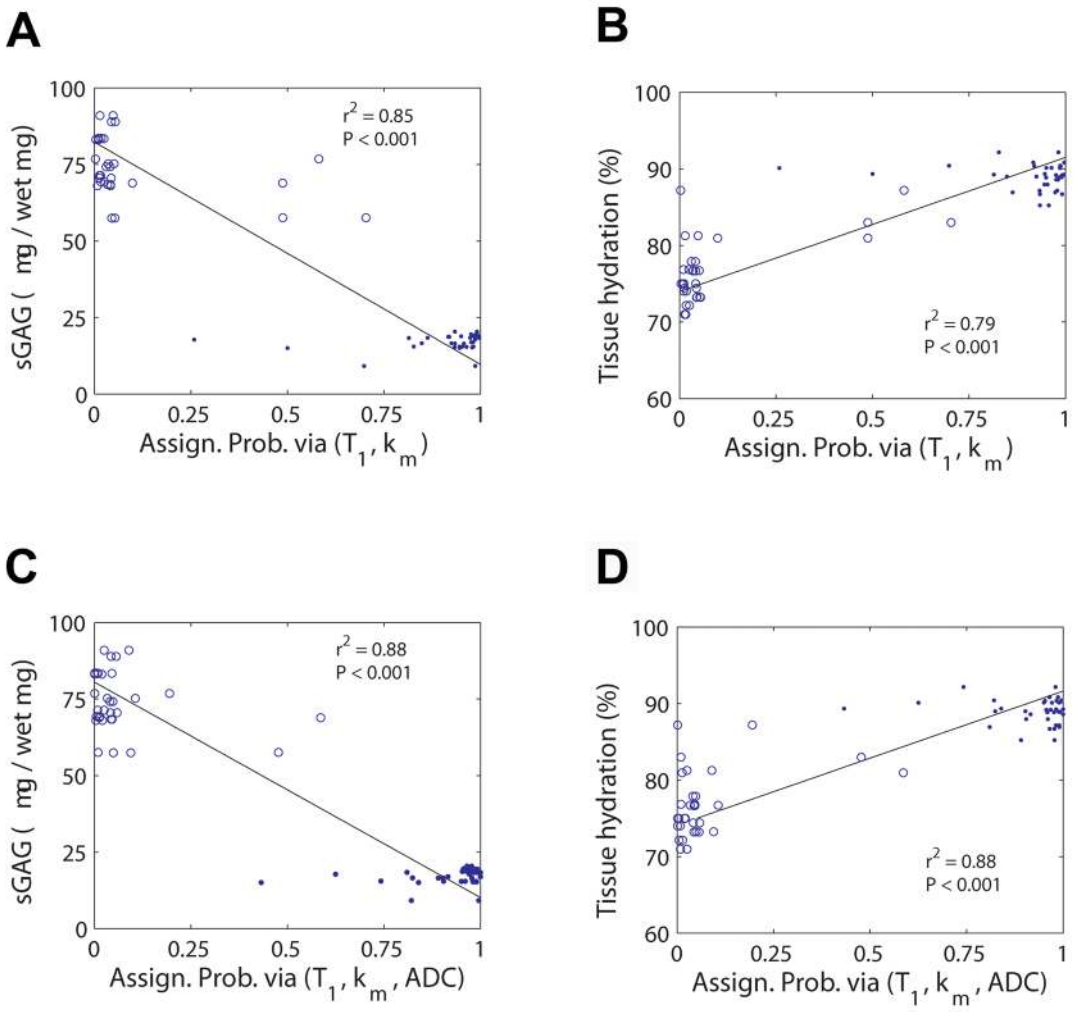

Figure 6.

Relationships between the biochemical content of pre- and post-24 hr trypsin-degraded BNC samples and their corresponding classification probabilities for the validation sets of the 10-fold cross-validation. Analysis was based on the two indicated MRI parameter combinations, (T_1, k_m) and (T_1, k_m, ADC), as described in the text. Each plot contains data from all 36 samples, before (open circles) and after degradation (filled circles). Panels A – B: sGAG concentration per wet weight and tissue hydration vs. assignment probability derived from the parameter set (T_1, k_m). Panels C – D: sGAG concentration per wet weight and tissue hydration vs. assignment probability derived from the parameter set (T_1, k_m, ADC).

Table 1

Biochemical characteristics and MRI outcome measures of pre- and post- 24-hr trypsin-degraded bovine nasal cartilage samples

	Pre-degradation	Post-degradation	p value
			Paired t-test
sGAG (μg/wet mg)	74.23 ± 9.86	16.98 ± 2.57	< 0.001
Collagen (μg/wet mg)	76.43 ± 5.79	138.71 ± 9.86	< 0.001
H_2O (%)	76.78 ± 4.36	88.94 ± 1.69	< 0.001
T_1 (ms)	1278.1 ± 100.7	1544.2 ± 102.0	< 0.001
T_2 (ms)	56.44 ± 11.10	58.97 ± 14.16	0.036
k_m (s^{-1})	0.63 ± 0.25	0.44 ± 0.11	*< 0.001
ADC (x 10^{-4} mm^2/s)	10.06 ± 0.89	12.45 ± 0.81	< 0.001

