

Optical tomographic detection of rheumatoid arthritis with computer-aided classification schemes

Christian D.Klose^a, Alexander D.Klose^b, Uwe Netz^c, Jürgen Beuthan^c, Andreas H.Hielscher^a

^aDepartment of Biomedical Engineering, Columbia University, New York, USA;

^bDepartment of Radiology, Columbia University, New York, USA;

^cInstitut für Medizinische Physik und Lasermedizin, Charité - Universitätsmedizin Berlin, Germany

ABSTRACT

A recent research study has shown that combining multiple parameters, drawn from optical tomographic images, leads to better classification results to identifying human finger joints that are affected or not affected by rheumatic arthritis RA. Building up on the research findings of the previous study, this article presents an advanced computer-aided classification approach for interpreting optical image data to detect RA in finger joints. Additional data are used including, for example, maximum and minimum values of the absorption coefficient as well as their ratios and image variances. Classification performances obtained by the proposed method were evaluated in terms of sensitivity, specificity, Youden index and area under the curve AUC. Results were compared to different benchmarks ("gold standard"): magnet resonance, ultrasound and clinical evaluation. Maximum accuracies (AUC=0.88) were reached when combining minimum/maximum-ratios and image variances and using ultrasound as gold standard.

Keywords: rheumatoid arthritis, classification, multi-parameter, computer-aided

1. INTRODUCTION

Over the last decade diffuse optical tomography (DOT) has increasingly been employed in preclinical and clinical studies, mainly in breast cancer, joint diseases, and brain imaging. While substantial advances have been made in instrumentation and image reconstruction algorithms, much less effort has been spent on developing image analysis tools. Other medical imaging fields such as magnet resonance imaging (MRI) and computer tomographic imaging (CT) frequently make use of advanced image analysis methods that enhance sensitivity and specificity in many cases. For example computer aided diagnostics (CAD) systems have been successfully employed in mammography.¹ In biomedical optics CAD has only been applied in two studies related to Optical Coherence Tomography (OCT) that explored its utility in esophageal and cervical cancer.^{2,3} To the best of our knowledge, no studies have been presented where CAD was employed in the analysis of DOT images.

In previous studies our group has evaluated the potential of using features such as minimal and maximal absorption or scattering coefficients or their ratios in an image region of interest, to distinguish between affected and not affected joints.⁴ Using the minimal absorption coefficient as a parameter, sensitivities and specificities of about 0.71 could be achieved in identifying affected joints assuming that ultrasound can be considered as a "gold standard". Subsequently we showed that combining several parameters increased sensitivities and specificities could be achieved.⁵ To deal with the problem of multi-parameter classification we employed a neural-network based interpretation method that explores and sorts through image data.⁶ Classifications were performed by an unsupervised artificial intelligence method called Self-Organizing Mapping (SOM).^{7,8} This method has been used in the past in other scientific fields for similar classification problems.^{6,9,10} It has shown to produce significant better results than approaches based on discriminant analysis and logistic regression.^{11,12}

Here, we go beyond our previous analysis in several ways. First, we consider image parameters that were previously not included in the analysis, such as the variance of optical properties in the images. Also, instead of

Further author information: (Send correspondence to C.D.Klose)
C.D.Klose: E-mail: ck2204@columbia.edu, Telephone: 1 212 854 2320

using only ultrasound as gold standard or ground truth, the classification performance is evaluated using additional gold standards including magnetic resonance imaging, clinical evaluation and visual inspection of optical tomographic imaging itself. Finally, a larger data set is used, that included most recent optical measurements of 100 finger joints.

In the following we will first describe the data used for the analysis. This is followed by a description of the neural-network based classification approach used in this work. Subsequently the results obtained with this approach are presented and discussed.

