Collagen morphology and texture analysis: from statistics to classification

Leila B Mostaço-Guidolin[1,2], Alex C-T Ko[1,2,], Fei Wang[1], Bo Xing[1], Mark Hewko[1], Ganghong Tian[1,3], Arkady Major[2], Masashi Shiomi[4], Michael G Sowa[1]

[1]National Research Council Canada, Medical Device Portfolio
435 Ellice Avenue, Winnipeg, MB, Canada R3B 1Y6

[2]Department of Electrical and Computer Engineering, University of Manitoba
75A Chancellor's Circle, Winnipeg, Manitoba, Canada R3T 5V6

[3]Department of Physiology, School of Medicine, University of Manitoba
745 Bannatyne Ave., Winnipeg, Manitoba, Canada R3E 0J9

[4]Institute for Experimental Animals, Kobe University, School of Medicine
Kobe 650-0017, Japan

Supplementary information

❖ **Details about texture parameters**

Table S1 – Texture features and its mathematical expressions.

Feature	Interpretation	Mathematical Expression
FOS		
Mean	Average gray value.	$\mu = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P_{i,j} = \sum_{j=0}^{N-1} j \sum_{i=0}^{N-1} P_{i,j}$
Standard deviation	Standard deviation of the gray values used to generate the mean gray value.	$\Sigma = \sum_{i=0}^{N-1} (i - \mu)^2 \sum_{j=0}^{N-1} P_{i,j}$
Integrated density	Product of image's area and mean gray value	$I = \sum_{i=0}^{N-1} n(Ni)$
Skewness	It quantifies how symmetrical the distribution is.	$\mu_3 = \sigma^{-3} \sum_{i=0}^{N-1} (i - \mu)^3 P_i$
Kurtosis	It quantifies whether the shape of the data distribution matches the Gaussian distribution.	$\mu_4 = \sigma^{-4} \sum_{i=0}^{N-1} (i - \mu)^4 P_i - 3$
GLCM		
Energy	Degree of image's texture directions according to the perception of human eyes (also called uniformity).	$\sum_{i,j=0}^{N-1} P_{i,j}^2$

Correlation	Linear dependency of grey levels on those of neighboring pixels.	$-\sum_{i,j=0}^{N-1} P_{i,j}\left[\dfrac{(i-\mu)(j-\mu)}{\sigma^2}\right]$
Inertia	Representation of pixels entirely similar to their neighbour.	$\sum_{i,j=0}^{N-1}(i-j)^2\,P_{i,j}$
Inverse Difference Moment	Measure of the amount of local uniformity present in the image (also called homogeneity).	$\sum_{i,j=0}^{N-1}\dfrac{P^2_{i,j}}{1+(i-j)^2}$
Entropy	Measure of "lack" of organization.	$-\sum_{i,j=0}^{N-1} P_{i,j}\log P_{i,j}$

FOS: first order statistics; GLCM: gray level co-occurrence matrix; P: probability density function; i and j: gray levels.

❖ **Complementarily statistical data tests**

Table S2 – p values obtained by comparing all groups through Kruskal-Wallis test.

Texture	p-values* (group a group)
	A vs.B	A vs.C	A vs.D	A vs.E	B vs.C	B vs.D	B vs.E	C vs.D	C vs.E	D vs.E
GLCM
Energy	**0.007**	**0.006**	**0.005**	**0.002**	**0.003**	**9.15x10⁻⁴**	0.159	0.189	**0.044**	0.302
Entropy	0.076	**0.007**	**0.005**	**0.001**	0.475	0.172	**0.016**	0.536	0.196	0.189
IDM	**0.008**	**0.004**	0.636	**0.002**	**0.007**	0.277	0.481	0.444	0.506	**0.005**
Correlation	**0.008**	**0.004**	0.228	**0.002**	0.501	0.401	0.321	**0.034**	0.622	0.270
Inertia	**0.001**	**0.003**	**0.006**	**0.007**	**0.004**	0.381	0.174	0.780	0.572	0.356
FOS
Mean	0.547	0.286	0.513	0.642	0.284	0.525	0.262	**0.084**	0.505	0.163
Std.Dev.	0.525	**0.039**	0.636	**0.025**	0.443	0.368	0.448	0.259	0.186	0.311
Kurtosis	**0.049**	0.491	0.923	0.718	0.341	0.495	0.729	0.371	0.478	0.403
Int.Dens.	0.755	**0.039**	**0.007**	0.590	0.780	0.514	0.110	0.596	0.289	**0.048**
Skewness	0.684	0.823	0.752	**0.035**	0.274	0.326	0.468	0.242	0.751	0.395

significant differences for p<0.05; FOS: first order statistics; GLCM: gray level co-occurrence matrix;

❖ **Classifier performance: specific cases**

Table S3 – Summary of the results obtained considering FOS and GLCM texture parameters.

