

Title	Computer Analysis of Bone Tumor Roentgenograms Using Discriminant Functions
Author(s)	松林, 隆; Lodwick, S. Gwilym
Citation	日本医学放射線学会雑誌. 1971, 31(9), p. 1007-1025
Version Type	VoR
URL	https://hdl.handle.net/11094/18116
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

Computer Analysis of Bone Tumor Roentgenograms Using Discriminant Functions

By

Takashi Matsubayashi, M.D.*, and Gwilym S. Lodwick, M.D.

Research Code No: 206

Key Words: Computer, Bone Tumors, Feature Selection, Discriminant Analysis

電子計算機による骨腫瘍 X 線像の判別関数分析

ミズーリ大学医学部放射線科

松 林 隆* Gwilym S. Lodwick

(G.S.L.: 主任教授, *現在北里大学医学部放射線医学教室)

(昭和46年 8 月30日 受付)

組織診断の確定している、ユーイング肉腫、線維肉腫、細網肉腫、骨肉腫の4種類の原発性骨腫瘍のX線所見をコード化し、電子計算機を使つてカイ自乗テストによる特徴選択を試みた。つぎに特徴選択の結果に基づいて線型判別分析を行ない、4種類の腫瘍鑑別のためのプログラムを作成

し、電算機診断を試みた。特徴選択、判別分析に使つた骨腫瘍 401例の Training group と、テスト用骨腫瘍 198例の Testing group について電算機による鑑別診断を試みたが、組織診断と電算機診断の一致率は前者では78%、後者では70%であつた。

The goal of computer research in diagnostic radiology is to make such machines that can recognize and interpret radiographic images and make correct diagnoses in place of radiologists. At first they should be trained by human beings, but eventually they will grow up to learn by themselves from their own experiences; that is, they are learning machines. For the sake of explanatory convenience, learning machines are considered to have two main processes⁽¹⁾⁽¹⁰⁾. The first is that of extracting features, or radiological findings, from the input radiographic images, and the second is that of making decisions on the diagnosis assignment to the input images in some way based on the radiological features extracted in the first process. Indeed both processes are inseparably related to each other, but computer approaches have been easier for the latter with some triable mathematical methods. So far clinical diagnostic radiology has contributed mainly to discovering and systematizing the radiological features for each disease category and the latter process has been inclined to be thought as intangible or intuitive, and the amount

From the Department of Radiology (G.S.L., Professor and Chairman), University of Missouri School of Medicine, Columbia, Mo., U.S.A.

*Present Address: Department of Radiology, Kitasato University School of Medicine, Sagami-hara, Japan. Presented at the Thirtieth Annual Meeting of the Japan Radiological Society, Tokyo, Japan, April 3-4, 1971.

of medical knowledge was so small as to make diagnosis almost intuitively. However, the number of new facts and methods we are now getting in medicine is increasing at an unprecedented rate. In this situation, the concept of computer-aided diagnosis⁵⁾⁸⁾, a trial to combine computers as a helper in decision-making with physicians as the best feature extractor, has been suggested and tried. This is a preliminary stage to complete learning machines in medical diagnosis.

For the purpose of making decisions most correctly and efficiently based on the present, vastly increasing amount of knowledge, we have to select the most effective set of symptoms and findings from what appear important. This is the problem of feature or measurement selection. A trial of feature selection was made in the roentgenologic examination of four categories of primary malignant bone tumors which are Ewing's sarcoma, fibrosarcoma, reticulum cell sarcoma and osteosarcoma, by analysing the coded information which had been punched into the IBM data cards for each patient, using 2×2 contingency tables and the chi-square test. Based on the results of feature selection, linear discriminant function analyses were tried for each pair of categories of bone tumors in two different ways: one is derived from assuming multivariate normal distributions¹⁰⁾¹⁵⁾, another is from using estimating equations²⁾.

The effect of feature selection should be evaluated from the results of the decision-making which is based on the selected features. For the purpose of this evaluation, features were selected at several tentative critical levels of the chi-square values, and the effect of selection was examined in association with the results of linear discriminant analyses.

Then diagnostic classification or differentiation was tried using the results of linear discriminant analyses in both the training group and the testing group of bone tumor cases. The differences of the results in differential diagnoses were examined between these two groups.

Material

This study was concerned with four categories of primary malignant bone tumors, which are Ewing's sarcoma, fibrosarcoma, reticulum cell sarcoma and osteosarcoma. First, a total of 599 histologically proven cases of these bone tumors was examined using the 2×2 contingency table analysis and the chi-square test for the purpose of feature selection, that is, estimating the importance of each radiological finding for discrimination between each pair of disease categories. Based on the results of feature selection, each pair of disease categories was examined using linear discriminant analyses, where 582 cases were available because 17 cases were discarded due to insufficient coding. Then using the results

Table 1. Grouping of material cases of bone tumors

Category	Total	Training	Testing
Ewing's sarcoma	179	119	60
Fibrosarcoma	151	101	50
Reticulum cell sarcoma	43	30	13
Osteosarcoma	226	151	75
Sum	599	401	198

Table 2. Items and questions about the radiological findings and the age of primary bone tumor cases. The random variable X_i has a value of unity when the response to the i -th question is yes, and it has a value of zero when the response to the i -th question is no.

No. (i)	Questions of the "yes-no" type about radiological findings plus the age
1. Age,	00—09 years
2.	10—14 years
3.	15—19 years
4.	20—29 years
5.	30—39 years
6.	40 years and over
7. Tumor size	01—30 mm.
8.	31—60 mm.
9.	61—90 mm.
10.	91 mm. and over
11. Shape, round (Length Less Than $1.5 \times$ Width)	$\left\{ \begin{array}{l} \text{if round, } x_{11} = 1. \\ \text{if elongated, } x_{11} = 0. \end{array} \right.$
12. Location,	central
13.	eccentric
14.	cortex or parosteal
15. Bone type,	tubular
16.	flat
17.	small
18.	sacrum or pelvis
19. Epiphysis,	involved
20. Growth plate,	involved
21. Articular cortex,	involved
22. Metaphysis	involved
23. Shaft,	involved
24. Radiolucency,	present
25. Bone destruction pattern,	absent
26.	geographic with a regular margin
27.	geographic with a lobulated margin
28.	geographic with a metamorphic margin
29.	geographic with moth-eaten zone more than 1 cm. in width
30.	geographic with moth-eaten zone less than or equal to 1 cm. in width
31.	geographic and permeated
32.	moth-eaten only
33.	moth-eaten and permeated
34.	permeated only
35.	geographic, moth-eaten and permeated
36. Fracture and displacement,	both absent
37. Fracture,	present
38. Displacement,	present
39. Penetration of cortex,	absent
40.	partial present

