ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 175, 2004

*Małgorzata Misztal**

THE USE OF SOME PATTERN RECOGNITION ALGORITHMS TO CLASSIFY PATIENTS UNDERGOING CABG

Abstract. The primary goal of pattern recognition is supervised or unsupervised classification in order to solve decision – making problems. Medical diagnosis brings about many practical problems, which may be interpreted as pattern recognition tasks. Making diagnosis of a given patient means to solve a classification problem – we must recognize patient's disease on the basis on some symptoms.

The aim of the article is to present the results of using selected pattern recognition algorithms to classify patients with Coronary Artery Disease undergoing Coronary Artery Bypass Grafting (CABG).

Key words: pattern recognition algorithms, classification trees, coronary artery disease.

1. INTRODUCTION

Although pattern recognition covers a very board spectrum of problems, roughly it consists in assignment of a pattern or object to one class from the finite set of classes:

$$\mathsf{K} = \{1, 2, ..., k\} \tag{1}$$

In decision – theoretical pattern recognition approach we assume that pattern is represented by a vector of numbers $\mathbf{x} = [x_1, x_2, .., x_p]^T$ – the so – called features values, e.g. obtained by scanning of an image at selected grid points or directly from measurements. This prompts us to introduce a pattern or feature space X with as many dimensions as the number of features and to think of an object as one point $\mathbf{x}$ in this space ($\mathbf{x} \in \mathsf{X}$).

Generally, pattern recognition can be defined as an information – reduction process involving measurements of the object to identify distinguishing

* PhD, Department of Statistical Methods, University of Łódź.

attributes, extraction of the features for the defining attributes and assignment of the object to a class based on these features.

The pattern recognition algorithm (or classification or decision rule) ψ maps the feature space into a set of class labels, that is:

$$\psi : \mathsf{X} \rightarrow \mathsf{K} \tag{2}$$

or equivalently partitions space X into the so – called decision regions containing points $\mathbf{x}$ which are assigned by algorithm ψ to the same class.

In statistical pattern recognition we assume that the vector of features describing the recognized pattern $\mathbf{x} \in \mathsf{X} \in R^p$ and its class number $i \in \mathsf{K}$ are observed values of a couple of random variables X and I so the main problem is to find such a decision rule which minimizes the expected value of loss function evaluating the loss in the case of misclassification.

Usually we are lack of exact knowledge of the probability distribution of the features and classes. That's why we are made to apply the only source of information, which is the set of observations (called the learning set):

$$U = \{(\mathbf{x}_1, i_1), (\mathbf{x}_2, i_2), ..., (\mathbf{x}_N, i_N)\}; \ \mathbf{x}_j \in \mathsf{X}; \ i_j \in \mathsf{K} \tag{3}$$

where $\mathbf{x}_j$ denotes the feature vector of the j-th learning pattern and i_j is its correct classification. Additionally, the set of learning patterns from the i-th class is denoted by:

$$U_i = \{\mathbf{x}_{i,l} \in \mathsf{X}, \ l = 1, 2, ..., \mathrm{N}_i\}, \ i \in \mathsf{K} \tag{4}$$

The recognition algorithm based on the learning set U will be denoted as $\psi_U(\mathbf{x})$.

Pattern recognition methods have many practical applications. One of them is medical diagnosis where the learning set consists of case records containing a description of patient's symptoms and corresponding reliable diagnosis.

2. DESCRIPTION OF RECOGNIZED POPULATION

In the Department of Cardiothoracic Surgery of Łódź Medical Academy the set of 762 case records of patients undergoing CABG during 1997–1999 was collected. The data form 1997–1998 constituted the learning set (NU = 407) and from 1999 – the test set (NT = 355).

Outcome after CABG is determined by the preoperative status of the patient so that 13 preoperative risk factors leading to postoperative morbidity and mortality were identified:

1. Age (in years);
2. BSA – body surface area;
3. RRs – systolic blood pressure (in mmHg);
4. RRd – diastolic blood pressure (in mmHg);
5. EF% – left ventricular ejection fraction (in %);
6. AspAt – aspartate aminotransferase (in U/L);
7. Family history of CAD (0 – no; 1 – yes);
8. Diabetes mellitus (0 – no; 1 – yes);
9. AO – arterial obstruction (0 – no; 1 – yes);
10. Left main stenosis $\geqslant$ 75% (0 – no; 1 – yes);
11. Hyperthyroidism (0 – no; 1 – yes);
12. Previous cardiac surgery (0 – no; 1 – yes);
13. Priority of operation (1 – elective; 2 – urgent; 3 – emergent).