2. DATA AND METHODS

2.1 Data and instrumentation

Data sets resulting from tomographic reconstructions of sagittal laser optical tomographic (SLOT) images were analyzed to determine best image interpretation results. SLOT images were obtained from two-dimensional sagittal cross sections through the proximal interphalangeal (PIP) finger joint.^{13–18} These images were generated by measuring the transmitted light intensities along the central axis of the index, middle and ring fingers on the left and right hand. The light source was a laser with wavelength $\lambda = 630\text{nm}$, which was focused to 1mm spot on 11 different positions on the back of each finger. For each position the transmitted light intensities were measured with a Si photo diode. This transmission data became input to a model-based iterative image reconstruction code that used the equation of radiative transfer as light propagation model.^{13–17}

In total, 100 optical tomographic images of human finger joints were used in this study. To prepare the images for CAD analysis, a region of interest (ROI) was defined within each image. Data were eliminated in the first 4 mm on the top and bottom of each image and 7 mm on the left and right. In this way the chosen ROI did not contain potential image artifacts, which are often encountered near source and detector positions (image boundaries). Within the ROI, different parameters of the absorption coefficients were extracted including the smallest value $\min(\cdot)$, mean value $\text{mean}(\cdot)$, largest value $\max(\cdot)$, mutual ratios (e.g., $\min(\cdot)/\text{mean}(\cdot)$) and statistical variance $\text{var}(\cdot)$. All extracted image features were combinatorically combined. Thus, each image was characterized by an n -dimensional feature vector \vec{x}_n consisting of a set of n image features/parameters. These feature vectors became input to an artificial neural network that classified each image as RA affected or not affected finger. Figure 1 shows cross plots of the image parameters.

To perform this analysis, one needs a gold standard or ground-truth benchmark, that identifies each patient as affected or not affected by RA. In previous works we used a gold standard based on ultrasound images.⁴ However, a majority of researchers currently consider magnetic resonance images of finger joints as the most accurate indicator for RA. However no studies have been presented that conclusively show that MRI is indeed the most accurate ground truth. This would require longitudinal studies spanning many year of follow up to establish the predictive and prognostic value of each imaging method. A study like this has not been performed yet. Therefore, we decided to report on the performance of our CAD system for different gold standards, including MRI, US, clinical evaluation, and optical inspection of SLOT images. For each modality experts were scoring the images and data on a 4 point score: 0 for definitely no synovitis, 1 for probably no synovitis, 2 for possibly synovitis and 3 for definitely synovitis. Subsequently each finger was labeled by only two different classes: class c_0 *not affected* = definitely and probably no synovitis and class c_1 *affected* = possibly and probably synovitis (Fig. 2). The feature vector $\vec{x}_{\{n,c\}}$ containing the various optical parameters was labeled accordingly and the performance of the CAD system evaluated.

2.2 Methodology

To deal with the problem of multi-parameter classification, a neural-network-based method of Self-Organizing Mapping (SOM) was employed as part of an agent-based unsupervised data exploration tool. In general, a SOM network is structured in two layers: an input layer and a Kohonen layer. The input layer is a one-on-one representation of the feature vector. The Kohonen layer represents a structure with a single 2-dimensional map (lattice) consisting of neurons arranged in rows and columns. Each neuron of this discrete lattice is fixed and is fully connected with all source neurons in the input layer.