Group	Test Set	Accuracy	Sensitivity	Specificity
FOS
A	308000	92%	100%	98%

B	326000	84%	79%	63%
C	332000	83%	65%	88%
D	318000	81%	86%	90%
E	224000	83%	95%	93%
Overall	**1508000**	**84%**	**81%**	**82%**

GLCM

A	308000	100%	100%	92%
B	326000	73%	100%	100%
C	332000	79%	86%	72%
D	318000	79%	75%	92%
E	224000	91%	98%	92%
Overall	**1508000**	**87%**	**92%**	**91%**

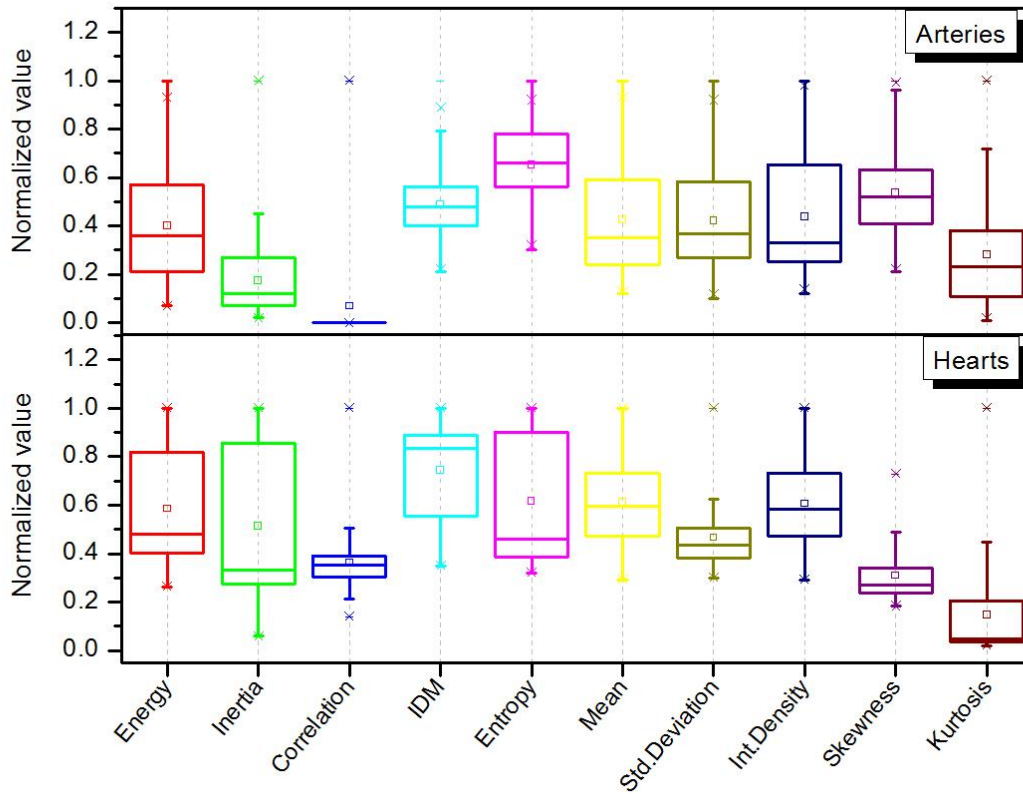❖ **Classification across data sets**

Figure S4 - Comparison of each texture parameters calculated for images acquired from atherosclerotic arteries and myocardium infarcted hearts. The top and bottom of each rectangular box denote the 75th and 25th percentiles, respectively, with the median shown inside the box. Vertical bars extending from each box represent the 90th and 10th percentiles.
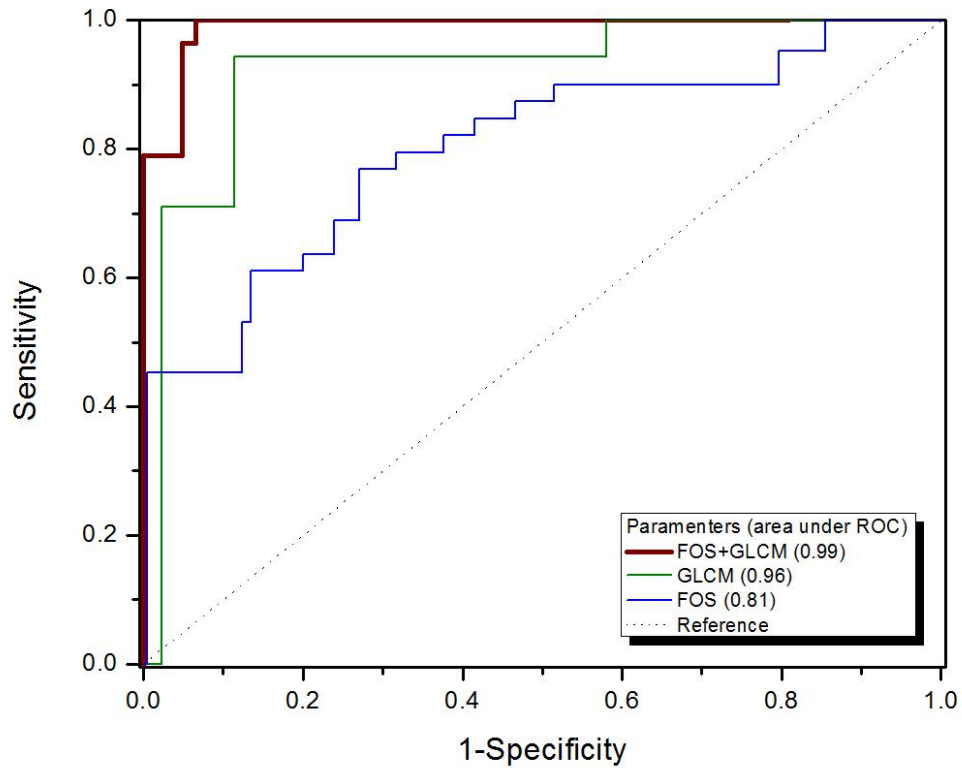
Figure S5 - ROC curves for all three texture sets tested: FOS, GLCM and FOS+GLCM. Values suggest that the classification across data sets has a good predictive value, as the area under the ROC curve was 0.99 for all texture parameters (FOS+GLCM), 0.81 for FOS parameters and 0.96 for GLCM parameters.