The outcome after CABG includes following two classes:

1. Good outcome with no cardiac complications (NU1 = 350; NT1 = 340);
2. Cardiac complications (myocardial infarction and/or low cardiac output) and death (NU2 = 57; NT2 = 15).

One of the problems occurring in medical diagnosis tasks is that some feature values are not available for every patient. That's why we have complete feature vectors for $N = 353$ case records (and NU1 = 149, NT1 = 146; NU2 = 46, NT2 = 15 respectively).

3. SELECTED PATTERN RECOGNITION ALGORITHMS

In order to study the usefulness of some recognition algorithms to classify patients to the risk subgroups we apply the following methods:

1. The Nearest Neighbour Algorithm (NN) – that classifies the unknown pattern vector x by calculating the distances between the object and all objects in the learning set and assigning it to the class that the nearest learning object belongs to. So:

$$\psi_U^{NN}(\mathbf{x}) = i \quad \text{if} \quad d(\mathbf{x};\ \mathbf{x}_{i,l_i}) = \min_{g \in \mathrm{K}} d(\mathbf{x};\ \mathbf{x}_{g,l_g}) \tag{5}$$

$$i \in \mathrm{K} \quad l_i = 1,\ \ldots,\ N_i, \quad l_g = 1,\ \ldots,\ N_g$$

where $d(\mathbf{x}_m;\ \mathbf{x}_n)$ is a distance measure between two object.

2. The α-Nearest Neighbours Algorithm (α-NN) – that classifies the unknown pattern vector x by assigning it to the class that is most common among its α nearest neighbours:

$$\psi_U^{\alpha-NN}(\mathbf{x}) = i \quad \text{if} \quad \alpha_i = \max_{g \in K} \alpha_g, \quad i \in K \tag{6}$$

3. The Distance – Based Algorithm (DB) – that classifies the unknown pattern vector $\mathbf{x}$ to the class scoring the lowest value among the k classifying functions:

$$\psi_U^{BD}(\mathbf{x}) = i \quad \text{if} \quad {}^{BD}D_i(\mathbf{x}) = \min_{g \in K} \{{}^{DB}D_g(\mathbf{x})\}, \quad i \in K \tag{7}$$

where:

$${}^{DB}D_i(\mathbf{x}) = \frac{1}{N_i} \sum_{m=1}^{N_i} d(\mathbf{x}; \mathbf{x}_m) - \frac{1}{2N_i^2} \sum_{m=1}^{N_i} \sum_{n=1}^{N_i} d(\mathbf{x}_m; \mathbf{x}_n), \quad i \in K \tag{8}$$

and $d(*)$ is a distance measure.

The distance measure applied in NN, α-NN and DB algorithms was (see Cessie, Houwelingen 1995):

$$d(\mathbf{x}_m; \mathbf{x}_n) = \sqrt{\sum_{r_1=1}^{p_1} \frac{(x_{mr_1} - x_{nr_1})^2}{2\mathrm{var}(x_{r_1})} + \sum_{r_2=1}^{p_2} \left(\frac{c_c}{c_c - 1}\right) \cdot I\{x_{mr_2} \neq x_{nr_2}\}} \tag{9}$$

where p_1 is the number of continuous variables, p_2 is the number of categorical variables, c_c is the number of different categories for the c-th categorical variable and $I\{A\}$ is the indicator function:

$$I\{A\} = \begin{cases} 1 & \text{if thfe proposition inside the brachets is true} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

4. Linear Classifying Functions – where k linear functions:

$$e_i : X \rightarrow R, \quad i \in K \tag{11}$$

one for each class, are defined. The unknown pattern x is classified to the class scoring the highest values among the k classifying functions:

$$\psi_U(\mathbf{x}) = i \quad \text{if} \quad e_i(x) = \max_{g \in \mathsf{K}} e_g(x), \quad i \in \mathsf{K} \tag{12}$$

In the particular case of normal class – conditional probability density functions we have four following formulas for the estimators of linear classifying functions:

$$_{(1)}\hat{e}_i(\mathbf{x}) = -\frac{1}{2} d_i^2(\mathbf{x}) + \ln q_i, \quad i \in \mathsf{K} \tag{13}$$

$$_{(2)}\hat{e}_i(\mathbf{x}) = -\frac{1}{2}\frac{N-k-p-1}{N-k} d_i^2(\mathbf{x}) + \frac{1}{2}\frac{p}{N_i} + \ln q_i, \quad i \in \mathsf{K} \tag{14}$$

$$_{(3)}\hat{e}_i(\mathbf{x}) = -\frac{1}{2} d_i^2(\mathbf{x}) - \frac{1}{2}\frac{p}{N_i} + \ln q_i, \quad i \in \mathsf{K} \tag{15}$$