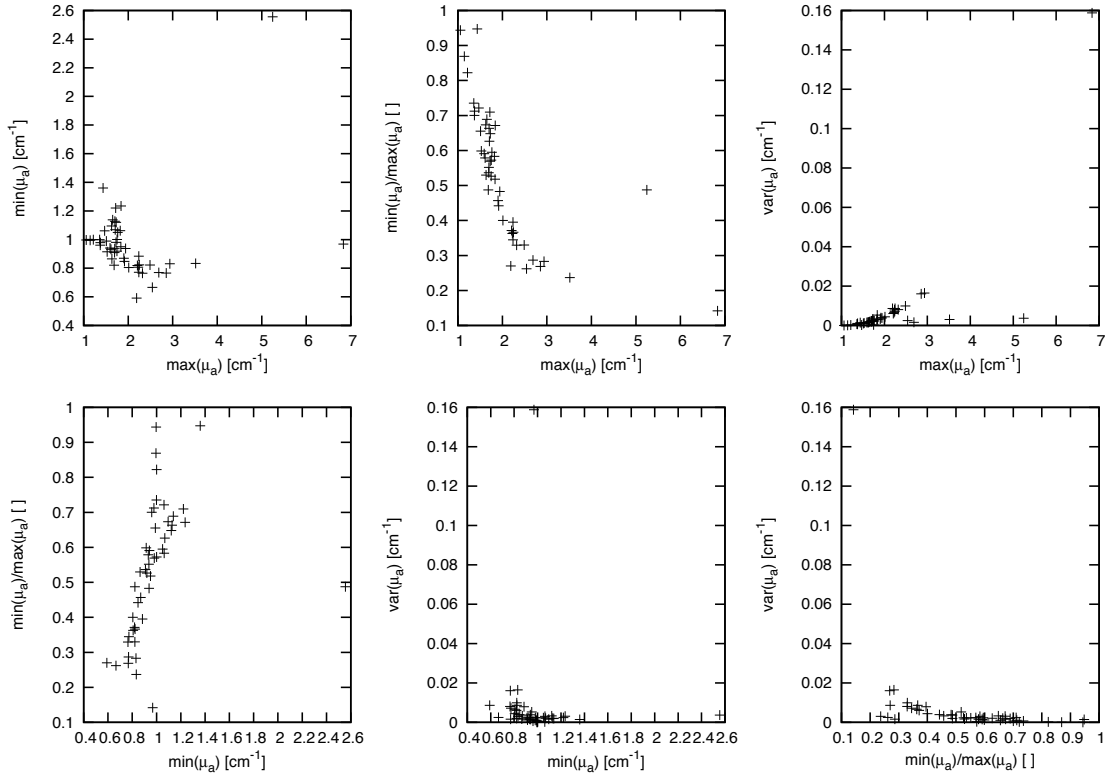


Figure 1. Cross plots of the image parameters. Each cross plot visualizes the raw data extracted from the optical tomographic images.

For the given task of interpreting three-dimensional optical data, each feature vector, which is presented to neurons of the input layer, typically activates (stimulates) a neuron in the Kohonen layer. Based on the given input data set, learning occurs during the self-organizing procedure as feature vectors, are presented to the input layer of the network. Neurons of the Kohonen layer compete to see which neuron will be activated by an input data vector. The weights between vectors and are used to determine only one activated neuron in the Kohonen layer after the winner-takes-all principle. After the learning process the SOM is considered as trained and its weights \vec{w}_m store the interrelations of all feature vectors \vec{x}_n . A more detailed overview about the learning process on tomographic image data can be found in Klose (2006).⁶

To illustrate the performance of such SOM network in our particular application, we consider the following example. Fig. 2 shows the distribution of a 2-dimensional feature vector, with the *variance*(μ_a) as first component (x-axis) and $\min(\mu_a)/\max(\mu_a)$ as the second component (y-axis). The red squares identify the feature vectors belonging to images of affected joints, while the blue circles identify the feature vectors belonging to images of not-affected joints, as determined by using US as ground truth. One can see that if one would attempt to classify a feature vector (representing an image) as affected or not affected using only threshold values for either $\text{var}(\mu_a)$ or $\min(\mu_a)/\max(\mu_a)$, would lead to a large number of miss-classifications. For example, postulating that all fingers with $\text{var}(\mu_a) > 0.2$ are affected would lead to 3 false negative and 17 true negative respectively. Similarly postulating that all finger with $\min(\mu_a)/\max(\mu_a) > 0.2$ are affected would lead to 12 false negative and 8 true negative. Hence, classifications on only one parameter would be highly flawed.

Therefore SOM methods uses a vector quantization approach to partitioning a given data set \vec{x}_n into m sub-regions (clusters). The size of the clusters is a variable of a given SOM network and will itself influence the classification of the performance of the network. When choosing different cluster size q and threshold p_t , will result in different classification results. In general, classification performance increases as the cluster size gets smaller (which is equivalent to the number of sub-groups getting larger) until an optimum is reached when mis-

