Scripta Scientifica Medica, vol. 29 (1996), pp.63-70

DETERMINATION OF DYSLIPIDEMIA BY DISCRETE AND PSEUDO-DISCRETE FEATURES

K. Tenekedjiev, A. Protopopova*, J. Mihova

*Department of Technique and Technology for Water and Air Protection, Technical University of Varna, and *Department of Endocrinology, Gastroenterology and Metabolic Diseases, Medical University of Varna, Varna*

The necessity of controlling the lipid state of the organism originates from the association of dyslipidemia with atherosclerosis, heart diseases, diabetes mellitus, etc. In the clinical practice, the determination of the exact type of dyslipidemia is not straightforward due to certain problems: daily biological fluctuations, dietary effects, strong overlapping in characteristic plasma lipid levels in some classes of type IIb, III and IV, frequent inavailability of a complete set of diagnostic measurements, as well as certain overlapping in the borderline values. These provide the objective to create a computer program for diagnostics of dyslipidemia in 16 classes. The apparatus of Bayes classification has been applied using 13 discrete and 7 pseudo-discrete features. The proposed system is suitable for monitoring the treatment of dyslipidemia as well as for the purposes of students' education and post doctoral training.

Key-words: Dyslipidemia, diagnostics, Bayes classification, discrete features, pseudo-discrete features

Dyslipoproteinemia represents a major risk factor for atherosclerosis . The specific factors increasing the risk for atherosclerosis include, elevated serum levels of low-density lipoproteins (LDL) carrying cholesterol from the liver to the periferal tissues and low serum levels of high-density lipoproteins (HDL) evacuating the cholesterol from the extrahepatic tissues (1). Hypertriglyceridemia is often regarded as an independent risk-factor for atherosclerosis especially in diabetes mellitus when the lipoprotein catabolism is impaired (9). Reduction of LDL-cholesterol (LDL_{ch}) by 12,6 % is accompanied by a 2% increase in HDL-cholesterol (HDL_{ch}), whereas a 25 % reduction in the total cholesterol (TC) results in a 50 % decrease in the vasular risk (7). Following 5 years of adequate diet and drug therapy the risk for cardiovascular diseases decreases by 34 % in patients with primary hyperlipoproteinemia and non-HDL_{ch} less than 5,2 mM (6).

Address for correspondence:
A. Protopopova, Dept. of Endocrinology, Gastroenterology and Metabolic Diseases, Medical University, 55 Marin Drinov St, BG-9002 Varna, BULGARIA

Modulation of the plasma lipoprotein levels by Colestipol® and Ac. nicotinici for about 2 years reduces LDL_{ch} by 43 % and increases HDL_{ch} by 37 % in patients with coronary arterial bypass. These changes are accompanied by a significant diminution in the number of angiographically proved new lesions in the arteries and by improvement of the whole coronary state (6,7). HDL affects some other aspects of atherogenesis related to cholesterol deposition (e.g.suppression of thrombocyte aggregation, stimulation of prostacyclin production in the arteries) (6,9). In screened males aged between 35 and 70 years the relative mortality risk from myocardial infarction at a TC of 6,5 mM (85^{th} percentile) is 3,8 - fold higher than at TC of 4,7 mM (20^{th} percentile). According to the Framingham Group Study, plasma cholesterol level is directly related to the mortality ratio in individuals younger than 50 years and particularly at the age of about 30 years (11). Hyperlipidemia is established in approximately 40-50 % of the diabetics and probably their increased morbidity and mortality rates of coronary disease are related to it (5,10,12-14,17). Triglyceride (TG) reduction to normal values results in a 50% lower coronary risk (1,6 instead 3,0) (10).

Hypertriglyceridemia represents the most essential anomaly of the lipid metabolism in diabetes mellitus (6,7,15,17). Hyperlipidemia that is secondarily determined by some other metabolic diseases can persist for a long time after the cessation of the effective treatment of the primary disease. All these data indicate the necessity of a regular monitoring of the lipid state in the general population with the objective of early diagnosis of dyslipidemia and its appropriate treatment . Plasma lipid-level shows daily biological fluctuations and affected by the diet. That is why the diagnosis of dyslipidemia should be based on at least two complete lipid analyses in the course of 2 weeks (6). In order to facilitate the evaluation of these disturbances, several classifications have been created (e.g., Frederickson, Schlierf, Kahlke, European Atherosclerosis Society, a Therapeutic Classification of Hyperlipidemiae, etc.) (1,2,4,6).