$$_{(4)}\hat{e}_i(\mathbf{x}) = -\frac{N-k+1}{2} \ln [1 + N_i(N_i+1)^{-1}(N-k)^{-1} d_i^2(\mathbf{x})] + $$

$$+ \frac{p}{2} \ln \frac{N}{N_i+1} + \ln q_i \tag{16}$$

where $d_i^2(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_i)^T \mathbf{S}^{-1} (\mathbf{x} - \overline{\mathbf{x}}_i)$; $\overline{\mathbf{x}}_i$ is the vector of means for the i-th class, $\mathbf{S}$ is the variance – covariance matrix and q_i is *a priori* probability that object $\mathbf{x}$ belongs to the i-th class. For more details see M. Krzyśko (1990).

5. The Mahalanobis Distance Algorithm – that classifies the unknown pattern vector $\mathbf{x}$ according to the decision rule:

$$\psi_U(\mathbf{x}) = I \quad \text{if} \quad {}^M D_i^2(\mathbf{x}) = \min_{g \in \mathsf{K}} \{ {}^M D_i^2(\mathbf{x}), \quad i \in \mathsf{K} \tag{17}$$

where

$${}^M D_i^2(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_i)^T \mathbf{S}_i^{-1} (\mathbf{x} - \overline{\mathbf{x}}_i), \quad i \in \mathsf{K} \tag{18}$$

6. Classification Trees – which are the rules for predicting the class of an object from the values of its predictor variables, constructed by recursively partitioning the learning set. At each node of the tree we do the following steps:

i. Examine every allowable split on each predictor variable.

ii. Select and execute the "best" of these splits.

iii. Stop splitting on a node when some stopping rule is satisfied.

Classification trees building methods are nonparametric techniques dealing with different kinds of variables (both ordered – continuous and discrete ordinal, and categorical), including missing values. For more details see E. Gatnar (2001), L. Breiman *et al.* (1984) (CART algorithm – Classification and Regression Trees), W.-V. Loh and V.-S. Shih (1997) (QUEST algorithm – Quick Unbiased Efficient Statistical Tree), H. Kim and W.-V. Loh (2000) (CRUISE algorithm – Classification Rule with Unbiased Interaction Selection and Estimation).

Some algorithms are not designed for categorical features so that for each categorical feature x, taking c values $\{c_1, c_2, ..., c_c\}$ we replaced it by $(c-1)$ – dimensional vector $(z_1, z_2, ..., z_{c-1})$, such that $z_i = 1$ if $x = c_i$ and $z_i = 0$ otherwise, for $i = 1, 2, ..., c-1$. If $x = c_c$, the vector consists of all zeros.

In the computations we used STATISTICA PL package; the author's own programmes in STATISTICA BASIC language for NN, α-NN, DB, Mahalanobis distance and linear classifying functions algorithms and classification trees building algorithms: QUEST (http://www.stat.wisc.edu/~loh/quest.html) and CRUISE (http://www.wpi.edu/~hkim/cruise/).

4. THE RESULTS OF CLASSIFICATION

The results of application of selected recognition algorithms are summarized in Tab. 1.

Table 1

Results of patients' classification

Pattern recognition algorithm	Frequency of incorrect diagnosis (%)		
	class	learning set	test set
1	2	3	4
NN	good outcome	x	15.75
	deaths	x	75.00
	total	x	20.25
5-NN	good outcome	x	4.11
	deaths	x	83.33
	total	x	10.13

Table 1 (condt.)

1	2	3	4
7-NN	good outcome deaths total	x x x	2.74 75.00 8.23
9-NN	good outcome deaths total	x x x	2.74 91.67 9.49
11-NN	good outcome deaths total	x x x	1.37 100.00 8.86
Mahalanobis Distances	good outcome deaths total	12.75 30.43 16.92	30.82 16.67 29.75
Linear classifying functions – (13)	good outcome deaths total	5.37 60.87 18.46	9.59 91.67 15.82
Linear classifying functions – (14)	good outcome deaths total	5.37 60.87 18.46	9.59 91.67 15.82
Linear classifying functions – (15)	good outcome deaths total	5.37 60.87 18.46	8.90 91.67 15.19
Linear classifying functions – (16)	good outcome deaths total	5.37 60.87 18.46	8.90 91.67 15.19
Distance – based algorithm	good outcome deaths total	x x x	37.67 16.67 36.08
CART • misclassification costs of predicting class "Deaths" as class "Good outcome": 3:1; • estimated priors; • univariate splits; • stopping rule – 1SE; • constructed tree – see Fig. 1	good outcome deaths total	26.17 36.96 28.72	31.51 33.33 31.65
CRUISE • misclassification costs of predicting class "Deaths" as class "Good outcome": 4:1; • estimated priors; • univariate splits; • stopping rule – 1SE; • constructed tree – see Fig. 2	good outcome deaths total	28.19 23.91 27.18	32.28 33.33 34.81

Table 1 (condt.)