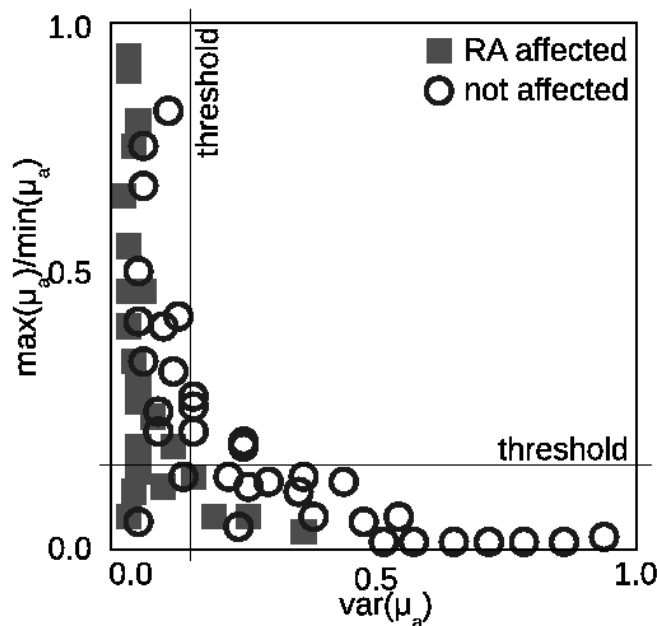


Figure 2. Schematic illustration of a clustering problem to identify RA affected finger joints. It can be seen the data distribution in a 2-dimensional feature space and the assignment of affected and not affected finger joints based on the ultra-sound benchmark. Single parameter classifications using parameter thresholds lead to miss-interpretations. This can be reduced when using multi-parameter classifications.⁵

classification is at a minimum. Thus, a too small number of clusters leads to "under-fitting" of a given data set whereas a too large number may lead to a data "over-fitting". In both cases, mis-classification is increasing.^{6,19}

In this study the cluster number varied from 1 to 100 and p_t from 0 to 100. For each combination classification performance was evaluated by using a leave-one-out approach. The n -dimensional given data manifold of $L = 100$ realizations was randomly split into q disjoint sub-sets (e.g., $q=10$). The accuracy was determined when performing a SOM classification with $q-1$ of q sub-sets (training step) and applying it to the remaining 1 of q sub-sets (validation step). By leaving one sub-set out, the procedure was conducted p times and the mean and standard deviation of various performance measures were calculated.

To quantify the classification performance the following measures were considered sensitivity S_e , specificity S_p , Youden index $J = S_e + S_p - 1$ and the area under the curve AUC . By varying p_t from 0 to 100 per cent receive operator curves (ROC) can be generated and analyzed as they are frequently used in the characterization of medical classification schemes. If p_t is set to 0, all images will be qualified as not affected leading to a sensitivity of 0, however, specificity is 1. If p_t is set to 100 percent the specificity will be 0. Intermediate p_t values lead to intermediate S_e and S_p values usually maximizing the J for a given pair. It should be pointed out that this approach differs from classical ROC analysis, which is typically applied to only one observable parameter, e.g., $\min(\mu_a)$.⁴ By varying the threshold of μ_a for which a patient is considered affected S_e and S_p can be calculated and ROC curves generated. By introducing p_t as threshold, we effectively extended the ROC analysis to multiple parameter interpretation in the frame work of SOM.

3. RESULTS AND DISCUSSION

First classification performances of the network were evaluated using different sets of optical parameters. Figure 3 shows the changes of J as a function of the frequency thresholds p_t for 11 different parameter configuration. Results are shown for 4 different benchmarks ("ground truth"), including clinical evaluation (CL), magnetic

resonance imaging (MRI), ultrasound (US) and visual SLOT interpretation by experts. The error bars in these figures represent the prediction errors (standard deviations), which result from the cross validation methods described in the previous section. Furthermore, to arrive at these particular error bars, the SOM neural network size (number of clusters) also varied for each parameter combination.

Overall, different combinations of parameters lead to classification results with varying Youden index J (sensitivity and specificity) and AUC . Combinations of parameters that lead to the highest Youden Index are given in (Table 1). Two-dimensional parameter pairs, yield the best results. All combinations involving 3 or more parameters resulted in worse performances. Looking at Figure 3 several significances can be observed. Parameter combinations of only 2 features ($min(\mu_a)/max(\mu_a), var(\mu_a)$ or $max(\mu_a), min(\mu_a)$) lead to higher accuracy measures than 3 or 4 feature combinations (shown in gray). One reason for this is that increasing the dimensionality of the parameter space may not necessarily increase the discrimination of the classes affected and not affected.

