

BusinessConnect

A New Era of Smart

June 12 2014

IBM SPSS Modeler – סיוע בזיהוי מחלת סרטן השד

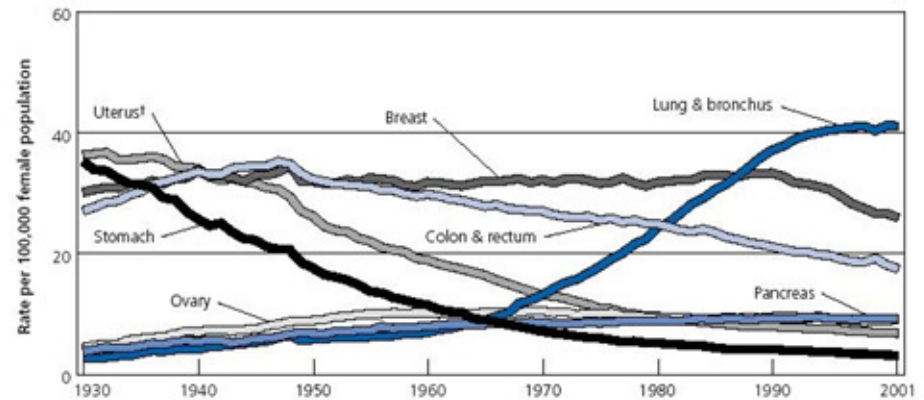
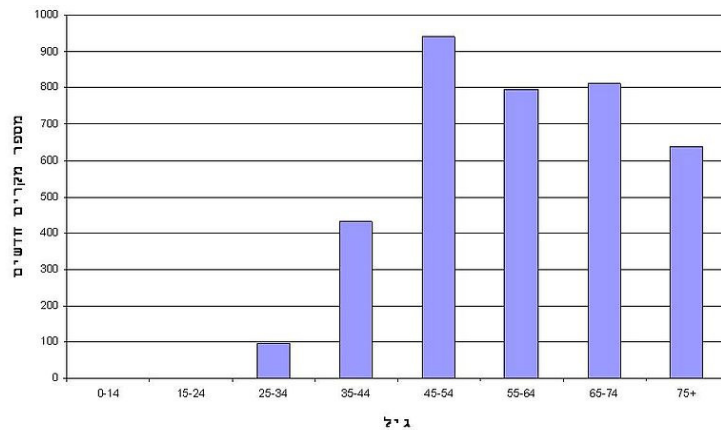
ד"ר ליאור שטראוס
מנהל מחלקת מו"פ - אוונטוס דיאגנוסטיקה



