

## Random\_Intensity\_150\_3a: 150 Cases, 150 Controls, 300 Peaks

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

### BMDK Analysis

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

Peak	catboot	student	dtgini	dtinfg	nnfeat	chisq	kruswal	kolsmir	extreme	vartest
5	2				4					
10	3									
24						2				
25							2			3
43	5									
86									3	
101		3	2	3	1		4	1		
103									3	
110						3			2	
111		1	1	1	2		3	1		2
161										5
176					5					
186	1						1	3		
190		4		4					1	1
199						5				
206					3		5			4
246		5		5					3	
257						1				
259						3				
268									3	
273	4									
285								5		
287			5					4		
293			4					5		
297		2	3	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
5										
10			X							X
24					X					
25										
43										
86		X								
101					X		X	X		X
103										X
110										
111			X			X				
161										
176						X			X	
186										
190										
199						X				
206										
246										
257							X			
259								X		
268									X	
273		X								
285										
287								X		
293										
297	X		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 25 putative biomarkers.

Metric	1-ad	2-ad	3-ad	1-rd	2-rd	3-rd	2-cr	3-cr	2-sd	3-sd
Sens	53.3	59.9	61.1	55.0	65.9	68.8	63.5	69.8	59.2	65.9
Spec	60.0	62.4	62.7	60.1	60.7	63.3	59.2	64.5	63.4	68.8
%Undet	0.0	1.3	0.3	1.0	11.0	26.3	1.7	12.0	2.7	15.3
Quality	115.3	120.9	123.4	114.2	115.6	105.7	121.0	122.3	120.0	119.3

Sensitivity, specificity, %undetermined, and quality (sensitivity + specificity - %undetermined) for each of the best distance-dependent 6-nearest neighbor classifiers using any of the 25 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	53.3	59.9	61.1	55.0	None	None	63.5	None	59.2	None
Spec	60.0	62.4	62.7	60.1	None	None	59.2	None	63.4	None
%Undet	0.0	1.3	0.3	1.0	None	None	1.7	None	2.7	None
Quality	115.3	120.9	123.4	114.2	None	None	121.0	None	120.0	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	69.3	73.3	77.3	75.3	77.3	76.0	85.3	84.7
Specificity	82.0	75.3	73.3	73.3	72.0	71.3	65.3	64.7
Quality	151.3	148.7	150.7	148.7	149.3	147.3	150.7	149.3

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 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	75.3	100.0	100.0	84.0	79.3	100.0	100.0	85.3	80.7
Spec	80.0	74.7	100.0	100.0	83.3	78.7	100.0	100.0	84.0	80.7	100.0	100.0
Quality	180.0	174.7	180.0	175.3	183.3	178.7	184.0	179.3	184.0	180.7	185.3	180.7

(Last updated 4/21/07)