Table 1. Results of the neural network based classification of all 2-parameter combinations that have shown best classification performances including sensitivity S_e , specificity S_p , the resulting Youden index $J = S_e + S_p - 1$ and the area under the curve AUC.

gold standard	data vector $\vec{x} = \{ \}$	S_e	S_p	J	AUC
CL	$max(\mu_a), min(\mu_a)$	0.99	0.71	0.70	0.44
MRI		1.00	0.73	0.73	0.44
US		0.97	0.82	0.79	0.75
SLOT		0.95	0.81	0.76	0.73
CL	$max(\mu_a), min(\mu_a)/max(\mu_a)$	0.97	0.82	0.79	0.54
MRI		0.99	0.81	0.80	0.55
US		0.95	0.89	0.84	0.81
SLOT		0.92	0.92	0.84	0.84
CL	$max(\mu_a), var(\mu_a)$	0.99	0.75	0.74	0.46
MRI		0.99	0.73	0.72	0.44
US		0.97	0.80	0.77	0.72
SLOT		0.96	0.84	0.80	0.77
CL	$min(\mu_a), min(\mu_a)/max(\mu_a)$	0.91	0.82	0.73	0.53
MRI		0.91	0.83	0.74	0.53
US		0.91	0.87	0.78	0.78
SLOT		0.82	0.97	0.79	0.82
CL	$min(\mu_a), var(\mu_a)$	0.96	0.79	0.75	0.58
MRI		0.96	0.81	0.77	0.55
US		0.89	0.91	0.80	0.80
SLOT		0.89	0.92	0.81	0.82
CL	$min(\mu_a)/max(\mu_a), var(\mu_a)$	0.98	0.85	0.83	0.65
MRI		0.96	0.85	0.81	0.60
US		0.96	0.91	0.87	0.86
SLOT		0.94	0.95	0.89	0.88

The combination of relative features, such as for instance, $min(\mu_a)/max(\mu_a)$ and $var(\mu_a)$ shows best prediction accuracies for all benchmarks. Particularly, when compared to visual interpretation results of the SLOT images where accuracies are maximum $J=0.89$. This may explain, that relative values among many images can be visually easier interpreted by eyes than absolute values (e.g., $min(\mu_a), max(\mu_a)$). Figure 3 also shows that the combination of $min(\mu_a)/max(\mu_a)$ and $var(\mu_a)$ is best when compared to US, MRI images and to the clinical diagnostics CL. It is interesting to note that the curves generated with US as "ground truth", look similar to curves generated with SLOT as "ground truth". The associated S_e and S_p values are all larger than 0.85 in this range. This suggest that US and SLOT are similar in the assessment of RA in finger joints.

