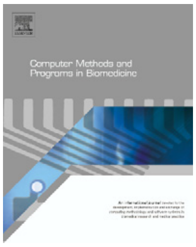
journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Comparative analysis of cell parameter groups for breast cancer detection

David Blokh[a], *Ilia Stambler*[a], *Elena Afrimzon*[a], *Max Platkov*[a], *Yana Shafran*[a], *Eden Korech*[a], *Judith Sandbank*[b], *Naomi Zurgil*[a], *Mordechai Deutsch*[a,*]

[a] *The Biophysical Interdisciplinary Jerome Schottenstein Center for the Research and the Technology of the Cellome, Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel*
[b] *Department of Pathology and Cytology, Assaf Harofeh Medical Center, Zerifin 70300, Israel*

ARTICLE INFO

Article history:
Received 12 August 2008
Received in revised form
5 January 2009
Accepted 7 January 2009

Keywords:
Selecting parameter groups
Information Theory
Nonparametric Statistics
Breast cancer (BC)
Fluorescence polarization (FP)
Peripheral blood mononuclear cells (PBMC)
Mitogenic stimulation

ABSTRACT

We present a method for the comparative analysis of parameter groups according to their correlation to disease. The theoretical basis of the proposed method is Information Theory and Nonparametric Statistics. Normalized mutual information is used as the measure of correlation between parameters, and statistical conclusions are based on ranking. The fluorescence polarization (FP) parameter is considered as the principal diagnostic characteristic. The FP was measured in fluorescein diacetate (FDA)-stained individual peripheral blood mononuclear cells (PBMC), derived from healthy subjects and breast cancer (BC) patients, under different stimulation conditions: by tumor tissue, the mitogen phytohemagglutinin (PHA) or without the stimulants. The FP parameters were grouped according to their correlation with breast cancer. It was established that the greatest difference between cells of BC patients and healthy subjects is found in the PHA test (parameter P1).

1. Introduction

The creation of formal models of cellular diagnostics is an important line of breast cancer research [1,2]. The construction of a formal diagnostic model necessitates the description of parameters relevant to the disorder, and selecting, out of the entire set of parameters, such parameters that are most relevant to the disorder, i.e. their comparative analysis [3,4]. The present work constructs a formal cell parameter group selection system, the mathematical basis of which is Information Theory [5,6] and Nonparametric Statistics [7].

Information Theory has been earlier employed for oncological investigations [8–10]. The informatic approach for the selection of parameter groups has been applied [11,12]. Alongside the informatic approach, there have been used “decision trees” [11] and “greedy selection” [12]. An optimization approach [13] and statistical approach [14] for selecting parameter groups have been considered. The application of Nonparametric Statistics in biology and medicine has been described earlier [15].

The initial choice of parameters is very important, since it determines the direction of subsequent research. The present study uses variants of cellular fluorescence polarization (FP) parameter, indicative of cell activation status, as the principal diagnostic characteristics. The possibility of breast cancer (BC) detection by lymphocyte activation status was first sug-

* *Corresponding author*. Tel.: +972 3 534 4675; fax: +972 3 534 2019.
E-mail address: rtcellom@mail.biu.ac.il (M. Deutsch).

0169-2607/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2009.01.005

gested by Wolberg [16]. FP is considered to be one of the first measures of cellular functionality [17]. In the course of cell activation, the processes linking early and late intracellular events involve conformational changes of cytosolic enzymes and/or their regulatory proteins, as well as intracellular matrix re-organization [18,19]. These early structural changes have been monitored by FP of cellular fluorescent probes [20,21].

The present study analyses the FP measured in fluorescein diacetate (FDA) stained individual peripheral blood mononuclear cells (PBMC), derived from healthy subjects and BC patients, under different stimulation conditions: incubated with or without the mitogen phytohemagglutinin (PHA) or tumor tissue. In each type of experiment, the FP was measured at 24 successive time points, in 2 emission wavelength regions: 530 nm and 580 nm.

Having established the parameters relevant to the diagnostic task, the next necessary step, i.e. the comparative analysis of parameters, is a crucial component in mathematical diagnostics (pattern recognition). This analysis is based on selecting, out of the entire set of parameters relevant to the diagnostic task, such subsets that are best correlated to the patterns studied [3,12,22].

The subject problem of the present study can be formulated as follows: given several groups of the FP parameters (FP measured in the subject cells at different/defined time points, under different/defined experimental conditions), we need to select the groups of most informative (discriminative) parameters. Or more precisely: we need to perform the ranking of the groups of parameters according to their estimated correlation with the patterns studied. In the present multiparametric study, the parameter groups are 12 time-series' of measurements, each series under a different stimulation condition and/or measured at a different FP wavelength. Thus, there are 12 groups, and each group (time-series) includes 24 parameters (measurements). The subject problem is to estimate how these measurement series' differ from each other (see Section 2).