Despite the merit of these classifications in the clinical practice often patients with lipid disturbances could hardly be classified into a definite class of dyslipidemia due to the overlapping of the lipid value-ranges for some classes (e.g., type IIb, III, and IV). That is the case with asymptomatic patients with subclinical atherosclerosis, ischemic heart disease, liver steatosis, diabetes mellitus type II, initial nephropathy (9,10). On the other hand, a few

health institutions can afford to examine the lipid state as a whole.

The purpose of the present paper is to elaborate a specialised algorithm for computer-aided diagnostics.

MATERIAL AND METHODS

Building of a classifier using discrete and pseudo-discrete features

Lipid profile parameters are continuous numbers with diagnostic information concentrated in different intervals of the numerical axis. They are particularly appropriate for representation with the so-called pseudo-discrete features (16). In addition to the lipid profile, information about the dyslipidemia type is found in external symptoms that are typical discrete features. Therefore, the diagnostics of dyslipidemia can be formulated as a problem for classification using discrete and pseudo-discrete features.

The following parameters are introduced: w_k (k=1,2,...,c), classes of state with *a priory* probabilities $P(w_k)$; the random vector of observation $\vec{X}$ consisting of a vector $\vec{X}^{(d)}$ with r discrete features and a vector $\vec{X}^{(p)}$ with s pseudo-discrete ones. The i-co-ordinate x_i^d of $\vec{X}^{(d)}$ adopts one of a number of b_i discrete values (x_i^d=1,2, ... j..., b_i) and its value is 0 for unknown x_i^d. The p_{ij}^k denotes the probability x_i^d to occupy j-discrete value, if the observation $\vec{x}$ belongs to class w_k ($\vec{x} \in w_k$).

Let i-co-ordinate x_i^p of $\vec{x}^{(p)}$ be a continuous random value that has typical ranges $[D_{j,i};U_{j,i}]$ and the limits of these ranges are as follows:

$$D_{1,i} \le U_{1,i} < D_{2,i} \le U_{2,i} < \ldots < D_{n_i,i} \le U_{n_i,i}$$

Then, either x_i^p belongs to the interval $[D_{1,i}; U_{n_i,i}]$, or it is NaN meaning an unknown value. The $q_{i,j}^k$ denotes the probability x_i^p to occupy j-typical range if $\vec{X} \in w_k$ and there no untypical range exists. Then $\sum_{j=1}^{n_i} q_{i,j}^k = 1$.

The latter condition enables the expert easily to determine these probabilities disregarding any atypical cases. The following subsiduary functions are introduced in order to determine the belonging of x_i^p to j-typical range:

$$\varphi_i^j(x_i^p) = \begin{cases} 0 & \text{for } x_i^p \le U_{j-1,i} \\ (x_i^p - U_{j-1,i}) / (D_{ji} - U_{j-1,i}) & \text{for } U_{j-1,i} < x_i^p < D_{ji} \\ 1 & \text{for } D_{ji} \le x_i^p \le U_{ji} \\ (D_{j+1,i} - x_i^p) / (D_{j+1,i} - U_{ji}) & \text{for } U_{ji} < x_i^p < D_{j+1,i} \\ 0 & \text{for } D_{j+1,i} \le x_i^p \end{cases}$$

The assumption for an independence of the features (8)

Table 1

Names and limits for the typical ranges of the pseudo-discretes features

Feature	Range	Limits (mM)
TG	normal	0.5-1.7
	increased	2.2-5
	extremely increased	10-70
TC	normal	2-5.2
	increased	8-10
	extremely increased	15-70
HDL_{ch}	decreased	0.1-0.7
	normal	0.9-1.3
	increased	1.6-3
LDL_{ch}	normal	2.5-4.9
	increased	5.5-7
	extremely increased	8-20
$VLDL_{ch}$	normal	0.05-0.3
	increased	0.5-5
ApoA	decreased	0-250
	normal	290-775
	increased	800-2000
ApoB	normal	30-150
	increased	180-800

(required to avoid "the curse of dimensionality") results in a Bayes-classifier (16):

$$A_k(\vec{x})=\ln P(w_k) + \sum_{i=1}^{r} \ln p^k_{i,x_i^d} + \sum_{i=1}^{s} \ln \sum_{j=1}^{b_i} q^k_{i,j} \varphi^i_j\left(x_i^p\right),$$

where $\varphi^i_j(NdN) = 1$ and $p_{i,0}=1$

From the so-called discriminators A_k (for k=1,2,...c) the desired *a posterior* probabilities $P(w_k / \vec{x})$ are obtained (8):

$$P(w_k / \vec{x}) = 1 / \sum_{j=1}^{c} e^{[A_j(\vec{x})-A_k(\vec{x})]}$$

where "e ≈ 2.7183" is the base of the natural logarithms.