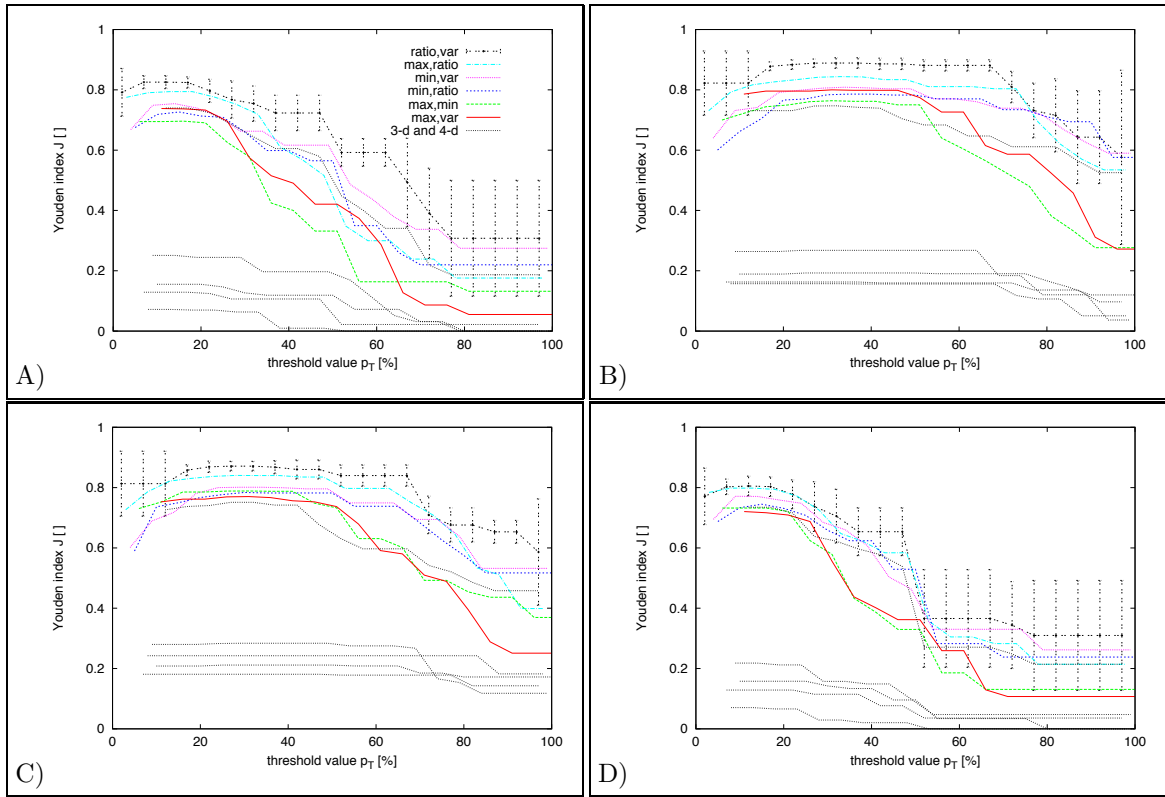


Figure 3. Youden index J as interpretation accuracy to identifying "arthritis affected" finger joints. Results show different image feature combinations with respect to "ground truth"-benchmarks (A-clinical diagnostics, B-visual SLOT image interpretation, C-ultrasound, D-magnet resonance). Feature combinations based on 2 features show higher J -values that base on 3 and 4 features (see key). Error bars are given only for the most reliable features. They result from uncertainties due to different SOM neural network sizes and the cross validation.

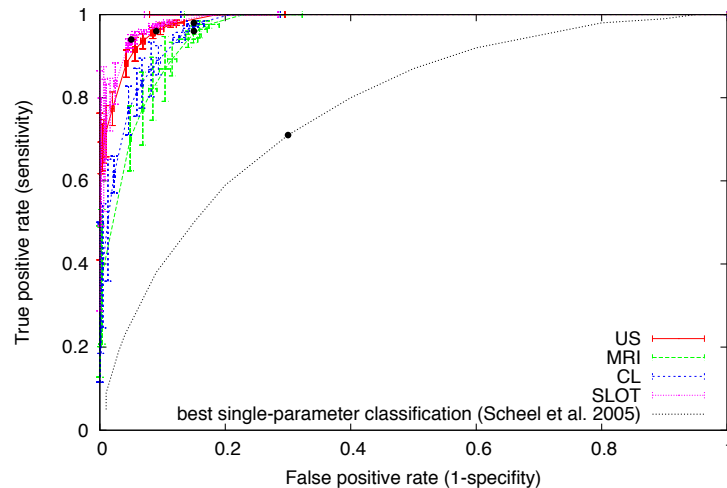


Figure 4. Sensitivity-specificity curves (ROC curves) illustrating image interpretation results based on the combination of $\min(\mu_a)/\max(\mu_a)$ -ratio and $\text{variance}(\mu_a)$. The curves show the best classification performances using J (black dots). The error bars of the curves result from cross validations. The results of all "ground truth" benchmarks are better when compared to the best single-parameter classifications reported by Scheel et al.

Using MRI as "ground truth", the highest Youden index was achieved with p_t values of 13 per cent. The associated sensitivities S_e and specificities S_p are 0.96, and 0.85, which is considerably higher than the values reported by Scheel et al. Relying on a single parameter ($\min(\mu_a)$) they reported $S_e = 0.71$ and $S_p = 0.71$. Finally, ROC curves are shown in Fig. 4, for the best cases found in this study with MRI, US, CL, and SLOT as ground truth.

Table 2 shows classification results based on discriminant analyses DA which show lower classification accuracies of class separabilities than based on the presented computer-aided classification approach in Table 1. The Youden index resulting from DA tends to be 50 per cent smaller than from SOM. However, DA confirms that two parameter combinations show better classification performances than 3 and 4 parameter combinations.