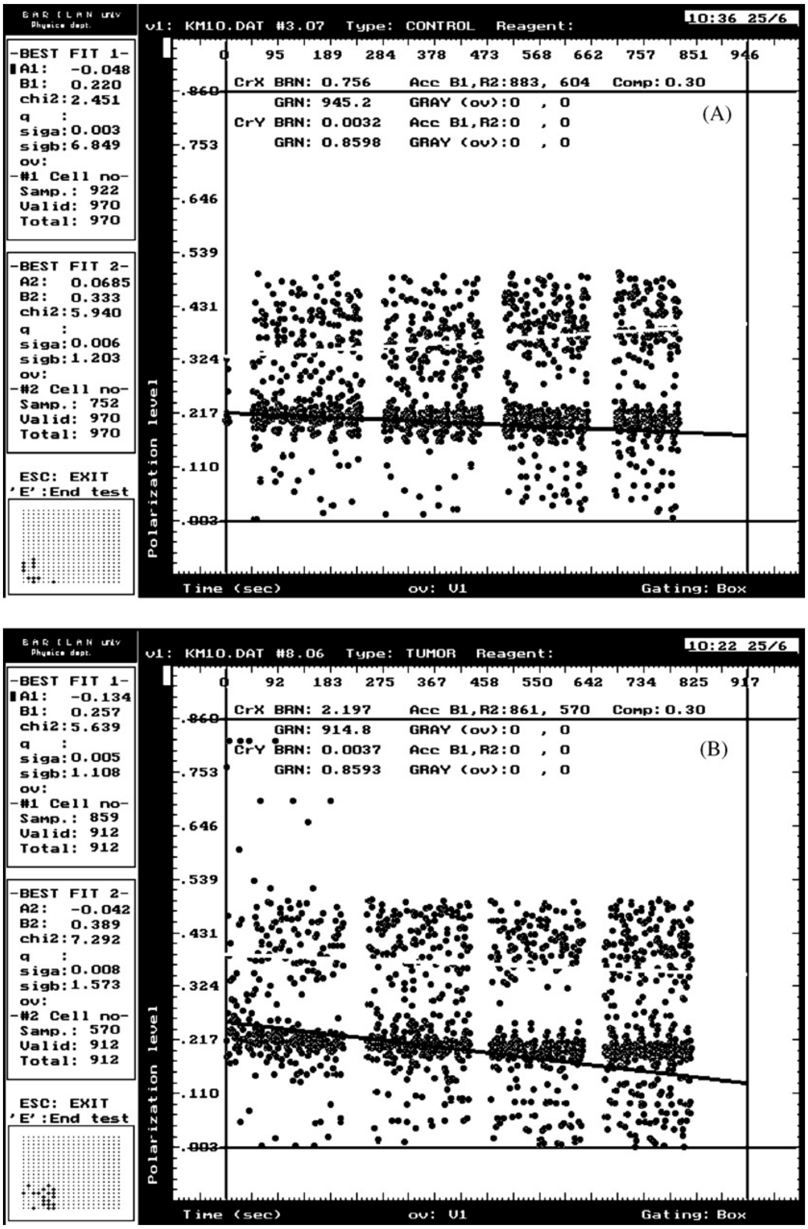
2. Materials and methods

2.1. Cell separation and stimulation

Twenty-three healthy volunteers and 27 age-matched breast cancer patients participated in the present study. The research was conducted in accordance with the principles of the Declaration of Helsinki, in collaboration with the surgical and pathological departments of Assaf Harofeh Medical Center, Zerifin, Israel, and Rabin Medical Center, Petach Tikva, Israel, which provided the blood and tumor tissue samples and carried out the clinical and pathological tumor tissue examinations. The age of subjects ranged 30–87. Among BC patients, 8 had first histopathological-grade tumor, 13 second-grade, and 6 third-grade tumor. 20 had ductal carcinoma, 7 intralesional intraductal cancer, and 10 ductal carcinoma *in situ* (Table 1).

Table 1 – Clinical and pathological characteristics of patients: *IDC*—invasive ductal carcinoma, *IIC*—intralesional intraductal cancer, *DCIS*—ductal carcinoma *in situ*, *ER*—estrogen receptor, *PgR*—progesterone receptor.

Patient number	Age	Tumor pathology	Tumor grade	Nodal status	Hormonal receptor	
					ER	PgR
1	77	IDC	2	No	++++	++
2	61	IDC	2	No	Negative	Negative
3	35	DCIS	1	No	Unknown	Unknown
4	66	DCIS	1	No	Unknown	Unknown
5	72	IDC	2–3	No	++++	++++
6	36	IDC	2	No	Negative	Negative
7	87	IDC, IIC	3	No	++++	++++
8	77	IDC, DCIS	2	No	Negative	Negative
9	71	IDC	3	Yes	++++	++++
10	83	IDC, IIC	2	No	++++	+
11	84	IDC, IIC	1	No	++++	+
12	71	IDC, DCIS	2	Yes	++++	Negative
13	61	IDC	3	Yes	++++	++++
14	74	IDC	2	Yes	++++	Negative
15	50	IDC	2	No	++++	++
16	67	DCIS	1	No	Unknown	Unknown
17	67	IDC	2	No	++++	Negative
18	30	IDC	3	No	Negative	Negative
19	69	DCIS	1	No	Unknown	Unknown
20	51	IDC, IIC	2	Yes	++++	+
21	76	IDC, IIC	2	No	++++	+
22	70	IDC, DCIS	2	No	Negative	Negative
23	50	DCIS	1	No	Unknown	Unknown
24	77	DCIS	1	No	Unknown	Unknown
25	37	IDC, IIC	3	No	Negative	Negative
26	56	IDC, IIC	2	No	++++	+
27	72	DCIS	1	No	Unknown	Unknown



```
BAR ILAN univ
Physics dept.
v1: KM10.DAT #3.07   Type: CONTROL   Reagent:          10:36 25/6

-BEST FIT 1-
 A1:   -0.048
 B1:    0.220
 chi2:2.451
 q    :
 siga:0.003
 sigb:6.849
 ov:
-#1 Cell no-
 Samp.:  922
 Valid:  970
 Total:  970

-BEST FIT 2-
 A2:    0.0685
 B2:    0.333
 chi2:5.940
 q    :
 siga:0.006
 sigb:1.203
 ov:
-#2 Cell no-
 Samp.:  752
 Valid:  970
 Total:  970

 ESC: EXIT
'E':End test

CrX BRN: 0.756     Acc B1,R2:883, 604   Comp:0.30
    GRN: 945.2     GRAY (ov):0   ,  0
CrY BRN: 0.0032    Acc B1,R2:0   ,  0
    GRN: 0.8598    GRAY (ov):0   ,  0

Time (sec)          ov: V1          Gating: Box
```

```
BAR ILAN univ
Physics dept.
v1: KM10.DAT #8.06   Type: TUMOR   Reagent:          10:22 25/6

-BEST FIT 1-
 A1:   -0.134
 B1:    0.257
 chi2:5.639
 q    :
 siga:0.005
 sigb:1.108
 ov:
-#1 Cell no-
 Samp.:  859
 Valid:  912
 Total:  912

-BEST FIT 2-
 A2:   -0.042
 B2:    0.389
 chi2:7.292
 q    :
 siga:0.008
 sigb:1.573
 ov:
-#2 Cell no-
 Samp.:  570
 Valid:  912
 Total:  912

 ESC: EXIT
'E':End test

CrX BRN: 2.197     Acc B1,R2:861, 570   Comp:0.30
    GRN: 914.8     GRAY (ov):0   ,  0
CrY BRN: 0.0037    Acc B1,R2:0   ,  0
    GRN: 0.8593    GRAY (ov):0   ,  0

Time (sec)          ov: V1          Gating: Box
```

Fig. 1 – The scatter diagram of FP (ordinate) vs. time (abscissa) in individual PBMC in a representative BC patient. (A) Cells incubated with PBS, (B) cells incubated with tumor tissue. Each cluster is defined by 6 FP measurements taken at 6 time points for the same individual cells when exposed to sequentially increasing FDA concentrations (0.6 μM, 1.2 μM, 2.4 μM, 3.6 μM). Thus 24 measurements are taken for each cell. The spaces between clusters represent the duration of the staining solutions replacement. Above: FP measured at 580 ± 10 nm emission wavelength (red spectrum area). Below: FP measured at 530 ± 10 nm emission wavelength (green spectrum area).

Peripheral blood was collected in heparinized tubes shortly before anesthesia and surgery. The mononuclear cells were separated by Ficoll-Hypaque (Sigma, St. Louis, Mo., USA) density gradient centrifugation. The cells were subsequently harvested, washed twice in phosphate-buffered saline (PBS; pH 7.2) and resuspended in complete RPMI-1640 medium. The obtained peripheral blood mononuclear cells (PBMC) of each examined person were incubated with or without the stimulants for 30 min at 37 °C: the first cell aliquot was incubated with PBS (the "Control test group"), the second – with the standard nonspecific mitogen phytohemagglutinin (PHA, 90 μg/mL) in PBS (the "PHA test group"), and the third – with a small piece (<1 mm^3) of intact non-fractionized tumor tissue (the "Tumor test group").