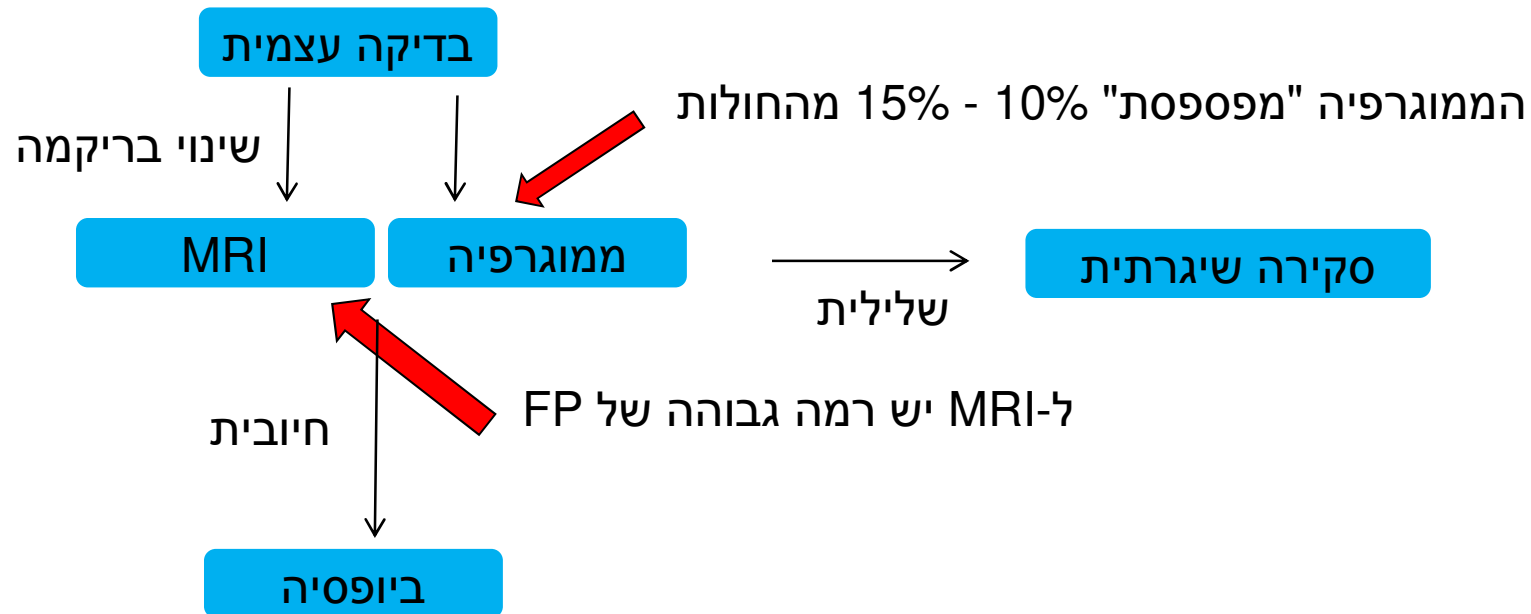
רקע כללי – על סרטן השד

- סרטן השד הוא הסרטן הנפוץ ביותר במדינת ישראל
- בממוצע בשנה מאובחנות כ-4000 נשים חולות בסרטן זה
- אחוז התמותה תלוי בשלב הגילוי
- בממוצע קצב התמותה נע באזור ה-1000 (שנת 2009 – 949 פטירות)
- סרטן השד נפוץ בעיקר באזור גילאי ה-50

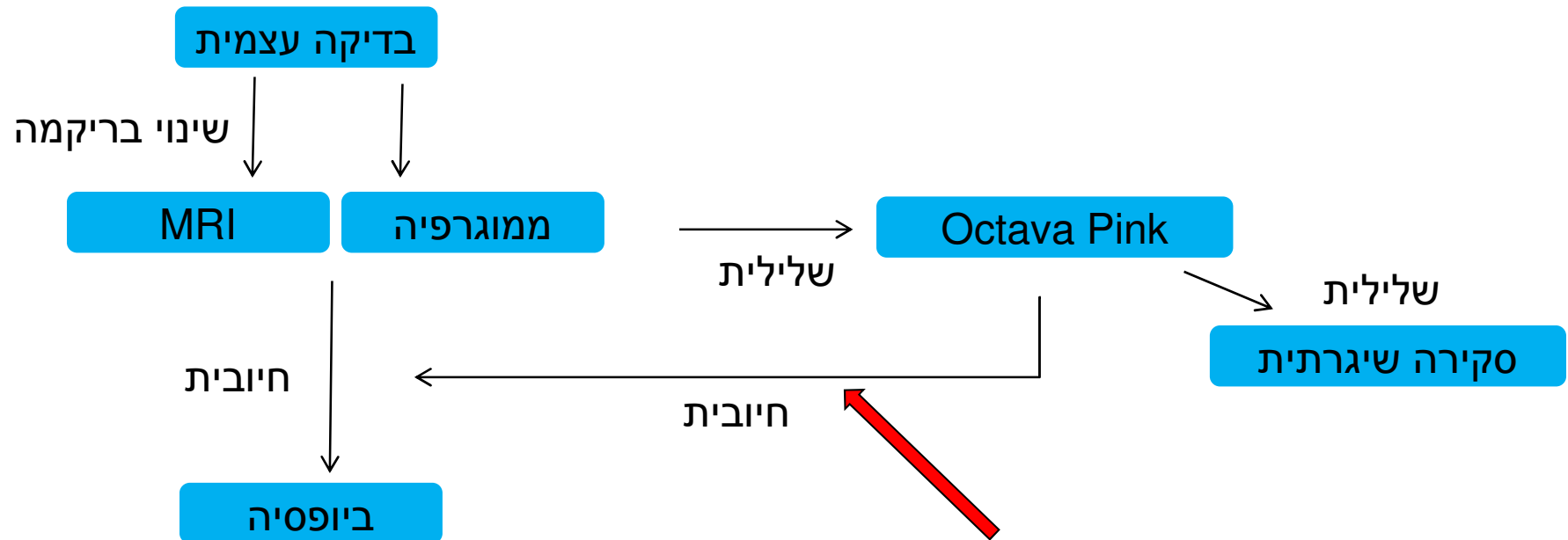
מספר נשים שאובחנו בסרטן השד (גידול ראשוני) בישראל בשנת 2004 (במספרים מוחלטים)