1		2	3	4
QUEST	learning set with missing values; • misclassification costs of predicting class "Deaths" as class "Good outcome": 4:1; univariate splits; stopping rule – 1SE; constructed tree – see Fig. 3	good outcome deaths total	18.59 43.86 22.36	26.18 26.67 26.20
CRUISE	learning set with missing values; • misclassification costs of predicting class "Deaths" as class "Good outcome": 4:1; linear combination splits; stopping rule – 1SE; constructed tree – see Fig. 4	good outcome deaths total	10.29 43.86 14.99	24.33 26.67 26.76

Source: author's calculations.

The classification rule from the CART algorithm (see Fig. 1) is very simple:

- Age ⩽ 62,5 years ⇒ Good outcome;
- Age > 62,5 years ⇒ Deaths.

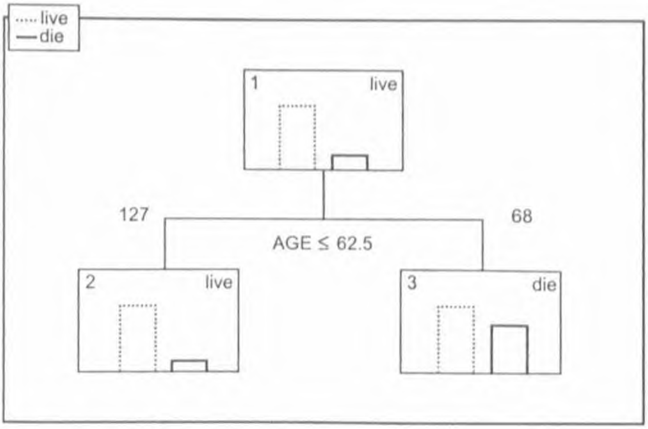

Fig. 1. Classification tree for patients undergoing CABG – the CART method

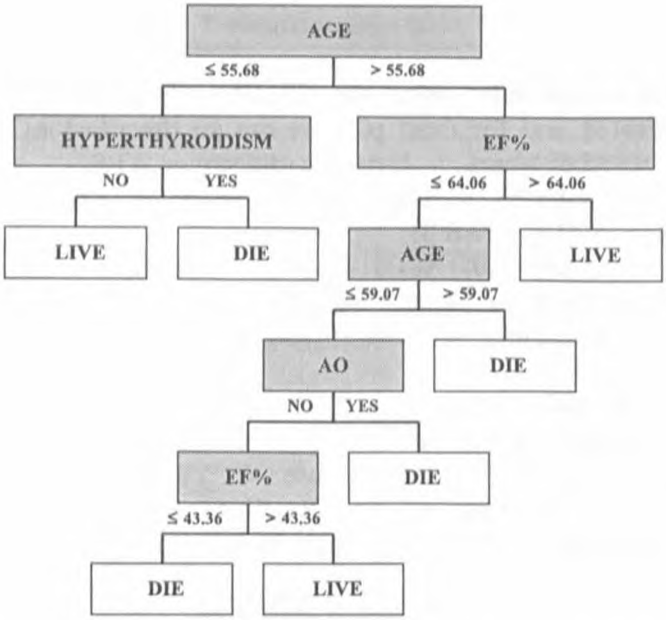

Fig. 2. Classification tree for patients undergoing CABG – the CRUISE method

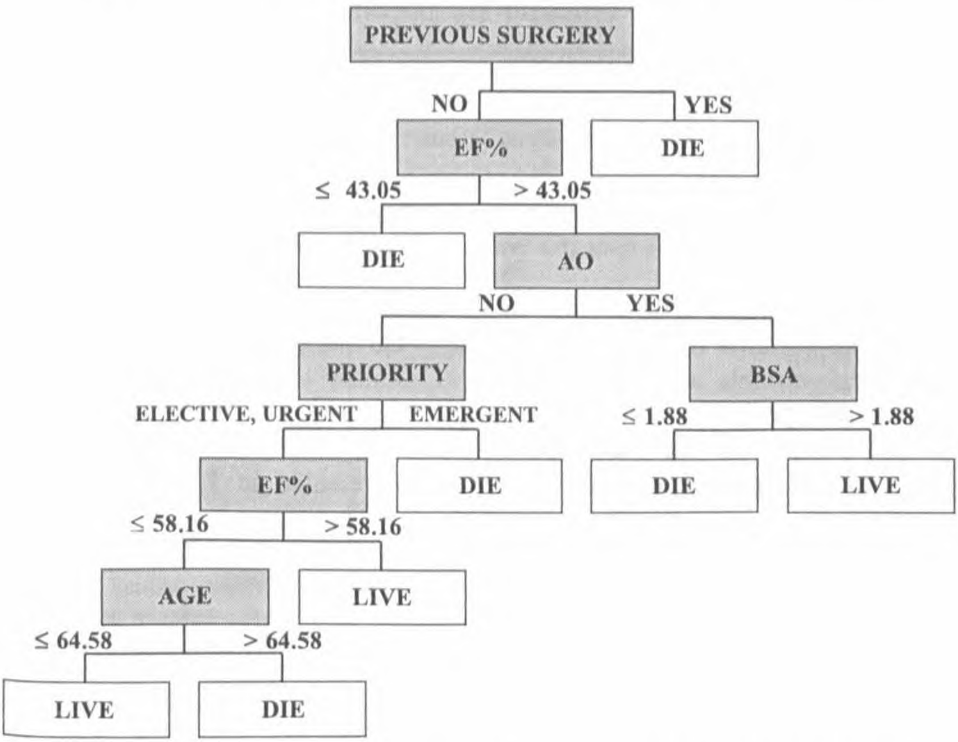

Fig. 3. Classification tree for patients undergoing CABG – the QUEST method (for learning set with missing values)

The classification rules using the CRUISE algorithm (see Fig. 2) are more complicated and for dead patients can be described as follows:

⇒ Age ⩽ 55.68 years ∧ Hyperthyroidism = YES;

⇒ EF % ⩽ 64.6 % ∧ {(Age > 59.07 years) or (Age ∈ (55.68; 59.07] years ∧ AO = "YES")};

⇒ Age ∈ (55.68; 59.07] years ∧ EF% ⩽ 43.36%.

The decision rules for patients who died, constructed using the QUEST algorithm for the learning set with missing values, can be described as (see Fig. 3):

⇒ Previous cardiac surgery = "YES" or:

