

## Random\_Intensity\_150\_5a: 150 Cases, 150 Controls, 300 Peaks

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

### BMDK Analysis

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

Peak	catboot	student	dtgini	dtinfg	nnfeat	chisq	kruswal	kolsmir	extreme	vartest
5									2	
11			3	4						
15		3					3			3
33						2				
37					4					
38	1	1	1	1	1		1	1	5	1
48		5	4				4			5
136							5			
144					5					
149						2				
167	3							3		
174	4									
176				2					1	
191		4								4
211	2	2	2	3	2	2	2	2		2
214						1				
219			5		3					
229								3		
236								3		
238			5						2	
239								3		
248	5					5				
264				5					2	



Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 23 putative biomarkers.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	56.7	65.0	62.3	60.3	63.6	70.5	63.9	62.5	63.4	62.7
Spec	59.3	61.9	63.4	56.8	64.6	66.7	61.7	70.2	62.4	70.4
%Undet	0.0	3.3	4.0	2.0	10.0	26.0	2.3	13.3	2.0	13.7
Quality	116.0	123.6	121.7	115.0	118.2	111.1	123.3	119.3	123.9	119.4

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 23 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	56.7	65.0	62.3	60.3	None	None	63.9	None	63.4	None
Spec	59.3	61.9	63.4	56.8	None	None	61.7	None	62.4	None
%Undet	0.0	3.3	4.0	2.0	None	None	2.3	None	2.0	None
Quality	116.0	123.6	121.7	115.0	None	None	123.3	None	123.9	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 400 and ran for 800 generations. A decision node became a terminal node when it contained 1% (1 sample) or 4% (6 samples) of a given State.

Metric	1%		1%		4%		4%	
	1 <sup>st</sup>	200 <sup>th</sup>						
Sensitivity	71.3	70.0	76.7	69.3	61.3	66.0	77.3	77.3
Specificity	81.3	81.3	77.3	83.3	88.0	80.7	71.3	69.3
Quality	152.7	151.3	154.0	152.7	149.3	146.7	148.7	146.7

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 1000 and ran for 2000 generations with the requirement that there are at most 100 Case-cells and 100 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	80.0	74.0	100.0	100.0	82.7	78.0	100.0	100.0	84.0	80.0
Spec	82.0	75.3	100.0	100.0	82.7	78.7	100.0	100.0	85.3	80.7	100.0	100.0
Quality	182.0	175.3	180.0	174.0	182.7	178.7	182.7	178.0	185.3	180.7	184.0	180.0

(Last updated 4/21/07)