Random_Intensity_300_2a: 300 Cases, 300 Controls, 300 Peaks

Brian T. Luke (lukeb@ncifcrf.gov)

BMDK Analysis

31 peaks selected as putative biomarkers by the 10 methods within BMDK

Peak	catboot	student	dtgini	dtinfg	nnfeat	chisq	kruswal	kolsmir	extreme	yip
7									5	
9					3					
25		4					4	5		4
28									5	
50			5							
57					5					
79	4									
82	5									
97			3	3		2			5	
105								4		
111					4					
120									5	
148									5	
156								5		
172		1	1	1			1	1		1
178		2					3			2
187					1					
188									5	
192	1	5	2	2			5	1	3	5
197	3			5					1	
222						3				
226			4	4						
227						4				
228		3					2			3
235					2				2	
255									3	
257						1				
261						4			5	
264								3		
281	2									
286									5	

Peaks used in each of the best distance-dependent 6-nearest neighbor classifiers

Peak	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
7										
9										
25						X		X		
28										
50								X		
57										
79										
82										
97		X	X		X		X		X	
105								X		X
111			X							
120										
148										
156										
172										
178						X				
187										
188										
192										
197										
222										
226						X				
227										X
228										X
235										
255										
257	X			X						
261										
264										
281		X	X		X		X		X	
286										

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 31 putative biomarkers.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	54.0	63.2	65.6	54.2	64.9	58.5	61.5	61.6	61.4	60.5
Spec	56.7	58.2	56.9	57.4	53.8	61.9	58.9	59.5	57.9	60.3
%Undet	0.0	0.3	0.3	0.8	7.2	20.7	0.3	11.3	0.5	11.2
Quality	110.7	121.1	122.1	110.8	111.5	99.8	120.1	109.7	118.8	109.7

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 31 putative biomarkers with the caveat that %Undetermined cannot exceed 5.0%.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	54.0	63.2	65.6	54.2	56.4	None	61.5	None	61.4	None
Spec	56.7	58.2	56.9	57.4	53.4	None	58.9	None	57.9	None
%Undet	0.0	0.3	0.3	0.8	5.0	None	0.3	None	0.5	None
Quality	110.7	121.1	122.1	110.8	104.8	None	120.1	None	118.8	None

Fingerprint Analysis

Sensitivity, specificity and quality (sensitivity + specificity) for the best and 200th best decision tree constructed from any of the 300 peak intensities. The evolutionary programming search used a population size of 400 and ran for 800 generations. A decision node became a terminal node when it contained 1% (3 samples) or 4% (12 samples) of a given State.

Metric	1%		1%		4%		4%	
	1st	200th	1st	200th	1st	200th	1st	200th
Sensitivity	67.0	67.7	67.7	66.3	72.7	71.3	76.3	70.7
Specificity	70.0	68.3	68.7	69.0	64.3	64.3	61.3	65.7
Quality	137.0	136.0	136.3	135.0	137.0	135.7	137.7	136.3

Sensitivity, specificity and quality (sensitivity + specificity) for the best and 200th best medoid classifier algorithm in each of the two runs using 5-, 6-, and 7-peak intensities from the set of 300. The evolutionary programming search used a population size of 1000 and ran for 2000 generations with the requirement that there are at most 200 Case-cells and 200 Control-cells.

Metric	5-Features		5-Features		6-Features		6-Features		7-Features		7-Features	
	1st	200th	1st	200th	1st	200th	1st	200th	1st	200th	1st	200th
Sens	100.0	100.0	69.3	63.3	100.0	100.0	79.0	72.7	100.0	100.0	80.3	75.7
Spec	69.7	63.3	100.0	100.0	75.3	71.3	100.0	100.0	78.7	74.0	100.0	100.0
Quality	169.7	163.3	169.3	163.3	175.3	171.3	179.0	172.7	178.7	174.0	180.3	175.7

(Last updated 4/21/07)