Classification of Sodium MRI Data of Cartilage with Machine Learning and Logistic Regression

Guillaume Madelin[1], James Babb[1], and Ravinder R Regatte[1]
[1]*Radiology Department, New York University Langone Medical Center, New York, NY, United States*

Purpose. To compare different methods of classification based on statistical/machine learning algorithms for differentiating subjects with and without osteoarthritis (OA) from sodium MRI data of articular cartilage at 7T. Methods are compared to logistic regression.

Methods. *Sodium MRI Data* - Sodium MRI was acquired on the knee cartilage of 19 healthy volunteers and 28 OA subjects at 7T (Siemens Healthcare), with 2 acquisition sequences: one without fluid suppression - radial 3D (R3D) - and one with fluid suppression by inversion recovery (IR) applied before R3D acquisition. Fluid suppression is expected to increase the sensitivity of sodium imaging to the characteristic loss of sodium content in cartilage with OA, by eliminating partial volume effect due to the presence of fluid signal within the large voxels of the sodium images [1]. Sodium quantification was performed by linear regression from the signal of agar gel phantoms with known sodium concentrations placed on top of the knee. The mean and standard deviation of the tissue sodium concentration (TSCmean and TSCstd) were measured in 3 regions of cartilage over 4 consecutive slices for each subject and each sequence (12 values/subject/sequence for each measurement). The minimum, maximum and mean of TSCmean and TSCstd was then calculated over the 12 measurements for each subject. In this study, X-ray joint space narrowing was used as gold standard for diagnosing OA patients. *Support Vector Machine (SVM)* [2] - Parameters: linear kernel function, quadratic programming optimization. *K-Nearest Neighbor (KNN)* [3] - Parameters: Euclidean distance metric, k=3 nearest neighbors (majority rule). *Naive Bayes (NB)* [4] - Parameters: normal probability distribution, empirical class prior probability calculated from the training set. *Discriminant Analysis (DA)* [5] - Parameters: linear type, empirical class prior probability calculated from the training set. *Training/Test Datasets* - All training datasets were composed of 50% of the original data chosen randomly, and test datasets were composed of the remaining 50%. Each algorithm was applied 100 times on different training and test datasets, and the mean and standard deviation of sensitivity, specificity and accuracy was calculated for each method and each combination of sodium measurements ('true positive' values were defined as 'OA'). *Logistic regression (LR)* - Stepwise variable selection was applied to identify the combinations of factors as predictors of OA. All machine learning processing was performed in Matlab (Mathworks, Inc), and LR processing in SAS (SAS Institute).

Table 1. Four best accuracy results of each method for classification of OA and control from sodium data at 7T. Results in %. Measurement numbers are at the table bottom.

	Accuracy	Sensitivity	Specificity
Support Vector Machine (SVM)			
[3]+[4]+[7]	**80.6** ± 7.6	89.6 ± 9.8	66.7 ± 17.9
[1]+[2]+[3]+[4]+[7]	**78.8** ± 7.3	86.4 ± 10.8	67.0 ± 17.0
[1]+[3]+[4]+[7]	**78.7** ± 7.3	87.2 ± 9.6	65.6 ± 19.6
[2]+[3]+[4]+[7]	**78.3** ± 7.4	87.4 ± 9.7	64.2 ± 18.4
K-Nearest Neighbor (KNN)			
[4]+[8]+[9]+[12]	**69.7** ± 8.2	76.2 ± 11.3	59.4 ± 16.7
[1]+[4]+[5]+[6]+[8]+[9]+[11]	**69.2** ± 8.5	72.1 ± 12.5	64.6 ± 17.4
[1]+[2]+[5]+[7]+[8]+[9]	**69.1** ± 8.0	73.7 ± 12.1	61.9 ± 15.7
[2]+[5]+[6]+[9]+[11]	**69.1** ± 8.9	74.9 ± 11.9	60.1 ± 20.5
Naive Bayes (NB)			
[1]+[2]+[3]+[4]+[5]+[6]+[9]	**73.4** ± 8.4	84.4 ± 9.9	56.3 ± 19.4
[1]+[2]+[5]+[7]	**73.1** ± 8.2	84.3 ± 10.2	55.8 ± 18.5
[1]+[2]+[3]+[5]+[6]+[7]+[9]+[12]	**72.7** ± 8.4	83.8 ± 10.1	55.3 ± 18.5
[2]+[5]+[6]+[7]+[8]+[12]	**72.7** ± 7.2	84.6 ± 9.0	54.0 ± 18.9
Discriminant Analysis (DA)			
[3]+[5]+[6]+[11]	**59.2** ± 10.6	59.1 ± 14.8	59.4 ± 19.0
[1]+[2]+[9]+[17]	**59.2** ± 10.4	59.8 ± 16.0	58.2 ± 17.1
[1]+[2]+[3]+[4]+[5]+[6]+[8]+[9]	**58.9** ± 10.2	60.6 ± 16.3	56.3 ± 15.9
[1]+[2]+[4]+[5]+[6]+[7]+[8]+[9]+[10]	**58.8** ± 11.6	60.0 ± 16.3	56.9 ± 19.8
Logistic Regression (LR)			
[7]	**78.7**	96.4	52.6
[12]	**76.6**	78.6	73.7
[9]	**74.5**	75.0	73.7
[10]	**74.5**	75.0	73.7

List of measurements numbers:

[1] R3D - min(TSCmean)	**[2]** R3D - max(TSCmean)	**[3]** R3D - mean(TSCmean)
[4] R3D - min(TSCstd)	**[5]** R3D - max(TSCstd)	**[6]** R3D - mean(TSCstd)
[7] IR - min(TSCmean)	**[8]** IR - max(TSCmean)	**[9]** IR - mean(TSCmean)
[10] IR - min(TSCstd)	**[11]** IR - max(TSCstd)	**[12]** IR - mean(TSCstd)

Results and Discussion The best accuracy results are presented in Table 1. An accuracy of 75-80% can be obtained with SVM and LR, while NB, KNN and DA generate much less efficient classifications (with the parameters used in this study). LR generate its best results by using only one measurement - min(TSCmean) or mean(TSCstd) from IR (fluid suppressed data) - compared to all machine learning algorithms (3 measurements or more are needed). SVM can slightly increase the accuracy of the classification compared to LR by combining data from both sequences (with and without fluid suppression), but at the expense of much lower specificity (64-67%) compared to sensitivity (86-90%). Except for the case with min(TSCmean) from IR where specificity is low and sensitivity very high, LR can classify the data with similar sensitivity and specificity (74-79%).

Conclusion. SVM and LR can classify OA and controls from sodium MRI data at 7T with an accuracy of 75-80%. SVM can classify the data with higher sensitivity on average, but lower specificity compared to LR. Further improvements of both methods can include a combination with principal component analysis (PCA), optimization of training and test datasets (for SVM) and acquisition of more data (in progress). Different parameters for KNN, NB and DA combined with PCA are also under investigation for accuracy improvement.

References. [1] Madelin G et al. JMR 207, 42-52, 2010. [2] Cortes C et al. Machine Learning 20(3), 273-297, 1995. [3] Cover T et al. IEEE Trans Inf Theory 13(1), 21-27, 1967. [4] Lewis D. Machine Learning:ECML-9, 4-15, 1998. [5] Fisher RA. Annual Eugenics 7, 179-188, 1936.
Funding. NIH 1R01AR053133, 1R01AR056260, 1R01AR060238