41.	total present
42. Tumor matrix mineralization,	flocculent present
43.	solid present
44.	lumps present
45.	clouds present
46. Sclerotic rim,	present
47. Mottling,	present
48. Endostosis,	present
49. Hyperostosis,	present
50. Buttress,	present
51. Septa,	present
52. Expanded shell,	absent
53. Codman's triangle,	absent
54.	one present
55.	two present
56.	three or more present
57. Periostosis,	absent
58.	laminated present
59.	amorphous present
60. Spiculation,	absent
61.	sunburst present
62.	hair-on-end present
63.	velvet present

of linear discriminant analyses, differential diagnosis or diagnostic classification was tried for 576 cases of these four categories, where 23 cases, 3.8 per cent of the total, were excluded because of insufficient coding for the classification.

Next, the total of 599 cases was divided into two parts, the training group and the testing group, according to only the order in which cases had been collected. The former has 401 cases and the latter has 198 cases. This relation is shown in Table 1. The training group was analysed using the 2×2 contingency table and the chi-square test for the purpose of feature selection. Based on the results, each pair of categories was examined using linear discriminant analyses, where 391 cases of the training group were used while 10 cases were unavailable because of insufficient coding. Using the results of discriminant analyses, both 387 cases of the training group and 189 cases of the testing group were classified into the four disease categories. Because of insufficient coding, 14 cases (3.5%) and 9 cases (4.5%) were excluded from each of the training group and the testing group respectively.

Selection of Radiological Features

As for each pair of bone tumor categories, the responses to 63 questions of the "yes-no" type were obtained, for 25 items of radiological findings and the age of patients, by having an IBM-360 computer manipulate the data cards into which the coded information had been punched according to the coding system devised and developed at the University of Missouri Medical Center⁽⁶⁾⁽⁸⁾. The items and questions are tabulated as Table 2. Each question is given a number (i) shown in the table. Then the

responses can be expressed mathematically, in other words, measured as that the random variable X_i has a value of unity when the response to the i -th question is yes, and it has a value of zero when the response to the i -th question is no. For example, if a patient is 8 years old and his left femur is involved, $x_1=x_{15}=1$ and $x_2=x_3=x_4=x_5=x_6=x_{16}=x_{17}=x_{18}=0$.

The frequencies of the response, yes or no, to each question were counted to make a 2×2 contingency table and the difference of the frequencies existing between each pair of disease categories was evaluated for each question by computing the chi-square value for the 2×2 contingency table. In Table 3 a 2×2

Table 3. A 2×2 contingency table

	Category 1	Category 2	Sum
Yes	a_1	a_2	N_A
No	b_1	b_2	N_B
Sum	N_1	N_2	N

contingency table is shown, and the simple formula for computing the chi-square value (χ^2) is as follows.

$$\chi^2 = \frac{(a_1 + a_2 + b_1 + b_2)(a_1 b_2 - a_2 b_1)^2}{(a_1 + b_1)(a_2 + b_2)(a_1 + a_2)(b_1 + b_2)} = \frac{N(a_1 b_2 - a_2 b_1)^2}{N_1 N_2 N_A N_B}$$

where a_1 and a_2 are the frequencies of cases in Category 1 and 2 whose responses to a given question are yes, and b_1 and b_2 are the frequencies of cases in Category 1 and 2 whose responses to the same question are no, $N = a_1 + a_2 + b_1 + b_2$, $N_1 = a_1 + b_1$, $N_2 = a_2 + b_2$, $N_A = a_1 + a_2$, $N_B = b_1 + b_2$.

The larger the chi-square value is, the greater the difference of frequencies existing between two categories is. Now then, we are not examining the difference of frequencies of each finding existing between the two categories, but we want to know the discriminating power of each finding for discrimination between the two categories. A computer program was made to pick up questions or random variables which have the larger chi-square values than several tentative critical levels. The values of 10.83, 6.63 and 3.84 were taken for all pairs of disease categories. These are the critical values in the chi-square test for one degree of freedom at the 0.1%, 1% and 5% levels of significance, respectively. As for the pair of Ewing's sarcoma and fibrosarcoma, the values of 40, 30 and 20 were added as the critical values. Thus, applying the 2×2 contingency table and the chi-square test to the questions listed in Table 2, a trial of radiological feature selection was made for discrimination between each pair of bone tumor categories. The results of this trial for the total of 599 cases are shown in Table 4 where the selected questions or random variables are listed by their number for each pair of categories and for each critical chi-square value.

Discriminant Analysis

Supposing that n radiological features are to be measured from each patient, each patient is represented as a set of n features or measurements. Each set of n features can be considered as a vector $X = (x_1, x_2, \dots, x_n)$, or a point in the n -dimensional Euclidean space E^n called a vector space, which may be referred to as a patient vector in a patient vector space⁽¹⁾. The rectangular coordinates of the point are the real numbers x_1, x_2, \dots , and x_n . In this study they are binary numbers, that is, each $x_i = 1$ or 0.

Table 4. Feature selection in the total group of 599 cases. Questions were selected for each pair of categories at various critical chi-square values. Each question is expressed by its number (i).