2.2. *Cell staining and measurements*

For intracellular FP measurements, cells were loaded onto an array of picoliter-volume wells (LiveCell Array, Molecular Cytomics, Boston, USA), designed to individually hold a few thousand cells, each held by a picoliter well. The retained cell can be then exposed to a variety of biochemical manipulations, while preserving its location and, as a result, its identity. Thus, data acquisition and analysis were carried out for each individual cell.

In the present study, cells retained in the array were exposed to increasing concentrations (0.6 μM, 1.2 μM, 2.4 μM, 3.6 μM) of non-fluorescent fluorescein diacetate (FDA). Neither electrically charged nor polarized, the hydrophobic FDA molecules easily diffuse through the plasma membrane into the cell. Within the cell, FDA is hydrolyzed by esterases into hydrophilic fluorescent fluorescein which accumulates in the cell.

In each individual cell, 6 sequential FP measurements per each of the 4 FDA concentrations were simultaneously taken at 2 emission wavelengths: 530 nm (hereafter referred to at the P1 parameter) and 580 nm (the P2 parameter). Thus, each cell was measured at 24 sequential time points, with an average interval of 40 s, yielding 24 × 2 (P1 and P2) variants of the FP parameters. This staining/measurement procedure was performed on each of the three tested cell groups (the Control, PHA, and Tumor test groups), generating the following 12 patterns: cells incubated without a stimulant, with FP measured at 530 nm emission (Control P1), cells incubated without a stimulant, with FP measured at 580 nm emission (Control P2), cells stimulated by PHA at 530 nm FP (PHA P1), cells stimulated by PHA at 580 nm FP (PHA P2), cells stimulated by tumor at 530 nm FP (Tumor P1), and cells stimulated by tumor at 580 nm FP (Tumor P2) – in healthy subjects as well as in patients. In other words, for each subject group – healthy and patients – 3 types of tests were performed, each at 2 wavelengths FP. Fig. 1 illustrates the measurement procedure and Fig. 2 presents mean FP values under different stimulation conditions and at different measurement time points.

2.3. *Mathematical analysis*

In order to determine the parameters and measurement conditions that can best differentiate between the test groups, i.e. select the most informative (discriminative) parameters for breast cancer diagnosis, it is necessary to perform their comparative analysis. The solution of the comparative analysis problem consists of two steps.

1. Estimating the correlation between each parameter in each group and the patterns studied.
2. Ranking the groups of parameters according to the estimation of the parameters' correlation with the patterns studied.

The first step consists in measuring the correlation between two random values.

In the present study, the normalized mutual information (the uncertainty coefficient) is chosen as the measure of correlation [5,6].

We convert the continuous random value into a discrete (binary) random value [23]. The conversion of quantitative parameters into binary parameters, using appropriately defined thresholds, allows studying the essential nature of the properties described by these quantitative parameters [24].

Let X be a discrete random value with a distribution function

X	x_1	x_2	...	x_n
Q	q_1	q_2	...	q_n

Entropy of random value X is

$$H(X) = -\sum_{i=1}^{n} q_i \log q_i$$

For 2 discrete random values: X, Y, the uncertainty coefficient (the normalized mutual information) equals [6,25,26]

$$r(X, Y) = \frac{H(X) + H(Y) - H(X, Y)}{H(Y)}$$

where $H(X)$, $H(Y)$, $H(X,Y)$ represent entropies of random values X, Y, and XY, respectively.

The uncertainty coefficient has the following properties [25,26]:

(i) $0 \le r \le 1$.

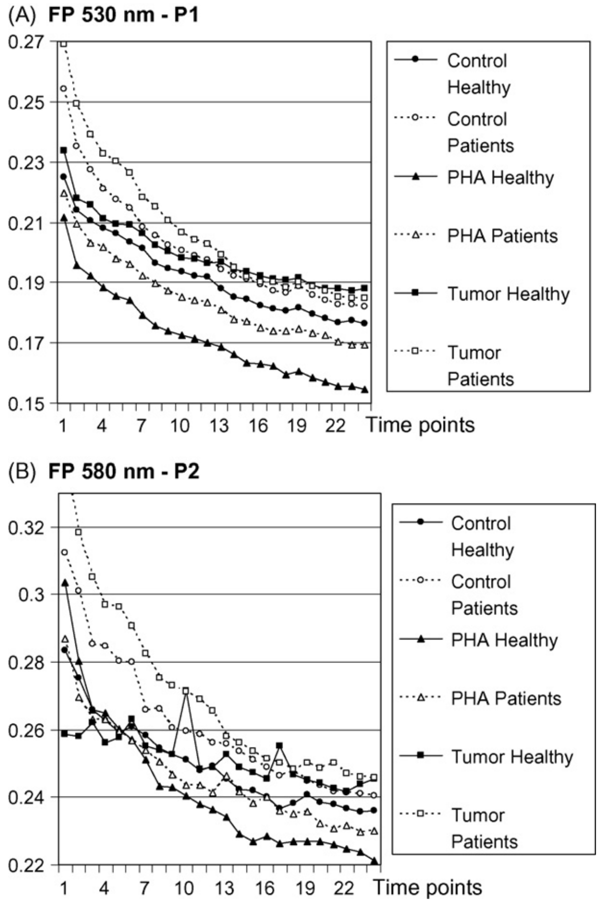


Fig. 2 – (A) Mean FP values determined at 530 nm (Parameter P1) and (B) mean FP measured at 580 nm (Parameter P2), in the "Control test" (circles), "PHA test" (triangles) and "Tumor test" (squares) groups at the 24 points of measurement, in the cells of healthy subjects (solid lines) and patients (dashed lines) (SD not shown). In either test, cells of healthy subjects tend to exhibit lower mean FP values. Different temporal divergence patterns can be visualized.

(ii) $r = 0$ if and only if X and Y are mutually independent (show a complete overlap).

(iii) $r = 1$ if and only if there exists a functional relationship between X and Y (a complete discrimination between the sets).

Let $P_0 = \{p_1, p_2, \ldots, p_m\}$ and $P_1 = \{p_{m+1}, p_{m+2}, \ldots, p_n\}$ be the measured parameter P in 2 patterns P_0, P_1 (for example, polarization P in PHA-activated cells derived from healthy subjects – P_0 vs. the patient group – P_1); $\bar{p}_0, \bar{p}_1, \sigma_0, \sigma_1$ – mean and standard deviation, correspondingly; and

$$b = \frac{\bar{p}_0\sigma_1 + \bar{p}_1\sigma_0}{\sigma_0 + \sigma_1}$$

We convert the continuous random value P into a discrete random value X:

$x_i = 0,\ \text{if}\, p_i \le b;\ \text{and}\, x_i = 1,\ \text{if}\ p_i > b.$

Considering the random values X, Y:

X	x_1	x_2	...	x_m	x_{m+1}	x_{m+2}	...	x_n
Y	0	0	...	0	1	1	...	1

"0" corresponds to pattern P_0, and "1" corresponds to P_1.