תהליך הדיאגנוסטיקה – סרטן השד



תהליך הדיאגנוסטיקה – סרטן השד

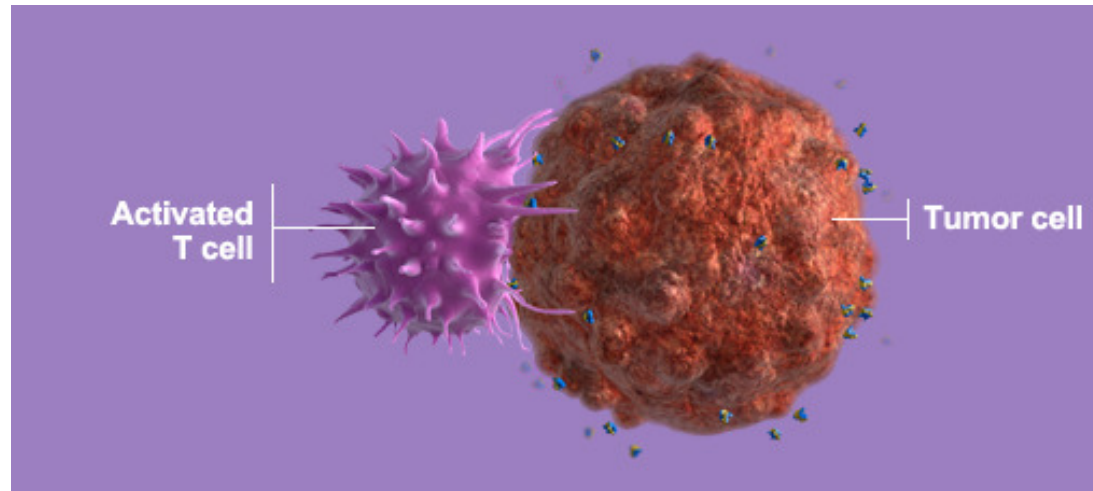
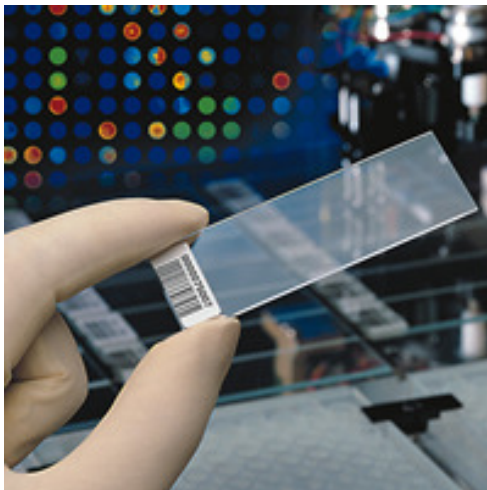