Critical χ^2	Ewing's sarcoma—fibrosarcoma																			
40	6	19	34	47	58															
30	6	19	27	30	34	46	47	58	59	60										
20	1 59	2 60	6 62	11	15	19	22	27	30	34	36	39	41	43	46	47	48	52	53	58
10.83 (0.1% level)	1 45	2 46	5 47	6 48	11 51	12 52	13 53	15 58	16 59	19 60	21 62	22	27	30	34	36	38	39	41	43
6.63 (1 % level)	1 42	2 43	5 45	6 46	11 47	12 48	13 51	15 52	16 53	19 55	21 58	22 59	27 60	30 62	34	36	37	38	39	41
3.84 (5 % level)	1 36 60	2 37 61	3 38 62	5 39	6 41	9 42	10 43	11 45	12 46	13 47	15 48	16 51	18 52	19 53	21 54	22 55	27 56	30 57	31 58	34 59
Critical χ^2	Ewing's sarcoma—reticulum cell sarcoma																			
10.83 (0.1% level)	6	19	21	59																
6.63 (1 % level)	1	3	5	6	19	21	43	53	55	59	60	62								
3.84 (5 % level)	1 60	2 62	3	5	6	11	12	13	19	21	33	36	37	43	47	48	53	55	58	59
Critical χ^2	Ewing's sarcoma—osteosarcoma																			
10.83 (0.1% level)	1	11	15	16	19	22	24	25	34	39	43	45	47	48	51	61	62			
6.63 (1 % level)	1 47	3 48	11 51	13 53	15 54	16 61	18 62	19	21	22	24	25	26	31	34	39	41	42	43	45
3.84 (5 % level)	1 40	3 41	11 42	12 43	13 45	15 47	16 48	17 51	18 53	19 54	21 60	22 61	23 62	24	25	26	27	31	34	39
Critical χ^2	fibrosarcoma—reticulum cell sarcoma																			
10.83 (0.1% level)	12	13	15	16	34	41	47													
6.63 (1 % level)	12	13	15	16	22	27	34	39	41	46	47	51	52	58						
3.84 (5 % level)	12	13	15	16	22	26	27	30	34	38	39	41	45	46	47	50	51	52	58	
Critical χ^2	fibrosarcoma—osteosarcoma																			
10.83 (0.1% level)	2 52	3 53	5 54	6 55	19 58	24 59	25 60	26 61	27	30	34	36	37	38	39	41	43	45	46	51
6.63 (1 % level)	2 52	3 53	5 54	6 55	19 58	24 59	25 60	26 61	27 62	30	34	36	37	38	39	41	43	45	46	51
3.84 (5 % level)	2 51	3 52	5 53	6 54	19 55	24 57	25 58	26 59	27 60	30 61	34 62	36	37	38	39	41	43	45	46	50
Critical χ^2	reticulum cell sarcoma—osteosarcoma																			
10.83 (0.1% level)	3	5	6	13	15	16	22	25	43	45	47	53	59	60						
6.63 (1 % level)	3	5	6	12	13	15	16	22	24	25	43	45	47	53	55	59	60	61		
3.84 (5 % level)	2 47	3 53	5 54	6 55	12 58	13 59	15 60	16 61	21	22	24	25	27	34	36	37	39	41	43	45

The problem of diagnostic classification or differential diagnosis is to assign each possible patient vector or point in the patient vector space to a proper pattern class or a disease category. As for the differential diagnosis of bone tumors, this can be interpreted as a partition of the vector space into mutually exclusive regions, each of which will correspond to a particular histological type of bone tumors. The boundary of partition, called the decision boundary, between regions in the vector space, can be expressed mathematically by various kinds of equations.

When only two regions, that is, only two categories of bone tumors are concerned, the decision on which category a patient is to be assigned to can be implemented by evaluating the sign of a single discriminant function $g(X)=g(x_1, x_2, \dots, x_n)$. If $g(X)$ is positive for a set of n features, that is, a vector $X=(x_1, x_2, \dots, x_n)$, the vector X or the patient is placed in one category; if $g(X)$ is negative, X is placed in another category. The equation $g(X)=0$ gives the decision boundary for separating the two categories. A two-dimensional illustration of discriminant function analysis or discriminant analysis is shown in Fig. 1 where only two features are measured from each patient.

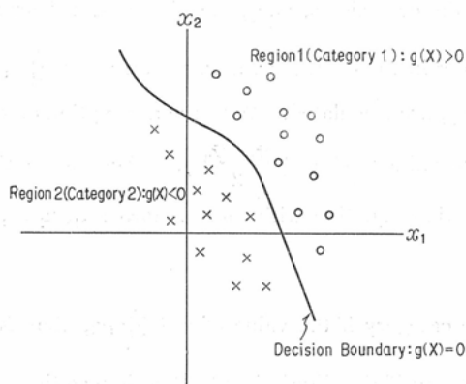


Fig. 1. An example of discriminant analysis in a two-dimensional space.

When a linear combination of the feature measurements x_1, x_2, \dots, x_n is selected for the discriminant function $g(X)$, this is called a linear discriminant function analysis or a linear discriminant analysis. The linear function $g(X)$ is expressed as follows.

$$g(X) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = \alpha_0 + \sum_{i=1}^n \alpha_i x_i$$

where α_i represents the weight or coefficient of x_i and α_0 is the constant.

A complete specification of any linear discriminant function is achieved by specifying the values of the weights or coefficients of the function family. The equation $g(X)=0$ is the equation of a hyperplane in the n -dimensional space E^n . This corresponds to a straight line in a two-dimensional space. So the decision boundary is a hyperplane in linear discriminant analysis.

Linear Discriminant Analyses

In this study two different methods were tried to specify the values of the weights or coefficients of linear discriminant functions. The first is the linear discriminant analysis assuming multivariate

normal distributions of patient vectors with identical covariance matrices. There are two major assumptions in this method. One is that random variables representing n radiological features shall be normally distributed, another is that the covariance matrices shall be identical in both categories of each pair. Under these assumptions, we can specify such values of the weights or coefficients that will minimize the probability of erroneous discrimination from the viewpoint of decision theory¹⁰⁾. In this method, the values of the coefficients, $\alpha_1, \alpha_2, \dots, \alpha_n$ are obtained from,

$$\alpha = W^{-1}d$$

where α is a $n \times 1$ column vector with elements α_i , d is a $n \times 1$ column vector with elements $d_i = (\bar{x}_{i1} - \bar{x}_{i2})$, where \bar{x}_{i1} and \bar{x}_{i2} are the means of x_i in Category 1 and 2 respectively, and W^{-1} is the inverse of the $n \times n$ covariance matrix with elements defined by

$$w_{ij} = \frac{1}{(N_1 + N_2 - 2)} \left[\sum_{c=1}^{N_1} (x_{i1c} - \bar{x}_{i1})(x_{j1c} - \bar{x}_{j1}) + \sum_{c=1}^{N_2} (x_{i2c} - \bar{x}_{i2})(x_{j2c} - \bar{x}_{j2}) \right]$$

for all $i, j = 1, \dots, n$, where N_1 and N_2 are the sample sizes of patients in Category 1 and 2 respectively, x_{i1c} and x_{j1c} are the values of x_i for the patients in Category 1, x_{i2c} and x_{j2c} are the values of x_i for the patients in Category 2. After the coefficients, $\alpha_1, \alpha_2, \dots, \alpha_n$, have been determined, the values of $\sum_{i=1}^n \alpha_i x_i$ are calculated for all sample cases in each category, then the means of $\sum_{i=1}^n \alpha_i x_i$, represented as \bar{y}_1 and \bar{y}_2 for Category 1 and 2 respectively, are calculated. With the assumption of identical covariance matrices, the value of the constant α_0 is determined to be $-(\frac{\bar{y}_1 + \bar{y}_2}{2})$. Then the decision boundary is a hyperplane which is expressed by the following equation with the unknowns x_1, x_2, \dots, x_n .

