

**Random\_Intensity\_150\_1a: 150 Cases, 150 Controls, 300 Peaks**  
 Brian T. Luke ([lukeb@ncifcrf.gov](mailto:lukeb@ncifcrf.gov))

**BMDK Analysis**

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

Peak	catboot	student	dtgini	dtinfg	nnfeat	chisq	kruswal	kolsmir	extreme	vartest
20	3									
31	4	1	3	4			1	2		1
48		5					5			5
49									4	
60				5		2			1	
63		4	2	2			4	1		4
77					4	4				
92									4	
94					3					
107					5					
121									2	
130								5		
138	5									
153						4				
221			5	3						
233									2	
240			1	1		3				
244		3					3		4	3
257									4	
282						1				
288	1		4		2			3		
294	2	2			1		2	4		2

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
20										
31										
48										
49							X			
60										
63								X		X
77								X		X
92										
94										
107										
121	X			X						
130										
138										
153										
221										
233		X								
240			X		X	X				
244		X			X					
257										
282			X			X			X	
288			X			X	X		X	
294								X		X

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 22 putative biomarkers.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	58.7	66.7	65.9	58.7	65.7	62.2	63.2	67.4	62.6	67.4
Spec	55.3	60.7	66.9	55.3	64.1	71.8	62.8	65.1	61.9	65.4
%Undet	0.0	0.0	11.7	0.0	10.7	23.7	2.7	14.0	2.0	13.7
Quality	114.0	127.3	121.2	114.0	119.1	110.3	123.4	118.5	122.5	119.1

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 22 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	58.7	66.7	58.9	58.7	None	None	63.2	None	62.6	None
Spec	55.3	60.7	52.8	55.3	None	None	62.8	None	61.9	None
%Undet	0.0	0.0	5.0	0.0	None	None	2.7	None	2.0	None
Quality	114.0	127.3	106.6	114.0	None	None	123.4	None	122.5	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	73.3	72.7	70.7	69.3	81.3	72.7	75.3	74.0
Specificity	78.7	77.2	81.3	80.7	68.7	75.3	77.3	76.0
Quality	152.0	150.0	152.0	150.0	150.0	148.0	152.7	150.0

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	82.0	78.0	100.0	100.0	84.7	79.3
Spec	82.0	76.0	100.0	100.0	83.3	79.3	100.0	100.0	86.7	81.3	100.0	100.0
Quality	182.0	176.0	180.0	174.0	183.3	179.3	182.0	178.0	186.7	181.3	184.7	179.3

(Last updated 4/21/07)