The measure of correlation between parameter P with patterns P_0 and P_1 is the uncertainty coefficient $r(X, Y)$.

3. Results

Sets of *uncertainty coefficients* were calculated for the FP values at each of the 24 measurement time points. The *uncertainty coefficients* are listed in Table 2. The closer is the *uncertainty coefficient* value to zero, the greater the similarity between the groups. On the other hand, the closer the *uncertainty* value to unity, the greater is the dissimilarity. 12 patterns are considered. For either the healthy subjects or patients, the patterns are: Control P1, Control P2, PHA P1, PHA P2, Tumor P1, Tumor P2.

Every coefficient in Table 2 corresponds to the parameters specified in the column heading, and measured at the time corresponding to the number of the row. Table 2 clearly indicates that the *uncertainty coefficients* have extrema at various time points. This raises the possibility of choosing the "most discriminative time point" of measurement through the implementation of Information Theory. Thus, notably, the difference between cells of healthy subjects and patients in PHA test (Parameter P1) tends to increase with time and FDA concentration (reaching an extremum at the twenty first measurement). Contrastingly, in Tumor test (Parameter P1), the difference (with an extremum at the 2nd measurement) is most marked shortly after the onset of measurements, at the lowest FDA concentration.

Nevertheless, the values of uncertainty coefficients at any time point do not exceed 0.2. This signifies that measurement at any single time point would not allow determining whether the measured cell belongs to a certain pattern. Therefore, it is necessary to consider the measurements over all 24 time points in every test.

Ranking corresponding to Table 2, with rank assignment by rows, yields Table 3. The greatest value of the uncertainty coefficient has the highest rank 1, while the smallest value has the rank 24.

Let us consider Table 3 as the Friedman statistical model [7] and examine the column effect of this table. Hypotheses: (H_0) There is no column effect ("null hypothesis"); (H_1) The null hypothesis is invalid.

Critical range. For the sample under consideration, $r = 24$, $c = 14$ (the numbers of rows and columns of Table 3, respectively). The sample is "large", therefore, the critical range is the upper 5%-range of χ^2_{13} distribution ($\text{df} = c - 1 = 13$).

Let us calculate the χ^2-criterion:

$$\chi^2 = \frac{12}{rc(c+1)}\left(\sum_{j=1}^{c} R_j^2 - \frac{\left(\sum_{j=1}^{c} R_j\right)^2}{c}\right)$$

where R_j is the sum of ranks of the jth column of Table 3. This gives $\chi^2 = 219.8$.

The critical range is $\chi^2_{13} > 22.4$. Since $219.8 > 22.4$, the null hypothesis with respect to Table 3 is rejected. Thus, according to the Friedman test, the column effect is found ($p = 0.05$). That is to say, there is a difference between the series' of measurements under consideration.

For multiple comparisons we use the method described by Conover [7]. Treatments i and j are considered different if the following inequality is satisfied:

$$\left|R_j - R_i\right| > t_{1-\alpha/2}\left[\frac{2\left(r\sum_{j=1}^{c}\sum_{i=1}^{r}[R(X_{ij})]^2 - \sum_{j=1}^{c}R_j^2\right)}{(r-1)(c-1)}\right]^{1/2}$$

where $t_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the **t** distribution with $(r-1)(c-1)$ degrees of freedom. The value for α is the same as used in the Friedman test. Further, we obtain $|R_j - R_i| > 31.6$.

By the multiple comparisons, we construct the clustering shown in Table 4. According to Table 4, based on the obtained coefficients and their ranks, the most informative parameters (i.e. parameters which can best discriminate between cells of BC patients and healthy individuals) are the FP parameters in PHA test at 530 nm – P1 (row 2.1) and in Tumor test at 580 nm – P2 (row 2.3). That is to say, the responses to PHA stimulation (as measured by FP at 530 nm) in cancer patients vs. healthy individuals differ most significantly; and also the responses to Tumor stimulation (as measured at 580 nm) differ very significantly.

Interestingly, the FP parameters in PHA test and Control test (both in the 530 nm "green" spectrum area) differ more significantly in the cells of healthy individuals as compared to a smaller difference in the patients (Table 4, row 1 vs. row 2.2, respectively). This signifies that cells of healthy subjects respond to PHA activation in a much greater degree than cells of BC patients. Similarly, the rating for Control P1 – Tumor P1 in the healthy subjects (row 3.2) is markedly different from the rating of Control P1 – Tumor P1 for the patients (row 6.1). That is to say, cells derived from healthy subjects respond to stimulation by tumor tissue in a much higher degree than cells of cancer patients. These phenomena may be also utilized for diagnostic purposes. Analogously, comparisons were conducted for groups of patients with DCIS only and IDC only. The

Table 2 – Uncertainty coefficients. Every uncertainty coefficient corresponds to the parameters specified in the column heading, and measured at the time corresponding to the number of the row.