$$-(\frac{\bar{y}_1 + \bar{y}_2}{2}) + \sum_{i=1}^n \alpha_i x_i = 0$$

A patient is classified into one category if the value of $\alpha_0 + \sum_{i=1}^n \alpha_i x_i$, that is, the discriminant value is positive, and into another category if the discriminant value is negative, where $\alpha_0 = -(\frac{\bar{y}_1 + \bar{y}_2}{2})$.

The second is the linear discriminant analysis using estimating equations. The values of the weights or coefficients of linear discriminant functions are specified as follows. The estimated value y_e shall be expressed as a linear function, i.e.,

$$y_e = \beta + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$$

Suppose that $y_e = 1$, if a patient belongs to one category, and $y_e = -1$, if he belongs to another category. Then an equation with the unknown $\beta, \alpha_1, \alpha_2, \dots, \alpha_n$, is made for each patient vector. If Category 1 and 2 have N_1 and N_2 cases respectively, a set of $N_1 + N_2$ simultaneous equations is composed. The set of simultaneous equations is approximately solved by solving the following system of n simultaneous equations in the unknowns, $\beta, \alpha_1, \alpha_2, \dots, \alpha_n$.

$$\begin{aligned} \sum y_e &= \beta(N_1 + N_2) + \alpha_1 \sum x_1 + \alpha_2 \sum x_2 + \dots + \alpha_n \sum x_n \\ \sum x_1 y_e &= \beta \sum x_1 + \alpha_1 \sum x_1^2 + \alpha_2 \sum x_2 x_1 + \dots + \alpha_n \sum x_n x_1 \\ \sum x_2 y_e &= \beta \sum x_2 + \alpha_1 \sum x_1 x_2 + \alpha_2 \sum x_2^2 + \dots + \alpha_n \sum x_n x_2 \\ &\vdots \\ \sum x_n y_e &= \beta \sum x_n + \alpha_1 \sum x_1 x_n + \alpha_2 \sum x_2 x_n + \dots + \alpha_n \sum x_n^2 \end{aligned}$$

where the symbol \sum is used to denote the sum over the total of $N_1 + N_2$ cases. After the coefficients $\alpha_1, \alpha_2, \dots, \alpha_n$ and the constant β have been determined, the value of $\beta + \sum_{i=1}^n \alpha_i x_i$ are calculated for all sample cases in each category, then the means and the standard deviations of $\beta + \sum_{i=1}^n \alpha_i x_i$, expressed as \bar{y}_{e1}, σ_1 and \bar{y}_{e2}, σ_2 for Category 1 and 2 respectively, are calculated. The value of the constant α_0 of a linear discriminant function is determined to be $\beta - (\frac{\sigma_2 \bar{y}_{e1} + \sigma_1 \bar{y}_{e2}}{\sigma_1 + \sigma_2})$. Then the decision boundary is a hyperplane whose equation is

$$\beta - (\frac{\sigma_2 \bar{y}_{e1} + \sigma_1 \bar{y}_{e2}}{\sigma_1 + \sigma_2}) + \sum_{i=1}^n \alpha_i x_i = 0$$

with the unknowns x_1, x_2, \dots, x_n . A patient is classified into one category if the value of $\alpha_0 + \sum_{i=1}^n \alpha_i x_i$, that is, the discriminant value is positive, and into another category if the discriminant value is negative, where $\alpha_0 = \beta - (\frac{\sigma_2 \bar{y}_{e1} + \sigma_1 \bar{y}_{e2}}{\sigma_1 + \sigma_2})$.

Results of Analyses

In Table 5 the results of linear discriminant analyses when assuming multivariate normal distribu-

Table 5. Linear discriminant analysis assuming multivariate normal distributions. Discriminations between Fwing's sarcoma and fibrosarcoma in the total group.

Critical χ^2 value (% level)	Number of selected questions and (items)	Rate of correct discriminations	
		Ewing's sarcoma	fibrosarcoma
40	5 (5)	125/176 (71%)	134/150 (89%)
30	10 (7)	152/176 (86%)	124/149 (83%)
20	23 (16)	152/176 (86%)	134/149 (90%)
10.83 (0.1% level)	31 (19)	156/176 (89%)	135/149 (91%)
6.63 (1 % level)	34 (19)	156/176 (89%)	134/149 (90%)
3.84 (5 % level)	43 (20)	(Discriminant coefficients not obtainable)	

tions are shown along with the number of selected questions for discrimination between Ewing's sarcoma and fibrosarcoma. Indeed the rates of correct discriminations increase as the number of selected questions, that is, feature measurements increases, but the improvement is not remarkable compared with the great increase of questions in number. Though the number of questions increases about 6 times, the rate of correct discriminations shows only about 10 per cent improvement, and appears to reach the highest plateau about the critical chi-square value of 10.83 which is the 0.1% level of significance in the chi-square test. At the critical value of 3.84, the coefficients of the linear function can not be obtained because the covariance matrix is singular for this pair of categories. In Table 6 the results of discrimination between each pair of 4 categories of bone tumors, except for the pair of Ewing's sarcoma and fibrosarcoma, are shown at each of the critical values 10.83, 6.63 and 3.84. Within this range of

Table 6. Linear discriminant analysis assuming multivariate normal distributions. Discriminations between each pair of Ewing's sarcoma, fibrosarcoma, reticulum cell sarcoma and osteosarcoma, except for the pair of Ewing's sarcoma and fibrosarcoma.