A key component of the algorithm is the parameter estimation

Table 2

Conditional probabilities for the presence of the discrete symptoms $p^k_{i,1}(p^k_{i,2} = 1 - p^k_{i,1})$

Feature Class	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.1	0.7	0.01	0	0	0.01	0.5	0.2	0.2	0	0.2	0.01	0.3
2	0.3	0.5	0	0.15	0.1	0.4	0.3	0.1	0.05	0.2	0.15	0.05	0.15
3	0.2	0.7	0.1	0.1	0.1	0.5	0.5	0.4	0.4	0.2	0.2	0.03	0.1
4	0.7	0.9	0	0.7	0.5	0.9	0.6	0.8	0.9	0.05	0.2	0.05	0.05
5	0.2	0.8	0.2	0.2	0.15	0.8	0.8	0.6	0.5	0.2	0.3	0.02	0.05
6	0.9	0.9	0.6	0.6	0.1	0.5	0.8	0.8	0.7	0.6	0.1	0.3	0.05
7	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
8	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
9	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
10	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
11	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
12	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
13	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
14	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
15	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0
16	0.05	0.3	0.02	0.05	0.01	0.1	0.2	0.01	0.01	0.05	0.01	0.005	0

Table 3
Conditional probabilities of the pseudo-discrete features

Feature	1			2			3			4			5		6			7	
Range / Class	1	2	3	1	2	3	1	2	3	1	2	3	1	2	1	2	3	1	2
1	0	0	1	07	0.3	0	1	0	0	0	1	0	0	1	0.5	0.5	0	0	1
2	0.6	0.4	0	0	1	0	0.2	0.8	0	0	0.7	0.3	0.5	0.5	0.1	0.9	0	0	1
3	0	1	0	0.2	0.4	0.4	0.9	0.1	0	0.3	0.3	0.4	0.1	0.9	0.7	0.3	0	0.3	0.7
4	0	1	0	0.5	0.5	0	0.6	0.4	0	0.4	0.6	0	0	1	0.9	0.1	0	0.5	0.5
5	0	1	0	0.5	0.5	0	0.5	0.5	0	0.7	0.3	0	0	1	0.75	0.25	0	0.5	0.5
6	0	0	1	0	1	0	1	0	0	0.4	0.5	0.1	0.3	0.7	0	1	0	0	1
7	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0
8	1	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0	1	0
9	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0
10	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0
11	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1	1	0
12	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	1
13	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0
14	1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0
15	1	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	1	0
16	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0

of the conditional probabilities $p_{i,j}^{k}$ and $q_{i,j}^{k}$. Three different approaches for their estimation are described in (16) and (3): a) by an expert and according to literature data; b) from a learning sample, consisting of examples for a given class (called learning observations), and c) by pseudo-Bayes estimates combining both of the aforementioned approaches. In any case, the conditional probabilities become more and more adequate to the problem under consideration with information accumulation in the course of exploitation of the diagnostic system. The process described is called learning. The relative insensitivity to the input inaccuracies is considered to be a typical feature of the algorithm described.

RESULTS AND DISCUSSION

Diagnostics of dyslipidemia

Based on the above presented algorithm a diagnostic system classifying in the following classes has been elaborated:

1. type I;
2. type IIa;
3. type IIb;
4. type III;
5. type IV;
6. type V;
7. normal lipid-state;
8. isolated hyper-LDL-emia;
9. isolated hyper-VLDL-emia;

10.isolated hypoapoprotein-A-emia;

11. isolated hyperapoprotein-A-emia;

12.isolated hyperapoprotein-B-emia;

13. isolated hypertryglycerid-emia;

14.isolated hypercholesterol-emia;

15. isolated hyper-HDL-emia;

16. isolated hypo-HDL-emia.