No.	Healthy				Patients				Healthy vs. patients					
	Control test P1 vs. PHA. test P1	Control test P2 vs. PHA. test P2	Control test P1 vs. Tumor test P1	Control test P2 vs. Tumor test P2	Control test P1 vs. PHA. test P1	Control test P2 vs. PHA. test P2	Control test P1 vs. Tumor test P1	Control test P2 vs. Tumor test P2	Control test P1	Control test P2	PHA test P1	PHA test P2	Tumor test P1	Tumor test P2
1	0.03748	0.00549	0.00983	0.00022	0.06210	0.02022	0.00011	0.00077	0.02607	0.01766	0.00193	0.00243	0.05319	0.11781
2	0.06723	0.00024	0.00445	0.00096	0.04493	0.05298	0.00043	0.00048	0.03433	0.01719	0.01879	0.00443	0.04891	0.17564
3	0.08311	0.00471	0.00866	0.00430	0.07269	0.04574	0.00001	0.00753	0.01414	0.01672	0.01158	0.00037	0.02673	0.13417
4	0.07347	0.00183	0.01181	0.00184	0.05479	0.04028	0.00011	0.00390	0.04026	0.01956	0.03582	0.00105	0.04496	0.13220
5	0.06495	0.00790	0.00779	0.00392	0.06637	0.02291	0.00027	0.00240	0.02083	0.00397	0.02228	0.00139	0.01632	0.14465
6	0.10654	0.00790	0.00638	0.00011	0.06613	0.04159	0.00138	0.00390	0.02083	0.01722	0.05257	0.00012	0.01001	0.15408
7	0.11455	0.01682	0.01518	0.00803	0.03153	0.02202	0.00187	0.00621	0.00929	0.00079	0.05998	0.00340	0.00892	0.14616
8	0.09900	0.02073	0.01889	0.00000	0.02825	0.02291	0.00026	0.0118	0.01627	0.00374	0.05357	0.01396	0.01144	0.12832
9	0.12206	0.02993	0.01456	0.00237	0.03174	0.01575	0.00109	0.01832	0.01391	0.00364	0.05909	0.00438	0.00559	0.08005
10	0.11283	0.02708	0.01181	0.00015	0.04587	0.01922	0.00059	0.01109	0.01185	0.00930	0.04442	0.00731	0.00491	0.09360
11	0.12422	0.03413	0.01634	0.00123	0.03778	0.02097	0.00085	0.01422	0.01366	0.00840	0.07663	0.00987	0.00321	0.09449
12	0.12881	0.03094	0.01529	0.00071	0.04580	0.03536	0.00129	0.00852	0.00550	0.00577	0.07352	0.00323	0.00401	0.08365
13	0.08998	0.03630	0.02666	0.00294	0.02935	0.02929	0.00182	0.00852	0.01347	0.01165	0.05030	0.02170	0.00007	0.02274
14	0.09516	0.04836	0.02287	0.00462	0.03662	0.02004	0.00036	0.00346	0.00712	0.03764	0.05215	0.02330	0.00123	0.02522
15	0.10603	0.04704	0.04146	0.00028	0.03917	0.03928	0.00036	0.00272	0.01719	0.01407	0.06988	0.01763	0.00091	0.01907
16	0.12852	0.03969	0.03076	0.00044	0.03894	0.02938	0.00076	0.00960	0.01915	0.01554	0.05526	0.00659	0.00638	0.02529
17	0.10053	0.04205	0.04624	0.00312	0.04319	0.03778	0.00066	0.00147	0.01935	0.01994	0.08062	0.01322	0.00157	0.01525
18	0.08245	0.02536	0.03204	0.00276	0.02832	0.02730	0.00036	0.00778	0.01349	0.01766	0.06167	0.00858	0.00368	0.01437
19	0.12464	0.04960	0.04023	0.00498	0.3905	0.01505	0.00129	0.00788	0.02192	0.01208	0.07507	0.02235	0.00220	0.02308
20	0.10224	0.04214	0.03880	0.00815	0.02601	0.00725	0.00346	0.00474	0.01650	0.00621	0.09652	0.01193	0.00003	0.01577
21	0.11395	0.04101	0.03010	0.00418	0.04580	0.01274	0.00384	0.02530	0.01078	0.00454	0.12131	0.00410	0.00005	0.02434
22	0.13092	0.03887	0.04574	0.00021	0.02718	0.00712	0.00000	0.02511	0.01306	0.00464	0.10028	0.01577	0.00399	0.05050
23	0.11410	0.04471	0.03199	0.00863	0.02528	0.01350	0.00303	0.01930	0.00774	0.01201	0.09052	0.00454	0.00061	0.00493
24	0.11589	0.05809	0.04681	0.00797	0.03786	0.00469	0.00310	0.02148	0.00788	0.00935	0.08649	0.01094	0.00150	0.01174

Table 3 – Ranking of data in Table 2 for the Friedman model. Ranking corresponding to Table 2, with rank assignment by rows, yields this table. The greatest value of the uncertainty coefficient has the highest rank 1, while the smallest value has the rank 24.

No.	Healthy				Patients				Healthy–patients					
	Control test P1. PHA. test P1	Control test P2. PHA. test P2	Control test P1. Tumor test P1	Control test P2. Tumor test P2	Control test P1. PHA. test P1	Control test P2. PHA. test P2	Control test P1. Tumor test P1	Control test P2. Tumor test P2	Control test P1	Control test P2	PHA test P1	PHA test P2	Tumor test P1	Tumor test P2
1	4	9	8	13	2	6	14	12	5	7	11	10	3	1
2	2	14	9	11	5	3	13	12	6	8	7	10	4	1
3	2	11	9	12	3	4	14	10	7	6	8	13	5	1
4	2	12	9	11	3	5	14	10	6	8	7	13	4	1
5	3	8	9	10	2	4	14	13	6	11	5	12	7	1
6	2	9	10	14	3	5	12	11	6	7	4	13	8	1
7	2	6	7	10	4	5	13	11	8	14	3	12	9	1
8	2	6	7	14	4	5	13	10	8	12	3	9	11	1
9	1	5	8	13	4	7	14	6	9	12	3	11	10	2
10	1	5	8	14	3	6	13	9	7	10	4	11	12	2
11	1	5	7	13	4	6	14	8	9	11	3	10	12	2
12	1	6	7	14	4	5	13	8	10	9	3	12	11	2
13	1	3	6	11	4	5	13	12	9	10	2	8	14	7
14	1	3	8	11	5	9	14	12	10	4	2	7	13	6
15	1	3	4	14	6	5	13	11	9	10	2	8	12	7
16	1	3	5	14	4	6	13	10	8	9	2	11	12	7
17	1	5	3	11	4	6	14	13	8	7	2	10	12	9
18	1	6	3	13	4	5	14	11	9	7	2	10	12	8
19	1	3	4	12	5	9	14	11	8	10	2	7	13	6
20	1	3	4	9	5	10	13	12	6	11	2	8	14	7
21	2	4	5	11	3	8	13	6	9	10	1	12	14	7
22	1	5	4	13	6	10	14	7	9	11	2	8	12	3
23	1	3	4	9	5	7	13	6	10	8	2	12	14	11
24	1	3	4	10	5	12	13	6	11	9	2	8	14	7
Total	36	140	152	287	97	153	322	237	193	221	84	245	252	101

Table 4 – Clustering for all healthy subjects and patients, according to data in Table 3.

	The degree of dissimilarity	Group of subjects	Testing parameter	Sum of ranks
1	Large difference	Healthy	Control P1 vs. PHA.P1	36
2.1	Intermediate difference 1	Healthy vs. patients	PHA.P1	84
2.2		Patients	Control P1 vs. PHA.P1	97
2.3		Healthy vs. patients	Tumor P2	101
3.1	Intermediate difference 2	Healthy	Control P2 vs. PHA.P2	140
3.2		Healthy	Control P1 vs. Tumor P1	152
3.3		Patients	Control P2 vs. PHA.P2	153
4.1	Small difference 1	Healthy vs. patients	Control P1	193
4.2		Healthy vs. patients	Control P2	221
4.3		Patients	Control P2 vs. Tumor P2	237
4.4		Healthy vs. patients	PHA.P2	245
4.5		Healthy vs. patients	Tumor P1	252
5	Small difference 2	Healthy	Control P2 vs. Tumor P2	287
6	Small difference 3	Patients	Control P1 vs. Tumor P1	322

Table 5 – Clustering for all healthy subjects and DCIS patients only.