Critical χ^2 value (% level)	Number of selected questions and (items)	Rate of correct discriminations	
Ewing's sarcoma and reticulum cell sarcoma			
		Ewing's sarcoma	reticulum cell sarcoma
10.83 (0.1% level)	4 (4)	159/178 (89%)	29/42 (69%)
6.63 (1 % level)	12 (7)	162/178 (91%)	30/42 (71%)
3.84 (5 % level)	22 (13)	162/176 (92%)	32/41 (78%)
Ewing's sarcoma and osteosarcoma			
		Ewing's sarcoma	osteosarcoma
10.83 (0.1% level)	17 (13)	153/177 (89%)	178/212 (84%)
6.63 (1 % level)	27 (16)	153/177 (89%)	185/212 (87%)
3.84 (5 % level)	33 (17)	162/177 (92%)	184/212 (87%)
Fibrosarcoma and reticulum cell sarcoma			
		fibrosarcoma	reticulum cell sarcoma
10.83 (0.1% level)	7 (5)	143/150 (99%)	25/42 (60%)
6.63 (1 % level)	14 (10)	147/150 (98%)	26/41 (63%)
3.84 (5 % level)	19 (13)	143/150 (99%)	27/41 (66%)
Fibrosarcoma and osteosarcoma			
		fibrosarcoma	osteosarcoma
10.83 (0.1% level)	23 (13)	129/149 (87%)	194/210 (92%)
6.63 (1 % level)	29 (13)	129/149 (87%)	194/210 (92%)
3.84 (5 % level)	31 (14)	129/149 (87%)	194/210 (92%)
Reticulum cell sarcoma and osteosarcoma			
		reticulum cell sarcoma	osteosarcoma
10.83 (0.1% level)	14 (10)	35/41 (85%)	203/210 (97%)
6.63 (1 % level)	18 (11)	35/41 (85%)	200/210 (95%)
3.84 (5 % level)	23 (14)	35/41 (85%)	201/210 (96%)

Table 7. Linear discriminant analysis using estimating equations. Discriminations between Ewing's sarcoma and fibrosarcoma in the total group.

Critical χ^2 value (% level)	Number of selected questions and (items)	Rate of correct discriminations	
		Ewing's sarcoma	fibrosarcoma
40	5 (5)	160/176 (91%)	105/150 (70%)
30	10 (7)	152/176 (86%)	124/149 (83%)
20	23 (16)	153/176 (87%)	130/149 (87%)
10.83 (0.1% level)	31 (19)	157/176 (89%)	134/149 (90%)
6.63 (1% level)	34 (19)	159/176 (90%)	132/149 (89%)
3.84 (5% level)	43 (20)	(Discriminant coefficients not obtainable)	

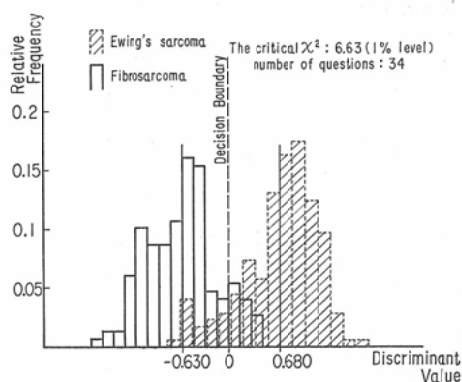


Fig. 2. Discriminant analysis using estimating equations. Relative frequency distributions of discriminant values, for the pair of Ewing's sarcoma and fibrosarcoma.

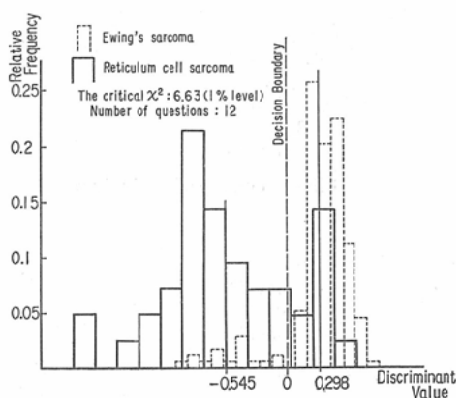


Fig. 3. Discriminant analysis using estimating equations. Relative frequency distributions of discriminant values, for the pair of Ewing's sarcoma and reticulum cell sarcoma.

the critical levels, the rate of correct discriminations shows no marked improvement with the increase of questions in number. All pairs including reticulum cell sarcoma as a partner show that their decision boundaries are relatively biased toward the prototype pattern of reticulum cell sarcomas.

The results of linear discriminant analyses using estimating equations are shown in Table 7 and 8 along with the number of selected questions for each pair of categories. All pairs of Ewing's sarcoma, fibrosarcoma and osteosarcoma show the results which are almost the same as those of analyses with the assumption of normal distributions. The pairs including reticulum cell sarcoma show the results which are more favorable than those of the former method of analysis, but the positional control of decision boundaries appears yet unsatisfactory. As for the latter method of analysis, some examples are shown as graphed relative frequency distributions of the discriminant values in Fig. 2, 3 and 4 which illustrate the situations described above.

Table 8. Linear discriminant analysis using estimating equations. Discriminations between each pair of Ewing's sarcoma, fibrosarcoma, reticulum cell sarcoma and osteosarcoma, except for the pair of Ewing's sarcoma and fibrosarcoma.

Critical χ^2 value (% level)	Number of selected questions and (items)	Rate of correct discriminations	
Ewing's sarcoma and reticulum cell sarcoma			
		Ewing's sarcoma	reticulum cell sarcoma
10.83 (0.1% level)	4 (4)	158/178 (89%)	30/42 (71%)
6.63 (1 % level)	12 (7)	160/178 (90%)	33/42 (79%)
3.84 (5 % level)	22 (13)	159/176 (90%)	34/41 (83%)
Ewing's sarcoma and osteosarcoma			
		Ewing's sarcoma	osteosarcoma
10.83 (0.1% level)	17 (13)	155/177 (87%)	180/212 (85%)
6.63 (1 % level)	27 (16)	157/177 (89%)	185/212 (87%)
3.84 (5 % level)	33 (17)	160/177 (90%)	187/212 (88%)
Fibrosarcoma and reticulum cell sarcoma			
		fibrosarcoma	reticulum cell sarcoma
10.83 (0.1% level)	7 (5)	135/150 (90%)	33/42 (79%)
6.63 (1 % level)	14 (10)	136/150 (91%)	32/41 (78%)
3.84 (5 % level)	19 (13)	139/150 (93%)	33/41 (80%)
Fibrosarcoma and osteosarcoma			
		fibrosarcoma	osteosarcoma
10.83 (0.1% level)	28 (13)	129/149 (87%)	194/210 (92%)
6.63 (1 % level)	29 (13)	129/149 (87%)	194/210 (92%)
3.84 (5 % level)	31 (14)	129/149 (87%)	194/210 (92%)
Reticulum cell sarcoma and osteosarcoma			
		reticulum cell sarcoma	osteosarcoma
10.83 (0.1% level)	14 (10)	36/41 (88%)	192/210 (91%)
6.63 (1 % level)	18 (11)	36/41 (88%)	194/210 (92%)
3.84 (5 % level)	28 (14)	35/41 (85%)	195/210 (93%)