Octava pink מזהה 70%-50% מהחולות שהממוגרפיה מספספת



המנגנון הביולוגי

- המערכת החיסונית נלחמת במחלה.
- המלחמה מתבטאת בשינוי ברמות הנודנים בזרם הדם.
- גידולים גורמים למוטציות בחלבון ספציפים ברקמה.
- החלבונים שרמות הנוגדנים כנגדם נמצאו "מפרידות" בין חולות ובריאות מודפסות על סלייד.
- פרופיל הנוגדנים של מאות נשים נלמד על מנת לאפיין את הפרופיל החולה והבריא



התהליך האנליטי

הכנת המידע

חלוקת המידע לקבוצות
למידה וביקורת

בניית מודלים לחיזוי

ניבוי נבדקות unknown

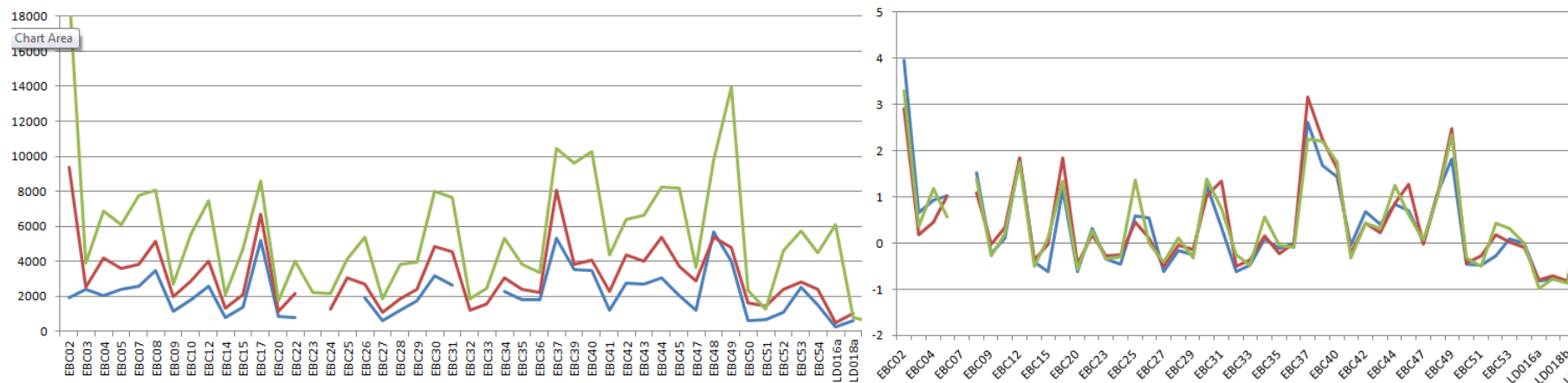
בחינת איכות המודלים



הכנת המידע

תהליכי הכנת המידע – נרמול הנתונים

- במערכת החיסונית יש רמות בסיס שונות של נוגדנים.
- על מנת לאפשר השוואה בין נבדקות יש צורך לנרמל את הערכים.
- המידע עובר מספר נרמולים מתמטיים.