	The degree of dissimilarity	Group of subjects	Testing parameter	Sum of ranks
1	Large difference	Healthy	Control P1 vs. PHA.P1	34
2.1	Intermediate difference 1	Healthy vs. patients	PHA.P1	73
2.2		Patients	Control P1 vs. PHA.P1	87
3.1	Intermediate difference 2	Healthy vs. patients	Tumor P2	126
3.2		Healthy	Control P1 vs. Tumor P1	138
3.3		Healthy	Control P2 vs. PHA.P2	143
3.4		Patients	Control P2 vs. PHA.P2	158
4.1	Small difference 1	Healthy vs. patients	Control P1	197
4.2		Healthy vs. patients	Control P2	225
4.3		Patients	Control P2 vs. Tumor P2	241
4.4		Healthy vs. patients	PHA.P2	246
4.5		Healthy vs. patients	Tumor P1	255
4.6		Healthy	Control P2 vs. Tumor P2	278
5	Small difference 2	Patients	Control P1 vs. Tumor P1	319

results of comparison are shown in Tables 5 and 6, correspondingly. The items in Table 6 that correspond to the "largest" differences between measurement series' (the first 3 lines of the Table) form separate clusters. That is to say, when considering IDC patients separately, the differences between series' "Healthy Control P1" vs. "Healthy PHA P1," "Healthy PHA P1" vs. "Patients PHA P1," "Patients Control P1" vs. "Patients PHA P1" are confidently distinct from each other. In the clustering involving both DCIS and IDC patients (Table 4), or only DCIS patients (Table 5), the differences between the series' "Healthy PHA P1" vs. "Patients PHA P1" and "Patients Control P1" vs. "Patients PHA P1" are in the same cluster. That is, for DCIS and IDC patients, as well as for DCIS patients only, the differences between the series' "Healthy PHA P1" vs. "Patients PHA P1" and "Patients Control P1" vs. "Patients PHA P1" are the same. In contrast, for IDC patients, the difference between "Patients Control P1" vs. "Patients PHA P1" is less than the difference "Healthy PHA P1" vs. "Patients PHA P1." The found

Table 6 – Clustering for all healthy subjects and IDC patients only.

	The degree of dissimilarity	Group of subjects	Testing parameter	Sum of ranks
1	Large difference	Healthy	Control P1 vs. PHA.P1	33
2	Intermediate difference 1	Healthy vs. patients	PHA.P1	71
3	Intermediate difference 2	Patients	Control P1 vs. PHA.P1	108
4.1	Small difference 1	Healthy vs. patients	Control P1	146
4.2		Healthy vs. patients	Control P2	149
4.3		Healthy vs. patients	Tumor P2	153
4.4		Healthy	Control P2 vs. PHA.P2	160
5.1	Small difference 2	Healthy	Control P1 vs. Tumor P1	204
5.2		Patients	Control P2 vs. PHA.P2	207
5.3		Patients	Control P2 vs. Tumor P2	216
5.4		Healthy vs. patients	PHA.P2	228
5.5		Healthy vs. patients	Tumor P1	235
6.1	Small difference 3	Healthy	Control P2 vs. Tumor P2	289
6.2		Patients	Control P1 vs. Tumor P1	321

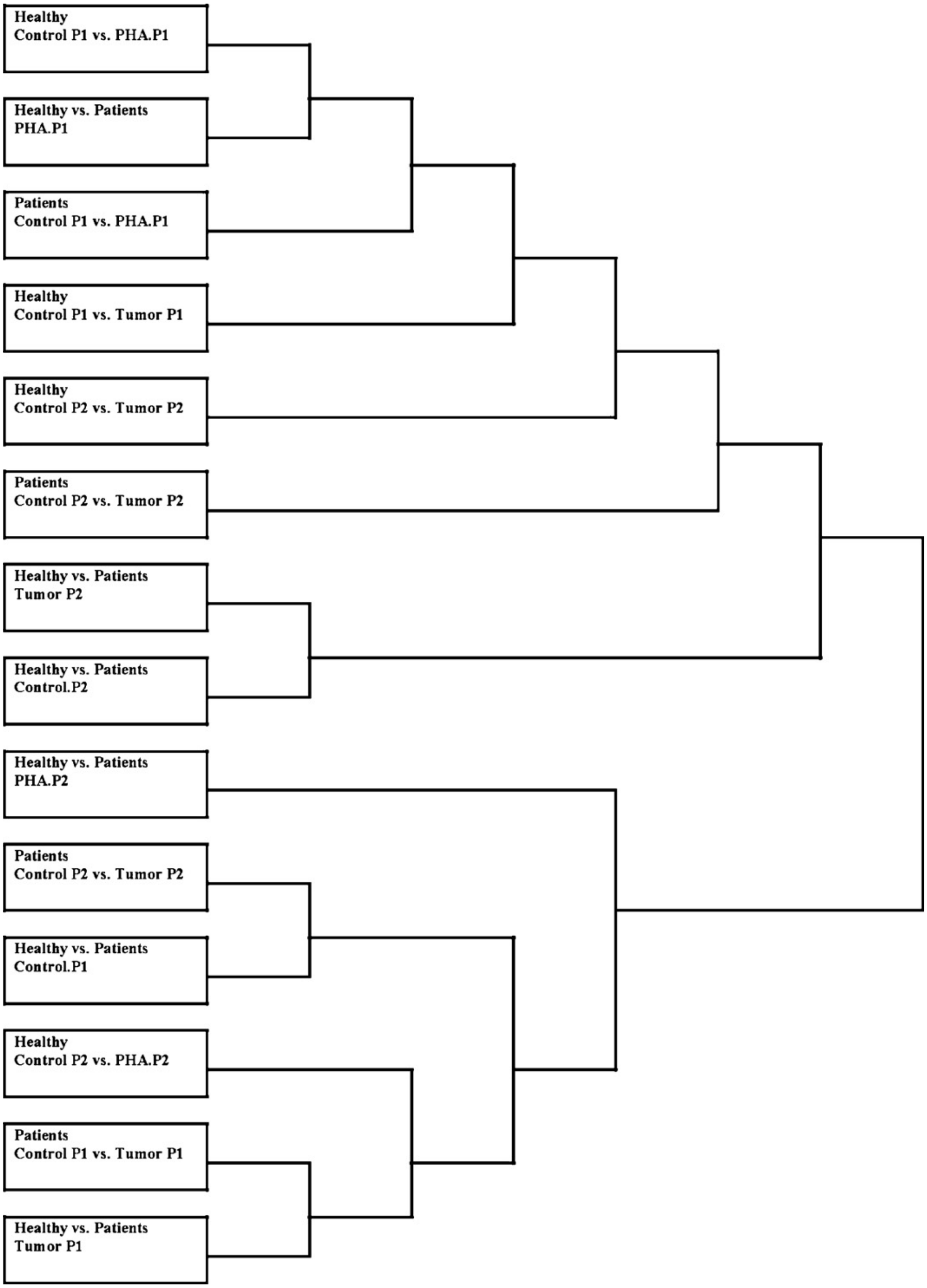

Healthy Control P1 vs. PHA.P1
Healthy vs. Patients PHA.P1
Patients Control P1 vs. PHA.P1
Healthy Control P1 vs. Tumor P1
Healthy Control P2 vs. Tumor P2
Patients Control P2 vs. Tumor P2
Healthy vs. Patients Tumor P2
Healthy vs. Patients Control.P2
Healthy vs. Patients PHA.P2
Patients Control P2 vs. Tumor P2
Healthy vs. Patients Control.P1
Healthy Control P2 vs. PHA.P2
Patients Control P1 vs. Tumor P1
Healthy vs. Patients Tumor P1