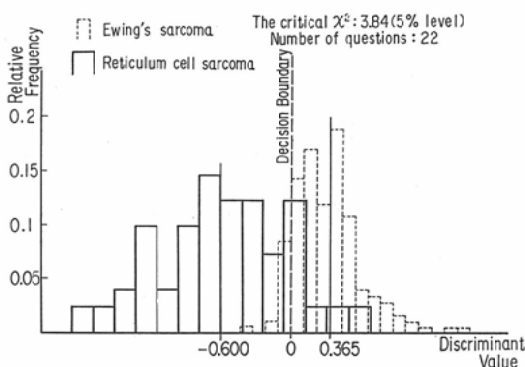


Fig. 4.. Discriminant analysis using estimating equations. Relative frequency distributions of discriminant values, for the pair of Ewing's sarcoma and reticulum cell sarcoma.

Differential Diagnosis

It can be said that discrimination between each pair of all considerable disease categories is the base of diagnostic classification or differential diagnosis. One method of classifying patients into R disease categories is to assign one category to a patient if the number of linear discriminant analyses which favor the category is maximum, after all pairs of R categories have been examined for the patient. When R categories are concerned, the total number of pairs is $\frac{R(R-1)}{2}$, and a system for differentiation consists of $\frac{R(R-1)}{2}$ linear discriminant functions. So when 4 categories are considered as possible, there are 6 pairs each of which can be divided into two categories by each specified linear discriminant function. In this method, however, it is possible that more than one category have the same maximum number of linear discriminant functions which favor each of them. In such cases, it is impossible to select only one category as a correct diagnosis.

First, as shown in the section of material, 576 cases were classified based on the results of linear dis-

Table 9 Differential diagnosis in the total group, using the questions selected at the critical chi-square value of 6.63.

		Computer's Diagnoses					Sum
		Ewing's sarcoma	fibro-sarcoma	reticulum cell sarcoma	osteosarcoma	undetermined	
Pathological Diagnoses	Ewing's sarcoma	135 (77%)	16	6	15	4	176 (100%)
	fibrosarcoma	6	119 (80%)	5	11	8	149 (100%)
	reticulum cell sarcoma	8	7	23 (56%)	2	1	41 (100%)
	osteosarcoma	20	12	2	168 (80%)	8	210 (100%)
Sum		169	154	36	196	21	576

Table 10. Feature selection in the training group of 401 cases. Questions were selected at the critical chi-square value of 6.63.

Pairs of categories	Selected questions															
Ewing's sarcoma -fibrosarcoma	1 39	2 41	5 43	6 45	11 46	15 47	19 48	21 51	22 52	27 53	30 55	32 57	34 58	36 59	37 60	38 62
Ewing's sarcoma -reticulum cell sarcoma	1	6	21	43	53	58	59	60	62							
Ewing's sarcoma -osteosarcoma	1 47	3 48	11 61	15 62	16	19	22	24	25	32	34	39	41	42	43	45
fibrosarcoma -reticulum cell sarcoma	12	34	39	41	47											
fibrosarcoma -osteosarcoma	2 43	3 45	5 46	6 51	19 52	24 53	25 54	26 55	27 57	30 58	34 59	36 60	37 61	38 62	39	41
reticulum cell sarcoma -osteosarcoma	3	5	6	12	13	22	24	25	37	43	45	47	53	59	60	61

criminant analyses using estimating equations where the feature selection was made at the critical chi-square value of 6.63, that is, the 1 per cent level of significance in the chi-square test. The results are shown in Table 9. On the average, 77 per cent of pathological diagnoses coincide with computer's diagnoses.

Second, as shown before, the total cases were divided into the training group of 401 cases and the testing group of 198 cases. After feature selection and linear discriminant analyses using estimating equations had been done at the critical chi-square value of 6.63 for the training group in the same way as described before, differential diagnoses were tried for both 387 cases of the training group and 189 cases of the testing group, as explained before. The results of feature selection and discriminant analyses are shown in Table 10 and 11. In Table 12 and 13 the results of differential diagnoses are shown for each group. As for the training group, the numbers of selected questions disagree with those selected in the total group, and the system of linear discriminant functions is different from that of the total group. However, differential diagnoses show almost the same results in the training group as in the total group. On the average, 78 per cent of pathological diagnoses coincide with computer's diagnoses in the training group.

In the testing group, the results of differential diagnoses are, on the average, about 10 per cent worse than those in the training group. On the average, 70 per cent of pathological diagnoses coincide with computer's diagnoses in the testing group.

Discussion

The problem of feature selection or measurement selection in diagnostic radiology consists in knowing what features or measurements should be taken from the input radiographic images for correct diagnosis. Selected features are supposed to be invariant with respect to the commonly encountered variations and also to contain less redundancies. Unfortunately, at present there is no established rule for the selection of features, even though some have studied this problem⁴⁾¹¹⁾¹⁶⁾. Usually, the decision on what to measure or detect is rather subjective and is often guided only by personal intuitive ideas.

Table 11. Linear discriminant analyses using estimating equations in the training group.

Critical χ^2 value (% level)	Number of selected questions and (items)	Rate of correct discriminations	
Ewing's sarcoma and fibrosarcoma			
		Ewing's sarcoma	fibrosarcoma
6.63 (1 % level)	32 (19)	105/117 (90%)	93/99 (94%)
Ewing's sarcoma and reticulum cell sarcoma			
		Ewing's sarcoma	reticulum cell sarcoma
6.63 (1 % level)	9 (6)	108/118 (91%)	21/29 (72%)
Ewing's sarcoma and osteosarcoma			
		Ewing's sarcoma	osteosarcoma
6.63 (1 % level)	20 (14)	106/118 (90%)	125/144 (87%)
Fibrosarcoma and reticulum cell sarcoma			
		fibrosarcoma	reticulum cell sarcoma
6.63 (1 % level)	5 (4)	89/100 (89%)	21/29 (72%)
Fibrosarcoma and osteosarcoma			
		fibrosarcoma	osteosarcoma
6.63 (1 % level)	30 (14)	89/99 (90%)	137/143 (96%)
Reticulum cell sarcoma and osteosarcoma			
		reticulum cell sarcoma	osteosarcoma
6.63 (1 % level)	16 (12)	25/28 (89%)	134/143 (94%)

Table 12. Differential diagnosis in the training group, based on the results of linear discriminant analyses shown in Table 11.