The computer system makes use of 13 discrete features :

1. hereditary predisposition; 2. obesity; 3. xanthomas; 4. Xanthelasms; 5."cholesterol ring"; 6. sclerotic manifestations; 7. diabetes mellitus; 8.pancreatitis; 9. liver disease; 10. renal failure; 11. gout; 12. nephritis; 13. myxoedema, as well as of 7 pseudo-discrete ones:

1. triglycerides (TG); 2. total cholesterol (TC); 3. HDL; 4. LDL; 5. very law density lipoproteins (VLDL); 6.apoprotein A (ApoA); 7. apoprotein B (ApoB).

All the discrete features have two discrete values (r=2, i. e., there is/there is not). The pseudo-discrete features are mainly ternary (s=3), however, the fifth and the seventh ones are binary (s=2). All these features can also be unknown during recognition. The names and the limits of the typical ranges of the pseudo-discrete features in mM units are presented in Table 1.The conditional probabilities for discrete features are given in Table 2 (for the discrete value "there is" only), but those for the pseudo-discrete ones are shown in Table 3. They have been determined according to literature data available and by expert's interviewing.

The software was written on MATLAB for Windows 4.2. c.1. Simulink and works on PC 486 / 66 MHz / 8 MB. A possibility exists for adaptation of the system to newly acquired data and for correction of the diagnostic value of the features, in the process of recognition and the diagnostic value of learning observations in the course of learning of the procedure.

The experiments carried out with 67 preliminary diagnosed patients (containing mainly pseudo-discrete features) show that the system detects the true class in 47 cases (70 %) with the greatest *a posterior* probability (the mean value of the *a posterior* probability in case of correct classification being 73 %). The second most probable classes have a mean *a posterior* probability of 27 % that can serve as an indicator for the quality of the classification. In cases when the true class is not the most probable one, the mean value of the *a posterior* probability is 34 % while the classes considered most probable have an average probability of 49 %. This indicates that the true

classification occurs rather often and it is performed with a relatively great reliability. On the other hand, even in the case of error, serious indications exist which prevent the ignoring the correct class.

CONCLUSIONS

The algorithm and the developed software-product are useful when the complete lipid profile of the patient is not available or there are uncertainties in defining the dyslipidemia type in order to prognostisise the atherogenic risk. It can also be used to monitor the treatment of dyslipidemia management. The software is, in our opinion, suitable for students' education and postdoctoral training.

REFERENCES

1. Горанов, И., О. Златев, П. Илинов.- В: Липиди. София, Медицина и физкултура, 1983, 99-186.- 2. Климов, А. Н., Н. Г. Никульчева.- В: Липопротеиды, дислипопротеидемии и атеросклероз. Ленинград, Медицина, 1984, 98-142.- 3. Недев, А., К. Тенекеджиев. -*Диагност. и безразруш. контрол* (*Варна*), 1996, № 3.- 4. Огнянов, М., М. Орбезов. Серумни липопротеини и хиперлипопротеинемии. София, Медицина и физкултура, 1979.- 5. Abbate, S., L., J. D. Brunzell. *J. Cardiovasc. Pharmacol.*, **16**, 1990, Suppl. 9, S1-S6.- 6. Assman, G. (ed). Lipid Metabolism Disorders and Coronary Hearth Disease. Mьnchen, MMV-Medizin Verlag,1993, 28-29; 93-96, 100-115.- 7. Betteridge, D. J. *Diabet. Med.*, **6**, 1989, 195-218.- 8. Duda, R., P. Hart. Pattern Classification and Scene Analysis. NewYork, Wiley, 1973.- 9. Gotto, A., M., R.Paoletti (ed). *Atheroscler. Rev.* 1991, 39-59; 75-93, 101-105.- 10. Kolopp, M. *Dietographie*, **37**, 1990, 35-52.- 11. Martin, M. J., S. B.Hulley, et al. *Lancet*, **2**, 1986, 933-936.- 12. Ravid, M., H. Savin, et al. *Kidney Internat.*, **45**, Suppl., 1994, S161-169.- 13. Reverter, J. L., M. Senti. *Diabet. Care*, **17**, 1994, No 3, 189-194.- 14. Schlueter, W., T. Keilani. *Amer. J. Cardiol.*, **72**, 1993, No 20, 374-444.- 15. Schwandt, P. *Drugs*, **40**, 1990, Suppl. 1, p. 38.- 16. Tenekedjiev, K., L. Altev, V. Naumov. - In: Proc. First Internat. Conf. on Marine Industry "MARIND'96". Varna, 1996, 263-276. 17. Winocour, P., H., M. F. Zaver. *Diabet, Med*, 7, 1990, 292.