Supplementary material for the paper

Biomarker discovery in MALDI-TOF serum protein profiles using discrete wavelet transformation

Theodore Alexandrov [1,*], Jens Decker [2], Bart Mertens [3], Andre M. Deelder [4],
Rob A.E.M. Tollenaar [5], Peter Maass [1], Herbert Thiele [2]

[1]Center for Industrial Mathematics,
University of Bremen, D-28334 Bremen, Germany

[2]Bruker Daltonik GmbH,
D-28359 Bremen, Germany

[3]Department of Medical Statistics and Bioinformatics,
[4]Department of Parasitology,
[5]Department of Surgery,
Leiden University Medical Center,
2300 RC Leiden, The Netherlands

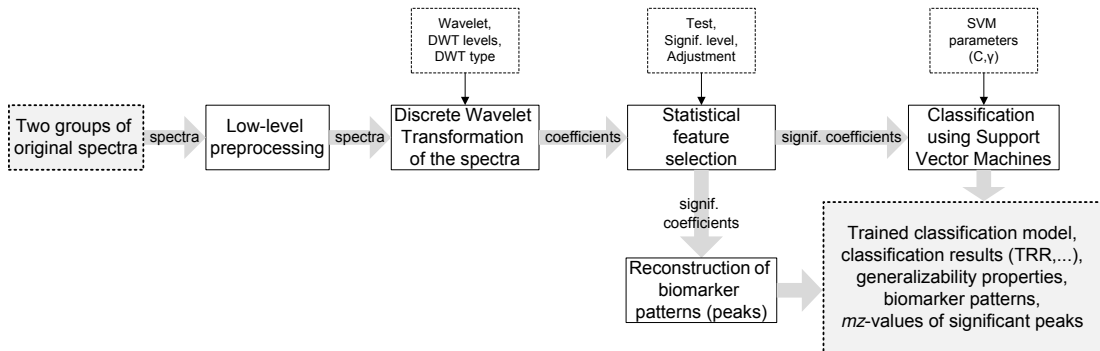[*]Corresponding author: theodore@math.uni-bremen.de

1

Figure 1: Overall scheme of the procedure proposed in the paper. Please note that this scheme does not include double cross-validation which is introduced for the simultaneous parameters estimation and predictor evaluation.
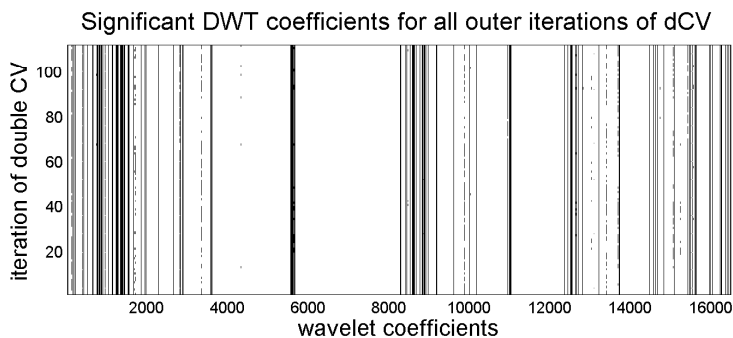

Figure 2: Indication of selection of the wavelet coefficients at each outer iteration of the double CV (selected coefficients are in black). Except of only tiny variations, the feature extraction is very stable, as the figure mostly contains vertical lines corresponding to the wavelet coefficients selected for all iterations.

Table 1: Total recognition rates and mean numbers of support vectors for classification of noised spectra using APPDWT coefficients. At each iteration of the outer loop of the double cross-validation the trained classifier is applied to the leftover spectrum with a doped normal noise. This simulates classification of new data with models trained on the original data. The noise standard deviation was 0.31 which is twice the real noise estimated in the m/z-interval $[1200, 2400]$ using the difference between the original data and a smoothed version (peak areas are excluded).

Test	TRR			Number of SV		
	BH	Bonf	BY	BH	Bonf	BY
KS	70.54	92.86	84.82	54.6	46.0	50.7
MW	66.07	96.43	86.61	56.1	43.6	50

Table 2: The most significant peaks determined by ClinProTools v.2.2 (Bruker Daltonics) which uses a t-test based on the peaks areas. All peaks have p-values less than 10^{-6}.

Peak significance order	Peak mz-value (Da)
1	1467.3
2	1264.9
3	1352.2
4	1208.0
5	1867.0
6	1898.5
7	1780.0
8	4055.6
9	1692.5
10	1520.4
11	2380.8
12	1451.3
13	3193.5
14	3264.6
15	2022.9

Table 3: The classification results, numbers of support vectors used and number of selected APPDWT coefficients for each DWT level considered individually.

DWT level	TRR		Mean number of SV		Number of coefficients	
	KS,Bonf	MW,Bonf	KS,Bonf	MW,Bonf	KS,Bonf	MW,Bonf
1	97.32	96.43	49.8	44.1	803	910
2	97.32	96.43	49.5	52.3	410	468
3	97.32	97.32	47.6	47.2	225	249
4	95.54	95.54	51.6	44.2	153	172
5	97.32	97.32	43.1	48.4	110	121
6	96.43	96.43	48.5	53.1	67	76
7	94.64	92.86	62.2	64.8	40	45
8	91.07	91.96	47.6	58.0	24	25
9	91.96	94.64	48.7	49.2	17	15
10	90.18	91.07	74.5	69.1	12	12