		Computer's Diagnoses					Sum
		Ewing's sarcoma	fibro- sarcoma	reticulum cell sarc.	osteo- sarcoma	undeter- mined	
Pathological Diagnoses	Ewing's sarcoma	92 (79%)	8	3	9	5	117 (100%)
	fibrosarcoma	4	81 (82%)	2	6	6	99 (100%)
	reticulum cell sarc.	8	7	12 (43%)	1	0	28 (100%)
	osteosarcoma	12	6	3	117 (82%)	5	143 (100%)
Sum		116	102	20	133	16	387

Table 13. Differential diagnosis in the testing group, based on the results of linear discriminant analyses in the training group.

		Computer's Diagnoses					Sum
		Ewing's sarcoma	fibrosarcoma	reticulum cell sarc.	osteosarcoma	undetermined	
Pathological Diagnoses	Ewing's sarcoma	36 (61%)	7	3	10	3	59 (100%)
	fibrosarcoma	1	36 (72%)	2	7	4	50 (100%)
	reticulum cell sarc.	3	2	4 (31%)	2	2	13 (100%)
	osteosarcoma	3	5	1	56 (83%)	2	67 (100%)
Sum		43	50	10	75	11	189

Generally speaking, this should depend upon the established statistics which are supported by other studies, such as chemical, histological, microbiological and so on.

Differential diagnosis can be considered as based on discrimination between each pair of all possible categories for a given set of features, that is, symptoms and findings. So the problem of feature selection is reduced to knowing how we can quantitatively express the importance, in other words, the discriminating power of a radiological feature for separating two categories. Now then, it will be helpful to review how physicians are used to determining the importance of a clinical finding for differential diagnosis. When we evaluate the importance of a finding for differentiation between two disease categories, we almost always compare the relative frequencies or the occurrence rates of the finding for patients in these two categories. The relative frequency is regarded as an estimate for the conditional probability of the finding with respect to each disease category. Actually we do not always rely upon the probability of each disease with respect to the finding. In fact, it is very hard to estimate it. It seems quite certain that the greater the difference of the conditional probabilities of a same finding between two different diseases is, the more important and powerful the finding is for discriminating the two diseases. It looks quite natural to express the discriminating power of a finding as the difference of the conditional probabilities whose estimates are the relative frequencies, because most of the time we are using this way of decision.

A chi-square test with a 2×2 contingency table is equivalent to a significance test of differences in proportions using the normal approximation. As the difference between two proportions or rates becomes greater, the chi-square value becomes larger. So the larger chi-square value can be considered as to indicate the more important finding for discriminating two disease categories.

In this study several tentative critical levels were tried to select features according to the chi-square values. Generally it can be said that the greater the number of selected features, which have large chi-square values, is for a pair of disease categories, the easier the discrimination between these two categories is. In this study feature measurements were selected from 63 questions about 25 items of radiological findings plus the age of patients. The numbers of items and questions to be considered may be increased and the members of selected features may be changed. In fact, this study shows that the results of feature selection in the total group of 599 cases are fairly different from those in the training group of 401 cases.

The problem of feature selection is associated not only with the present but also with the future. At present we cannot but select features from what we can consider, but new facts and methods are incessantly appearing. We must always seek the most powerful set of features.

Moreover, the problem of feature selection is closely related to the decision-making scheme used. In this method of feature selection, it can not be examined if there are mutually dependent features or redundant features in the set of selected features, and it is rather evident that there are fairly many items and questions which do not seem mutually independent at all. But we have no established method for settling this problem as yet, even though some works have been reported¹³⁾¹⁴⁾. However, how much the existence of mutual dependence of selected features influences the results of a discriminant analysis depends upon what is taken in this analysis as a probabilistic model for the distribution of patient vectors. As for the former method of discriminant analysis assuming normal distributions, it is favorable that normal distributions encompass some situations in which the feature measurements or variables are not statistically independent. Furthermore, in the latter method of discriminant analysis, where there is no supposed probabilistic model, mutual dependence of feature measurement is of no concern.

Then what should be taken as the critical chi-square value for feature selection? As for the pair of Ewing's sarcoma and fibrosarcoma, the rate of correct discriminations appears to reach the highest plateau about the critical value of 10.83 in both methods of linear discriminant analysis as shown in Table 5 and 7, and however smaller the critical value is, that is, however more questions are picked up, it increases only redundant information. As for other pairs also, the rates of correct discriminations were improved, at most, only 3% in all pairs except for the pair of Ewing's sarcoma and reticulum cell sarcoma, as the critical value decreases from 6.63 to 3.84 (Table 6 and 8). The chi-square value of 3.84 corresponds to the differences between, for example, such couples of relative frequencies as 0.92 to 0.85, 0.19 to 0.12, and so forth. Considering these figures, more improvement can not be expected from making the critical value smaller. Especially as for the pair of fibrosarcoma and reticulum cell sarcoma, the number of questions selected in the training group at the critical value of 6.63 is only 5 compared with 14 questions selected in the total group as shown in Table 4 and 10. But it is interesting to find that the results of discriminant analyses using estimating equations show insignificant differences, as shown in Table 8 and 11. This fact suggests the existence of redundancies.

In this study the scale of radiological features or measurements is a nominal or classificatory scale, and the variables representing features take only the value of either 1 or 0 for each patient. Clearly, this does not satisfy the first assumption of the former method of analysis assuming normal distributions. The second assumption of identical covariance matrices also seems hard to be generally admitted. So it means only approximation to take a multivariate normal distribution as a probabilistic model in this study. The latter method of analysis using estimating equations is not based on these questionable assumptions. This study shows that the rates of correct discriminations between all pairs of three bone tumors except for reticulum cell sarcoma, which has a small number of cases compared with others, are almost the same between these two different methods of analysis. So it appears that the assumptions in the former method of analysis might be nearly satisfied for Ewing's sarcoma, fibrosarcoma and osteosarcoma. But the patient vectors of reticulum cell sarcoma are distributed with a greater within-dispersion than those of other categories. When examining the standard deviations of discriminant values

in the latter method of analysis for all pairs of bone tumor categories, it is found that the differences of the standard deviations between each pair of Ewing's sarcoma, fibrosarcoma and osteosarcoma are at most 20 per cent, while as for the pairs including reticulum cell sarcoma, most of the standard deviations of reticulum cell sarcoma are more than twice those of its partners. This suggests that the assumption of identical covariance matrices in the former method of analysis cannot be admitted for the pairs including reticulum cell sarcoma. Concerning the difference of covariance matrices, the latter method of analysis is clearly better than the former because the decision boundary is determined by using the standard deviations.