N = 36 samples for each measurement, with each sample analyzed before and after degradation

Table 2

Areas under the ROC curves for the validation sets

Parameter(s) used for deriving ROC	Extensive (24 hr) Degradation		Mild (6 hr) Degradation	
	Training set	Validation Set	training set	Validation set
T_1	0.969	0.965	0.600	0.600
T_2	0.538	0.434	0.513	0.465
k_m	0.541	0.535	0.512	0.438
ADC	0.965	0.975	0.580	0.567
(T_1, T_2)	0.976	0.950	0.742	0.718
(T_1, k_m)	0.999	0.997	0.936	0.915
(T_1, ADC)	0.999	0.995	0.734	0.715
(T_2, k_m)	0.946	0.813	0.814	0.471
(T_2, ADC)	0.970	0.971	0.791	0.673
(k_m, ADC)	0.999	0.985	0.838	0.704
(T_1, T_2, k_m)	0.999	0.995	0.963	0.932
(T_1, T_2, ADC)	0.995	0.989	0.849	0.711
(T_1, k_m, ADC)	1.000	0.999	0.980	0.940
(T_2, k_m, ADC)	0.970	0.965	0.901	0.657
(T_1, T_2, k_m, ADC)	0.998	0.995	0.984	0.913

Calculated areas under ROC curves shown in Fig. 3 (columns two and three) and Fig. 4 (columns four and five)

Table 3

Multivariate classification analysis of Control vs 6-hr trypsin degraded BNC samples

Method	Parameter	Training set		Validation set	
		Sensitivity	Specificity	Sensitivity	Specificity
Arithmetic means *	T_1	0.67 ± 0.05	0.63 ± 0.06	0.67 ± 0.12	0.62 ± 0.13
	(T_1, T_2)	0.70 ± 0.11	0.60 ± 0.15	0.68 ± 0.14	0.58 ± 0.16
	(T_1, k_m)	0.89 ± 0.07	0.73 ± 0.08	0.86 ± 0.14	0.75 ± 0.11
	(T_1, ADC)	0.68 ± 0.10	0.65 ± 0.14	0.67 ± 0.14	0.63 ± 0.16
	(T_2, k_m)	0.55 ± 0.35	0.53 ± 0.34	0.51 ± 0.35	0.49 ± 0.36
	(T_2, ADC)	0.64 ± 0.12	0.78 ± 0.16	0.52 ± 0.18	0.71 ± 0.21
SVM	(k_m, ADC)	0.81 ± 0.08	0.81 ± 0.06	0.65 ± 0.13	0.67 ± 0.15
	(T_1, T_2, k_m)	0.92 ± 0.04	0.90 ± 0.05	0.87 ± 0.10	0.81 ± 0.11
	(T_1, T_2, ADC)	0.70 ± 0.15	0.74 ± 0.19	0.62 ± 0.18	0.69 ± 0.22
	(T_1, k_m, ADC)	0.91 ± 0.04	0.97 ± 0.03	0.84 ± 0.10	0.86 ± 0.09
	(T_2, k_m, ADC)	0.87 ± 0.09	0.86 ± 0.09	0.62 ± 0.22	0.59 ± 0.18
	(T_1, T_2, k_m, ADC)	0.92 ± 0.03	0.94 ± 0.04	0.84 ± 0.11	0.84 ± 0.09
	(T_1, T_2)	0.75 ± 0.10	0.77 ± 0.10	0.65 ± 0.18	0.58 ± 0.18
	(T_1, k_m)	0.88 ± 0.04	0.87 ± 0.06	0.79 ± 0.16	0.79 ± 0.15
	(T_1, ADC)	0.72 ± 0.10	0.78 ± 0.10	0.60 ± 0.17	0.60 ± 0.22
	(T_2, k_m)	0.74 ± 0.11	0.71 ± 0.12	0.51 ± 0.19	0.53 ± 0.19
	(T_2, ADC)	0.67 ± 0.11	0.78 ± 0.08	0.53 ± 0.17	0.66 ± 0.18
Model-based (MCLUST) *	(k_m, ADC)	0.75 ± 0.10	0.76 ± 0.09	0.58 ± 0.17	0.53 ± 0.18
	(T_1, T_2, k_m)	0.94 ± 0.04	0.92 ± 0.06	0.78 ± 0.23	0.72 ± 0.26
	(T_1, T_2, ADC)	0.84 ± 0.12	0.89 ± 0.10	0.54 ± 0.32	0.64 ± 0.31
	(T_1, k_m, ADC)	0.94 ± 0.04	0.95 ± 0.04	0.74 ± 0.29	0.72 ± 0.26
	(T_2, k_m, ADC)	0.81 ± 0.13	0.86 ± 0.11	0.52 ± 0.31	0.57 ± 0.27
	(T_1, T_2, k_m, ADC)	0.96 ± 0.05	0.96 ± 0.05	0.48 ± 0.39	0.79 ± 0.29

* Results are from Tables 2 and 5 of Ref. (6), the discriminant analysis using univariate MRI parameter and multi-component Gaussian clusters, respectively.