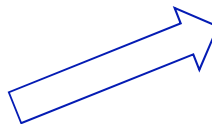


חלוקת המידע לקבוצות למידה וביקורת

כריית מידע – בניית מודלים

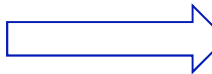
	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105	0	49	1046.87	896.5	1865.1	1218.19	1772
3	B0106	0	66	438.1167	347.9917	563.55	591.6167	1081.133
4	B0110	0	46	1284.61	1267.18	2402.34	1996.4	2256
5	B0113	0	58	1057.2	1067.11	2113.92	1450.93	2129.6
6								2131.73
7								2966.38
8								2114.07
9								4976.52
10								2927.3
11		0	58	1057.2	1067.11	2113.92	1450.93	1420.9
12	B0134	0	62	2254.6	1884.642	5491.483	2663.933	4478.2
13	B0138	0	46	6542.83	7950.4	14183.2	15807.5	4284.78
14	B0149	0	51	848.211	1222.88	1281.3	715.25	1264.55
15	B0154	0	58	1973.93	2304.27	4783.55	2913.72	3953.77
16	B0161	0	62	1198.83	2774.6	2541.05	2136.5	2272.52

~1000 subjects Patient/Control



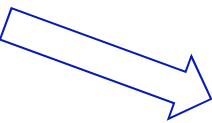
	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105	0	49	1046.87	896.5	1865.1	1218.19	1772
3	B0106	0	66	438.1167	347.9917	563.55	591.6167	1081.133
4	B0110	0	46	1284.61	1267.18	2402.34	1996.4	2256
5	B0113	0	58	1057.2	1067.11	2113.92	1450.93	2129.6
6	B0118							2131.73
7	B0127							2966.38
8	B0128							2114.07
9	B0130							4976.52
10	B0131							2927.3
11	B0132	0	58	1057.2	1067.11	2113.92	1450.93	1420.9
12	B0134	0	62	2254.6	1884.642	5491.483	2663.933	4478.2
13	B0138	0	46	6542.83	7950.4	14183.2	15807.5	4284.78
14	B0149	0	51	848.211	1222.88	1281.3	715.25	1264.55
15	B0154	0	58	1973.93	2304.27	4783.55	2913.72	3953.77
16	B0161	0	62	1198.83	2774.6	2541.05	2136.5	2272.52

55% - train



	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105	0	49	1046.87	896.5	1865.1	1218.19	1772
3	B0106							1081.133
4	B0110							2256
5	B0113							2129.6
6	B0118							2131.73
7	B0127	0	52	2156.67	2561.3	2481.88	2702.7	2966.38
8	B0128	0	63	616.383	969.5	1536.3	983.217	2114.07
9	B0130	0	51	2000.57	1980.41	3029.41	3002.8	4976.52

25% - test



	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105	0	49	1046.87	896.5	1865.1	1218.19	1772
3	B0106							133
4	B0110							256
5	B0113							19.6
6	B0118	0	63	964.483	1079.5	3051.81	1863.68	2131.73
7	B0127	0	52	2156.67	2561.3	2481.88	2702.7	2966.38