Diagnostic classification or differential diagnosis was tried using a system of 6 linear discriminant functions for 4 bone tumors as shown before with the results. This method has a disadvantage that there may be cases whose diagnoses can not be determined in the event that more than one categories have the same maximum number of discriminant functions which favor each of them. The diagnoses of 4 to 5 per cent of cases were undetermined through the differential trials in this study. Reticulum cell sarcoma whose cases are distributed with a greater within-dispersion shows the worse results of computer differentiation. This corresponds to the difficulty of this bone tumor in clinical differential diagnosis.

This study shows that the testing group was differentiated with the results which are about 10 per cent worse on the average than the training group, nevertheless the training and the total groups have almost the same results of differentiation. It will be certain that a system for differentiation derived from a training group can be more improved and lessen the difference of the results existing between the training group and the testing group, as the training group becomes larger. It will be interesting to observe how the gap will become smaller in the future.

So far several kinds of decision-making schemes have been proposed and tried in application of computers in medical diagnosis, which are based on Boolean algebra³⁾⁴⁾, Bayes' rule⁵⁾⁶⁾, maximum-likelihood decision¹²⁾, discriminant functions, some combinations of these methods⁷⁾ and so on. Indeed they are triable in clinical medicine, but there is no established general method as yet. A lot of new theories and methods will probably be proposed in the future. The only way to find the best decision-making scheme for computer diagnosis is to try what appears promising, using the largest possible amount of reliable data.

Summary

Using the 2×2 contingency tables and the chi-square test, a method of feature selection was tried in the roentgenologic examination of four categories of primary malignant bone tumors, which are Ewing's sarcoma, fibrosarcoma, reticulum cell sarcoma and osteosarcoma, whose coded data of 599 histologically proven cases had been punched into the IBM cards. Based on the results of feature selection, two different methods of linear discriminant analysis were tried and compared for each pair of categories of bone tumors, one is derived from assuming multivariate normal distributions and another is from using estimating equations. The results of discriminant analyses were examined and discussed along with the results of feature selection.

A system of linear discriminant functions for differential diagnosis was composed, based on the results of feature selection and linear discriminant analyses in the training group of 401 cases, and used

to differentiate the testing group of 198 cases from which 9 cases (4.5%) were excluded because of insufficient coding. On the average, 70 per cent of pathological diagnoses coincided with computer's diagnoses in the testing group, while the average rate of coincidence was 78 per cent in the training group.

Acknowledgment: The authors express their gratitude to Dr. Samuel J. Dwyer III, Professor, and Mr. Joseph K. Bryan, Graduate Student, Department of Electrical Engineering, University of Missouri, for their helpful suggestions and observations during the course of this work.

References

- 1) Fu, K.S.: Sequential Methods in Pattern Recognition and Machine Learning. New York, Academic Press, 1968.
 - 2) Watanabe, S.: Recognition and Information, Tokyo, Japan Broadcasting Association (NHK) Press, 1968.
 - 3) Ledley, R.S., and Lusted, L.B.: Reasoning Foundations of Medical Diagnosis; Symbolic Logic, Probability, and Value Theory Aid Our Understanding of How Physicians Reason. Science, 130: 9-21, July 3, 1959.
 - 4) Lodwick, G.S.: Solitary Malignant Tumors of Bone: The Application of Predictor Variables in Diagnosis. Seminars in Roentgenology, 1: 293-313, July, 1966.
 - 5) Lodwick, G.S., Haun, C.L., Smith, W.E., Keller, R.F., and Robertson, E.D.: Computer Diagnosis of Primary Bone Tumors: A Preliminary Report. Radiology, 80: 273-275, February, 1963.
 - 6) Lodwick, G.S., Keats, T.E., and Dorst, J.P.: The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. Radiology, 81: 185-200, August, 1963.
 - 7) Lodwick, G.S., and Reichertz, P.L.: Computerunterstützte Diagnostik von Tumoren und tumorähnlichen Veränderungen des Knochens. Das begrenzte Bayes-Konzept. Röntgen-Blätter, 22: 162-168, April, 1969.
 - 8) Lodwick, G.S., Turner, A.H., Jr., Lusted, L.B., and Templeton, A.W.: Computer-Aided Analysis of Radiographic Images. J. Chron. Dis., 19: 485-496, 1966.
 - 9) Miyawaki, K.: Medical and Biological Information Processing, 1st ed., Tokyo, Corona Co., 1966, p. 180-237.
 - 10) Nilsson, Nils J.: Learning Machines, New York, McGraw-Hill, 1965.
 - 11) Pipberger, H.V., Klingeman, J.D., and Cosma, J.: Computer Evaluation of Statistical Properties of Clinical Information in the Differential Diagnosis of Chest Pains. Meth. Inform. Med., 7: 79-92, April, 1968.
 - 12) Takahashi, K., Miyata, M., Miyahara, H., and Dohmae, A.: Localization of Cerebellar Tumors. Clinic All-Round 17: 19-25, January, 1968.
 - 13) Templeton, A.W., Jansen, C., Lehr, J.L., and Hufft, R.: Solitary Pulmonary Lesions: Computer-Aided Differential Diagnosis and Evaluation of Mathematical Methods. Radiology 89: 605-614, October, 1967.
 - 14) Templeton, A.W., Simmons, C., and Lehr, J.L.: Computer Diagnosis of Heart Disease: The Public Model. The Am. J. Roentgenol., 102: 865-874, April, 1968.
 - 15) Torii, T., Takahashi, K., and Doi, I.: Inferential Statistics for Medicine and Biology, 1st ed., Tokyo, Tokyo Univ. Press, 1954, p. 73-88.
 - 16) Wilson, W.J., Templeton, A.W., Turner, A.H., Jr., and Lodwick, G.S.: The Computer Analysis and Diagnosis of Gastric Ulcers. Radiology, 85: 1064-1073, December, 1965.
-