

## Random\_Intensity\_60\_2a: 60 Cases, 60 Controls, 300 Peaks

Brian T. Luke ([lukeb@ncifcrf.gov](mailto:lukeb@ncifcrf.gov))

### BMDK Analysis

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

Peak	catboot	student	dtgini	dtinfg	nnfeat	chisq	kruswal	kolsmir	extreme	vartest
9	3	2					3	5		2
30	1									
32					2	1		5	4	
43			1	1					1	
48									4	
50						3				
67				3						
71			2					1		
75		4			1		5	5		4
76		5						5		5
79		3					2			3
82						5				
96						4				
109								1		
111					4					
124	5									
135			5					1		
136								5		
173			4	2					2	
176		1	2				1	5		1
178									4	
179	2									
189								5	4	
221									4	
230	4					1				
236				4						
239					5			1		
274				5					3	
300					3		4			



Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 29 putative biomarkers.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	55.0	64.9	67.2	55.9	78.0	70.6	68.4	74.0	69.0	72.5
Spec	66.7	69.0	63.3	66.1	61.1	87.2	69.8	75.6	69.8	76.1
%Undet	0.0	4.2	1.7	1.7	13.3	39.2	8.3	20.8	7.5	19.2
Quality	121.7	129.7	128.9	120.4	125.8	118.6	129.9	128.7	131.3	129.5

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 29 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	55.0	64.9	67.2	55.9	None	None	61.4	None	61.4	None
Spec	66.7	69.0	63.3	66.1	None	None	65.5	None	66.7	None
%Undet	0.0	4.2	1.7	1.7	None	None	4.2	None	5.0	None
Quality	121.7	129.7	128.9	120.4	None	None	122.8	None	123.1	None

## Fingerprint Analysis

Sensitivity, specificity and quality (sensitivity + specificity) for the best and 200<sup>th</sup> best decision tree constructed from any of the 300 peak intensities. The evolutionary programming search used a population size of 200 and ran for 400 generations. A decision node became a terminal node when it contained 1% (no samples) or 4% (2 samples) of a given State.

Metric	1%		1%		4%		4%	
	1 <sup>st</sup>	200 <sup>th</sup>						
Sensitivity	93.3	88.3	85.0	83.3	88.3	85.0	93.3	88.3
Specificity	83.3	83.3	90.0	88.3	83.3	83.3	80.0	80.0
Quality	176.7	171.7	175.0	171.7	171.7	168.3	173.3	168.3

Sensitivity, specificity and quality (sensitivity + specificity) for the best and 200<sup>th</sup> 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 400 and ran for 800 generations with the requirement that there are at most 40 Case-cells and 40 Control-cells.

Metric	5-Features		5-Features		6-Features		6-Features		7-Features		7-Features	
	1 <sup>st</sup>	200 <sup>th</sup>										
Sens	100.0	100.0	93.3	83.3	100.0	100.0	93.3	85.0	100.0	100.0	93.3	86.7
Spec	88.3	81.7	100.0	100.0	93.3	85.0	100.0	100.0	95.0	86.7	100.0	100.0
Quality	188.3	181.7	193.3	183.3	193.3	185.0	193.3	185.0	195.0	186.7	193.3	186.7

Note that scrambling Cases and Controls evens out the difference between sensitivity and specificity. For example, one scrambled dataset yielded the following 5-Feature results:

Sensitivity = 96.7% (2 Cases misclassified) with 37 Case cells

Specificity = 91.7% (5 Controls misclassified) with 34 Control cells

This allows the following combinations of sensitivity and specificity between training sets and testing sets when 20 Cases and Controls are removed to form the testing set.

Sensitivity		Specificity	
Training	Testing	Training	Testing
100.0	90.0	100.0	75.0
97.5	95.0	97.5	80.0
95.0	100.0	95.0	85.0
		92.5	90.0
		90.0	95.0
		87.5	100.0

(Last updated 4/27/07)