Fig. 3 – Clustering by the Hierarchical Cluster Analysis.

differences may emphasize the well acknowledged fact of a changed responsiveness of peripheral lymphocytes of cancer patients to different stimulants (reflecting a suppression of immune response) [27–29], and also correspond to the widely discussed assumption that invasive (IDC) and non-invasive (DCIS) forms of breast cancer represent pathological states with different origin, pathogenesis and behavior [30,31].

The data were also processed by the Hierarchical Cluster Analysis method, using the SPSS package [32]. In this procedure, the clustering parameters are the normalized means differences, the inclusion method is the “Nearest neighbor” and the distance measure is the “Euclidian distance”. Fig. 3 shows the scheme of Hierarchical Cluster Analysis. According to this analysis, the “Healthy. Control P1 vs. PHA.P1” and the “Healthy vs. Patients. PHA.P1” groups are in the same cluster. However, the difference between normal lymphocytes and lymphocytes that underwent mitogenic stimulation in healthy subjects should be greater than the difference between PHA-stimulated lymphocytes in healthy subjects and PHA-stimulated lymphocytes in BC patients, since one of the important steps of tumorigenesis is the specific immune response favoring tumor invasion, which involves the suppression of functional activity of peripheral blood mononuclear cells (PBMC): the inhibition of specific T-cell response, suppression of effector antitumor mechanisms, violation of cell signaling, cell cycle and apoptosis regulation. However, nonspecific stimulation of peripheral lymphocytes by PHA is partially maintained even in advanced cancer stages [27–29]. That is, the "Healthy. Control P1 vs. PHA.P1" and the "Healthy vs. Patients. PHA.P1" should be in different clusters. Therefore the proposed method provides a more adequate clustering than Hierarchical Cluster Analysis.

4. Discussion

The analysis of the correlation measure between various parameters is an important part of many biomedical studies. However, as has been noted earlier [33], the majority of statistical measures of correlation do not have a theoretical substantiation. Many classification methods are heuristic [34,35]. The method proposed here offers a theoretical substantiation of the selection process within the framework of Information Theory and Nonparametric Statistics. The major statistic measure of correlation between various parameters is the correlation coefficient. This coefficient serves in estimating the linear correlation between parameters. In complex heterogeneous biophysical systems, however, the correlations between parameters are much more complex than merely a linear correlation [24,36–38], in oncology in particular [39]. Therefore, the application of the correlation coefficient in the study of bio-systems often yields unsatisfactory results. In contrast, the informatic measures of correlation/linkage permit estimating the non-linear relations between parameters. This is the reason why Information Theory can be successfully applied in the research of biomedical systems. The advantage of ranking criteria consists in their independence of normality assumption [40]. The latter property entails the wide application of ranking criteria in the analysis of biomedical data [41–43].

The present study investigates the correlation between the values of FP measured at different wavelengths, under different stimulation conditions (by PHA and tumor), and the disease status. By establishing the correlation, we can estimate a boundary between the FP-indicated responses in healthy subjects and cancer patients, and, at a later stage of research, determine whether a particular cellular FP value can be attributed to a distinct disease pattern.

Despite the limited sample size, the present comparative analysis has shown that the most informative (best discriminating between the healthy and patient groups) and thus probably the most suitable for diagnostic purposes are the parameters FP in PHA test (measured at 530 nm) and Tumor test (measured at 580 nm). The fact that the significant differences between the populations are detected in different stimulation tests (PHA and Tumor) at different emission wavelengths (530 nm and 580 nm, respectively) may be due to intracellular domain specific response to the different kinds of stimulation. The domain specific pH, polarity, polarizability, viscosity, etc. can cause the fluorescence emission red shift. As a result, a cell, stained with a mono-fluorophore, appears as if stained with an ensemble of 'dissimilar probes.' In particular, if these seemingly dissimilar probes experience different levels of viscosity, then an emission wavelength FP dependency for the entire stained cell may be anticipated [44].

It was also found that the difference between cells of healthy subjects and patients, as measured in the PHA test at 530 nm, is most marked at a later stage of measurement, at the highest fluorophore concentration; whereas in the Tumor test at 580 nm, the difference was largest at the onset of measurements, at the lower fluorophore concentration. These findings may also suggest differential pathways of activation, possibly leading to local changes in probe mobility restrictions that are detected by FP changes. In either of the three tests, the cells of healthy subjects tended to exhibit lower FP values than patients. This phenomenon may indicate a greater responsiveness to activation by the cells of healthy subjects.

In addition, the analysis has confirmed that the FP parameters in PHA test and Control test (both in the 530 nm "green" spectrum area) differ significantly in the cells of healthy individuals as compared to a smaller difference in the patients (Table 4, row 1 vs. row 2.2). Similarly, the rating for Control P1 – Tumor P1 for the healthy subjects (Table 4, row 3.2) is significantly different from the rating of Control P1 – Tumor P1 for the patients (Table 4, row 6.1). Thus, in both cases, cells derived from healthy subjects respond to stimulation (either by PHA or tumor tissue) in a much higher degree than cells of cancer patients. This behavior may be utilized as yet another diagnostic feature.

Once the most informative (discriminative) parameters are established, the following step of research should concentrate on determining the rules for cancer detection according to the particular values of FP parameters in PHA test (P1) and Tumor test (P2), obtained in the patient cells. Other formal parameters can be used for diagnosing other kinds of disease in a way methodologically similar to the proposed in the present study.

Conflict of interest

None.

Acknowledgements

This research was supported by The Komen Foundation, The Horowitz Foundation and by US Army Medical Research and Material Command Grant DAMD17-01-1-0131.

REFERENCES

[1] W.H. Wolberg, O.L. Mangasarian, Multisurface method of pattern separation for medical diagnosis applied to breast cytology, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 9193–9196.

[2] O.L. Mangasarian, W.N. Street, W.H. Wolberg, Breast cancer diagnosis and prognosis via linear programming, Oper. Res. 43 (1995) 570–577.

[3] I.M. Gelfand, B.I. Rosenfeld, M.A. Shifrin, Essays on Collaboration of Mathematicians and Physicians, Nauka, Moscow, 1989 (in Russian).