20% - blind test



בניית מודלים לחיזוי

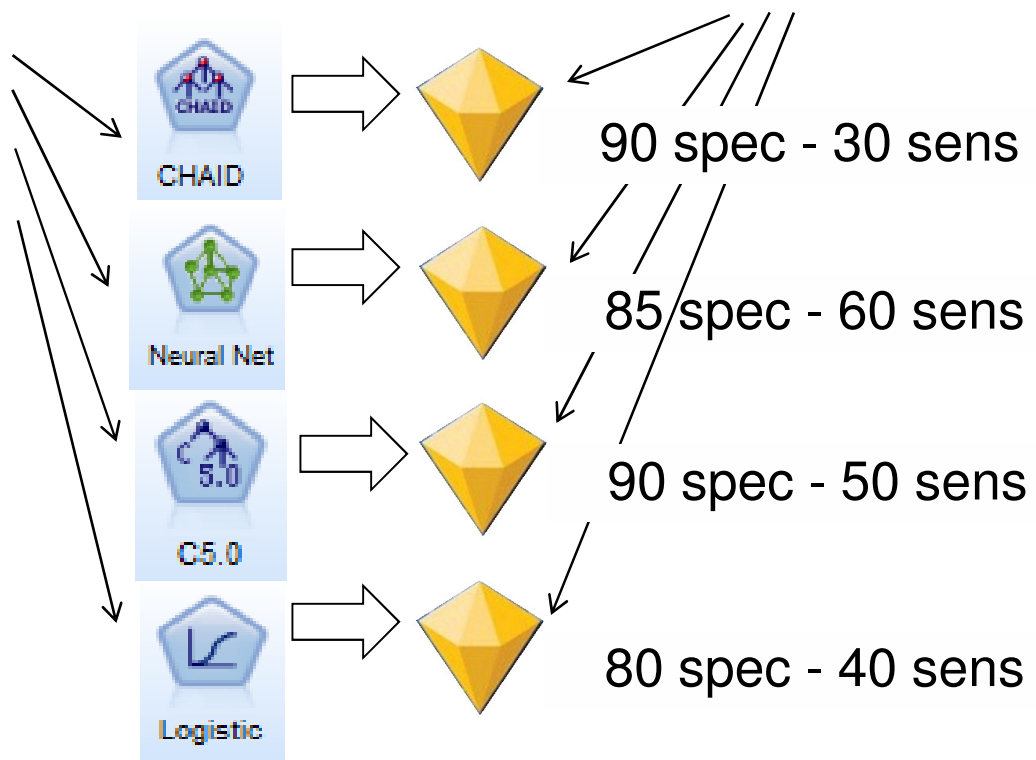
כריית מידע – בניית מודלים

1	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105	0	49	1046.87	896.5	1865.1	1218.19	1772
3	B0106	0	66	438.1167	347.9917	563.55	591.6167	1081.133
4	B0110	0	46	1284.61	1267.18	2402.34	1996.4	2256
5	B0113	0	58	1057.2	1067.11	2113.92	1450.93	2129.6
6	B0118							131.73
7	B0127							966.38
8	B0128							114.07
9	B0130							976.52
10	B0131							2927.3
11	B0132	0	58	705.1	867.083	1656.93	1003.77	1420.9
12	B0134	0	62	2254.6	1884.642	5491.483	2663.933	4478.2
13	B0138	0	46	6542.83	7950.4	14183.2	15807.5	4284.78
14	B0149	0	51	848.211	1222.88	1281.3	715.25	1264.55
15	B0154	0	58	1973.93	2304.27	4783.55	2913.72	3953.77
16	B0161	0	62	1198.83	2774.6	2541.05	2136.5	2272.52

55% - train

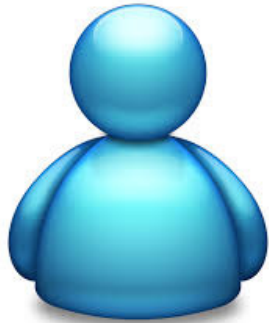
1	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B0105							1772
3	B0106							1081.133
4	B0110							2256
5	B0113							2129.6
6	B0118							2131.73
7	B0127	0	52	2156.67	2561.3	2481.88	2702.7	2966.38
8	B0128	0	63	616.383	969.5	1536.3	983.217	2114.07
9	B0130	0	51	2000.57	1980.41	3029.41	3002.8	4976.52