Table 2. Results of the traditional discriminant analysis with respect to different parameter combinations and the MRI and US ground truth. The table shows the sensitivity S_e , specificity S_p , Youden index $J = S_e + S_p - 1$ and the F-values. The critical F-value for a statistical significance of $\alpha=5\%$ is 4.15. Obtained F-values need to exceed this critical value in order to discriminate the affected and not affected groups with statistical significance.

data vector $\vec{x} = \{ \}$	MRI				US			
	S_e	S_p	J	F	S_e	S_p	J	F
$\min(\mu_a)$	0.2	0.92	0.13	3.68	0.80	0.47	0.27	18.13
$\max(\mu_a)$	0.12	0.88	0.00	0.01	0.75	0.34	0.09	0.73
$\min(\mu_a)/\max(\mu_a)$	0.17	0.93	0.10	2.85	0.81	0.40	0.20	12.26
$\text{var}(\mu_a)$	0.33	0.93	0.26	13.08	0.81	0.70	0.51	27.24
$\min(\mu_a), \min(\mu_a)$	0.22	0.93	0.15	6.05	0.82	0.50	0.32	11.37
$\min(\mu_a), \min(\mu_a)/\max(\mu_a)$	0.21	0.92	0.13	0.81	0.80	0.47	0.27	8.98
$\min(\mu_a), \text{var}(\mu_a)$	0.16	0.93	0.24	1.09	0.87	0.61	0.47	20.20
$\max(\mu_a), \min(\mu_a)/\max(\mu_a)$	0.19	0.92	0.11	5.18	0.85	0.49	0.34	11.22
$\max(\mu_a), \text{var}(\mu_a)$	0.37	0.93	0.30	0.07	0.81	0.74	0.55	13.80
$\min(\mu_a)/\max(\mu_a), \text{var}(\mu_a)$	0.32	0.93	0.25	1.35	0.85	0.62	0.47	19.15
$\min(\mu_a), \max(\mu_a), \min(\mu_a)/\max(\mu_a)$	0.21	0.92	0.13	0.16	0.83	0.50	0.33	8.55
$\min(\mu_a), \max(\mu_a), \text{var}(\mu_a)$	0.32	0.93	0.24	1.26	0.85	0.59	0.44	14.43
$\min(\mu_a), \min(\mu_a)/\max(\mu_a), \text{var}(\mu_a)$	0.32	0.93	0.25	0.01	0.87	0.60	0.47	13.42
$\max(\mu_a), \min(\mu_a)/\max(\mu_a), \text{var}(\mu_a)$	0.42	0.97	0.39	3.66	0.87	0.64	0.51	15.89
$\min(\mu_a), \max(\mu_a), \min(\mu_a)/\max(\mu_a), \text{var}(\mu_a)$	0.44	0.97	0.41	0.01	0.43	0.97	0.40	12.03

4. CONCLUSIONS

The article presents an advanced computer-aided diagnostic (CAD) methods to detect rheumatic arthritis in human finger joints. In particular an unsupervised interpretation system was employed based on Self-Organizing Maps (SOM) to distinguish between finger joints affected and not affected by rheumatoid arthritis. Different parameters (e.g., smallest and largest absorption and respective ratios) drawn from sagittal laser optical tomography (SLOT) images became input to the CAD algorithm, and Youden index, specificity and sensitivity, and area under the curve were used as classification performance measures. They were calculated for 4 different benchmarks ("gold standards"), including MRI, US, clinical evaluation and optical inspection of SLOT images.

Combining the minimal-maximum ratio of absorption coefficient and the variance in the image, specificities and sensitivities of 0.96 and 0.85 could be achieved, assuming MRI provides a good ground truth. If US is chosen as ground truth, these values increase to 0.96 and 0.91. At these levels of specificity and sensitivity optical tomographic reconstruction methods are an attractive tool for the evaluation of arthritis in finger joints.

4.1 Acknowledgments

The authors thank Prof. Scheel, University of Göttingen, Germany for providing experimental data used in this analysis. This work was supported in part by a grant (2R01 AR46255) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), which is part of the National Institutes of Health.