[4] C.J.D.M. Verhayen, R.P.W. Duin, F.C.A. Groen, J.C. Joosen, P.W. Verbeek, Progress report on pattern recognition, Rep. Prog. Phys. 43 (1980) 785–831.

[5] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.

[6] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[7] W.J. Conover, Practical Nonparametric Statistics, Wiley, New York, 1999.

[8] R.A. Gatenby, B.R. Frieden, Application of Information Theory and extreme physical information to carcinogenesis, Cancer Res. 62 (2002) 3675–3684.

[9] D. Blokh, N. Zurgil, I. Stambler, E. Afrimzon, Y. Shafran, E. Korech, J. Sandbank, M. Deutsch, An information-theoretical model for breast cancer detection, Methods Inf. Med. 47 (2008) 322–327.

[10] D. Blokh, I. Stambler, E. Afrimzon, Y. Shafran, E. Korech, J. Sandbank, N. Zurgil, M. Deutsch, The information-theory analysis of Michaelis-Menten constants for detection of breast cancer, Cancer Detect. Prev. 31 (2007) 489–498.

[11] I.K. Sethi, G.P.R. Sarvarayudu, Hierarchical classifier design using mutual information, IEEE Trans. Pattern Anal. Mach. Intell. 4 (1982) 441–445.

[12] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Trans. Neural Netw. 5 (1994) 537–550.

[13] P.S. Bradley, O.L. Mangasarian, W.N. Street, Feature selection via mathematical programming, INFORMS J. Comput. 10 (1998) 209–217.

[14] H. Lu, R. Setiono, Feature selection via discretization, IEEE Trans. Knowledge Data Eng. 9 (1997) 642–645.

[15] S.A. Glantz, Primer of Biostatistics, Fourth edition, McGraw-Hill, New York, 1994.

[16] W.H. Wolberg, Inhibition of migration of human autogenous and allogeneic leukocytes by extracts of patients' cancers, Cancer Res. 31 (1971) 798–802.

[17] H.M. Shapiro, Practical Flow Cytometry, 3rd ed., Alan R. Liss, New York, 1995.

[18] A. Ben-Ze'ev, A.D. Bershadsky, The role of the cytoskeleton in adhesion-mediated signaling and gene expression, Adv. Mol. Cell. Biol. 24 (1997) 125–163.

[19] B. Geiger, D. Rosen, G. Berke, Spatial relationships of MTOC and the contact area of cytotoxic T lymphocytes, J. Cell Biol. 95 (1982) 137–143.

[20] K. Dimitropoulos, J.M. Rolland, R.C. Nairn, Flow cytofluorimetry of fluorescein fluorescence polarization to assay lymphocyte activation, Biochem. Biophys. Res. Commun. 136 (1986) 1021–1029.

[21] M. Kaplan, E. Trebnyikov, G. Berke, Fluorescence depolarization as an early measure of T lymphocyte stimulation, J. Immunol. Methods 201 (1997) 15–24.

[22] G.D. Tourassi, E.D. Frederick, C.E. Floyd Jr., Application of the mutual information criterion for feature selection in computer-aided diagnostics, Med. Phys. 28 (2001) 2394–2402.

[23] W.C. Chou, M.A. Neifeld, R. Xuan, Information-based optical design for binary-valued imagery, Appl. Opt. 39 (2000) 1731–1742.

[24] G. Nicolis, I. Prigogine, Exploring Complexity, W.H. Freeman, New York, 1990.

[25] A. Renyi, On measures of dependence, Acta Math. Acad. Sci. Hung. 10 (1959) 441–451.

[26] J. Zvarova, M. Studeny, Information theoretical approach to constitution and reduction of medical data, Int. J. Med. Inf. 45 (1997) 65–74.

[27] J.F. Head, R.L. Elliott, J.L. McCoy, Evaluation of lymphocyte immunity in breast cancer patients, Breast Cancer Res. Treat. 26 (1993) 77–88.

[28] C. Wiltschke, M. Krainer, A.C. Budinsky, A. Berger, C. Muller, R. Zeillinger, et al., Reduced mitogenic stimulation of peripheral blood mononuclear cells as a prognostic parameter for the course of breast cancer: a prospective study, Br. J. Cancer 71 (1995) 1292–1296.

[29] R.H. Schwartz, A cell culture model for T lymphocyte clonal anergy, Science 248 (1990) 1349–1356.

[30] M. Ignatiadis, C. Sotiriou, Understanding the molecular basis of histologic grade, Pathobiology 75 (2008) 104–111.

[31] R. Roylance, P. Gorman, W. Harris, R. Liebmann, D. Barnes, A. Handy, D. Sheer, Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer, Cancer Res. 59 (1999) 1433–1436.

[32] A. Buhl, P. Zofel, SPSS Version 10, Addison-Wesley, New York, 2001.

[33] G.J.G. Upton, The Analysis of Cross-tabulated Data, Wiley, New York, 1978.

[34] B.S. Duran, P.L. Odell, Cluster Analysis. A Survey, Springer-Verlag, Berlin, 1977.

[35] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley InterScience, New York, 2001.

[36] U. Rajendra Acharya, O. Faust, N. Kannathal, Tji Leng Chua, S. Laxminarayan, Non-linear analysis of EEG signals at various sleep stages, Comput. Methods Programs Biomed. 80 (2005) 37–45.

[37] S. Cerutti, G. Carrault, P.J.M. Cluitmans, A. Kinie, T. Lipping, N. Nikolaidis, I. Pitas, M.G. Signorini, Non-linear algorithms for processing biological signals, Comput. Methods Programs Biomed. 51 (1996) 51–73.

[38] D. Blokh, E. Afrimzon, I. Stambler, E. Korech, Y. Shafran, N. Zurgil, M. Deutsch, Breast cancer detection by Michaelis–Menten constants via linear programming, Comput. Methods Programs Biomed. 85 (2007) 210–213.

[39] R.Y. Chandawarkar, D.P. Guyton, Oncologic mathematics: evolution of a new specialty, Arch. Surg. 137 (2002) 1428–1434.

[40] B.L. van der Waerden, Mathematical Statistics, Springer-Verlag, Berlin, 1969.

[41] B.L. Strom, P.L. Hibberd, K.A. Soper, P.D. Stolley, W.L. Nelson, International variations in epidemiology of cancers of the extrahepatic biliary tract 1, Cancer Res. 45 (1985) 5165–5168.

[42] M.R. Sheldon, M.J. Fillyaw, W.D. Thompson, The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs, Physiother. Res. Int. 1 (1996) 221–228.

[43] V. Bewick, L. Cheek, J. Ball, Statistics review 10: further nonparametric methods, Crit. Care 8 (2004) 196–199.

[44] Y. Yishai, D. Fixler, M. Cohen-Kashi, N. Zurgil, M. Deutsch, Ratiometric fluorescence polarization as a cytometric functional parameter: theory and practice, Phys. Med. Biol. 48 (2003) 2255–2268.