25% - test



ניבוי נבדקות unknown

פרדיקציה – ניבוי נבדקות unknown

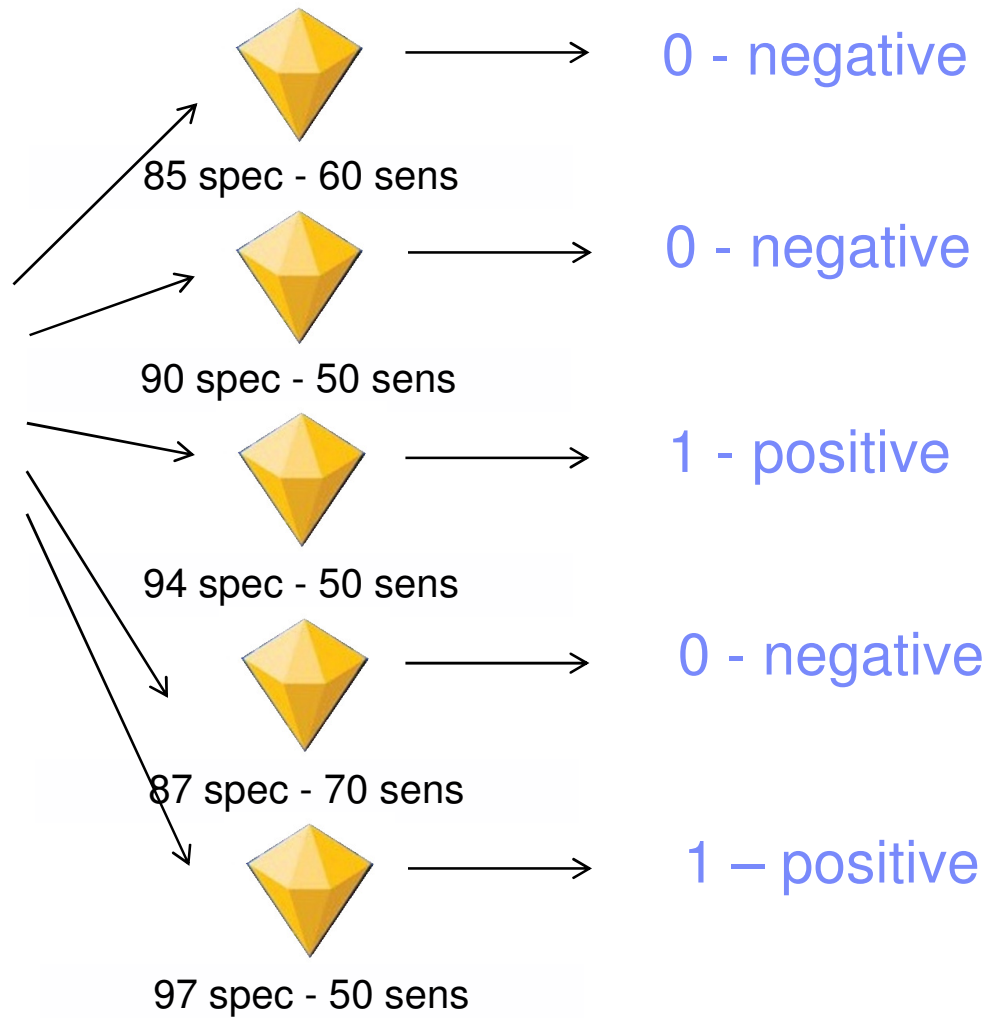


Unknown

סה"כ הופיעה כחולה: 2 פעמים

Threshold = 3

נבדקת -- תוצאה: ה בריאה

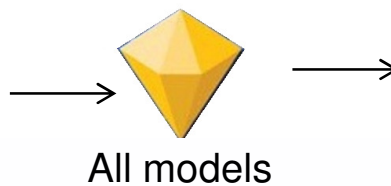


בחינת איכות המערכת

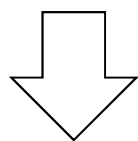
פרדיקציה – קביעת איכות ה-Training set

	A	B	C	D	E	F	G	H
1	id	type	age	dld001	dld002	dld007	dld008	dld011
2	B1							.72
3	B1							.33
4	B1							.56
5	B1							.7
6	B0118	0	63	964.483	1079.5	3051.81	1863.6	
7	B0127	0	52	2156.67	2561.3	2481.88	2702.7	

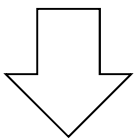
20% - blind test



פרדיקציה לכלל הנבדקות כ- unknown



בחינת הפרדיקציות אל מול נתוני האמת



קביעות יכולות ההפרדה של המערכת

96 – specificity
65 - sensitivity

תודה על ההקשבה