REFERENCES

- [1] Meinel, L.A., A.H. Stolpen, K.S. Berbaum, L.L. Fajardo, J.M. Reinhardt (2006) Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system, *Journal of Magnetic Resonance Imaging* 25(1), 89 - 95.
- [2] Qi X, Sivak M, Insberg G, Willis JE, Rollins AM, (2006) Computer-aided diagnostics of dysplasia in Barrett's esophagus using endoscopic optical tomography, *Journal of Biomedical Optics*, 11(4).
- [3] Bazant-Hegemark F.,N. Stone, M.D. Read, K. McCarthy, R.K. Wang (2006) Optical coherence tomography (OCT) imaging and computer aided diagnosis of human cervical tissue specimens, *Proc. SPIE* 6627, 66270F.
- [4] Scheel A K, Backhaus M, Klose A D, Moa-Anderson B, Netz UJ, Hermann K-GA, Beuthan J, Mller GA, Burmester GR, Hielscher AH (2005) First clinical evaluation of sagittal laser optical tomography for detection of synovitis in arthritic finger joints, *Ann Rheum Dis* 64: 239-245
- [5] Klose C.D., Klose A.D., Beuthan J., Hielscher A. (2006). Multi-parameter classifications of optical tomographic images. *Journal of Biomedical Optics*, 13(5), 050503.
- [6] Klose C.D. (2006). Self-Organising Maps for Geoscientific Data Analysis: Geological Interpretation of Multi-dimensional Geophysical Data, *Computational Geosciences*, 10(3), 265-277.
- [7] Kohonen T., (1982). Self-organizing formation of topologically correct feature maps, *Biol. Cyb.* 43(1), 59-69.
- [8] Kohonen T., (2001). *Self-Organizing Maps*, 3rd edition, Springer, Berlin.
- [9] Pascual-Montano, A Taylor KH, Winkler H, Pascual-Marqui RD, Carazo J-M. Quantitative self-organizing maps for clustering electron tomograms, *Journal of Structural Biology*, 138, 114122, (2002).
- [10] Nattkemper TW, Wismüller A, Tumor feature visualization with unsupervised learning, *Medical Image Analysis*, 9, 344351, (2005).
- [11] Schönweiler R, Wübbelt P, Tolloczko R, Rose C, Ptok M. Classification of Passive Auditory Event-Related Potentials Using Discriminant Analysis and Self-Organizing Feature Maps, *Audiology and Neurotology*, 5, 69-82, (2000).
- [12] Veltri R W, Chaudhari M, Miller MC, Poole EC, ODowd GJ and Partin AW. (2002) Comparison of Logistic Regression and Neural Net Modeling for Prediction of Prostate Cancer Pathologic Stage, *Clinical Chemistry* 48, 10, 18281834.
- [13] Klose AD Beuthan J Müller G, (1997) Investigations of RA-diagnostics applying optical tomography in frequency domain, In: *Optical and Imaging Techniques for Biomonitoring*, H.J. Foth, R. Marchesini, and H. Podbielska, eds., *SPIE Proc*; 3196, 194-204.
- [14] Klose AD, Hielscher AH, Hanson KM, Beuthan J. (1998) Two and three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis. *Proc SPIE Int Soc Opt Eng*; 3566, 15160.
- [15] Klose AD (2002) *Optical Tomography Based on the Equation of Radiative Transfer*, Ph.D. Thesis, Freie Universitt Berlin, Germany, ISBN 3-89825-565-4, <http://www.diss.fu-berlin.de/2002/135/indexe.html>
- [16] Klose AD, Netz U, Beuthan J, Hielscher AH (2002) Optical tomography using the time-independent equation of radiative transfer. Part I: Forward model, *J Quant Spectrosc Radiat Transfer*; 72(5), 691-713.
- [17] Klose AD, Hielscher AH (2002) Optical tomography using the time-independent equation of radiative transfer. Part II: Inverse model, *J Quant Spectrosc Radiat Transfer*; 72(5), 715-732.
- [18] Hielscher AH, Klose AD, Scheel A, Moa-Anderson B, Backhaus M, Netz U, Beuthan J (2004) Sagittal Laser Optical Tomography for Imaging of Rheumatoid Finger Joints," *Physics in Medicine and Biology*; 49(7), 1147-1163.
- [19] Haykin, S. (1999) *Neural Networks*, 2nd ed., Prentice Hall, pp./ 842.