⇒ Ejection fraction EF % ⩽ 43.05% or:

⇒ Femoral popliteal vascular disease = "Yes" and BSA ⩽ 1.88 or:

⇒ Priority of operation = "EMERGENT" or:

⇒ Ejection fraction EF% ∈ (43.05; 58.16]% and Age > 64.58 years.

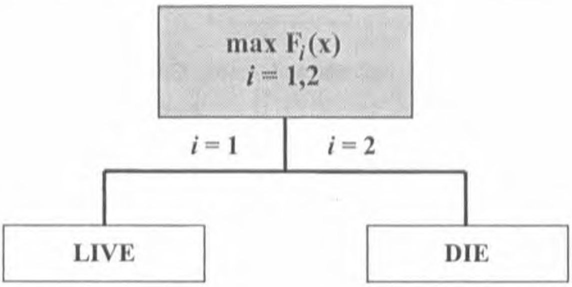

Fig. 4. Classification tree for patients undergoing CABG – the CRUISE method with linear combination splits (for learning set with missing values)

The use of linear combination splits requires the transformation of every categorical variable into an ordered one – the sample values taken by the categorical variable are mapped into 0–1 dummy vectors and the dummy vectors are projected onto their largest discriminate coordinate (called CRIMCOORD; see e.g. Loh and Shih 1997). The CRIMCOORD values for categorical preoperative risk factors are presented in Tab. 2.

Table 2

CRIMCOORD values for categorical risk factors

Variable		Crimcoords
CAD	no yes	–0.04959 0.04959
Diabetes mellitus	no yes	–0.07250 0.07250

Table 2 (condt.)

Variable		Crimcoords
AO	no	–0.09792
	yes	0.09792
Hyperthyroidism	no	–0.1359
	yes	0.1359
Previous cardiac surgery	no	–0.1601
	yes	0.1601
Left main stenosis	no	–0.07903
	yes	0.07903
Priority of operation	elective	–0.160
	urgent	–0.09205
	emergent	0.2529

Source: author's calculations.

The classification rule (see Fig. 4) can be described as follows: we go to the corresponding node if the discriminant score is the maximum. The values of discriminant coefficients are presented in Tab. 3.

Table 3

Discriminant coefficients

Variable	Coefficients – node 1 – good outcome	Coefficients – node 2 – deaths
Constant	–176.40	–173.30
Age	1.134	1.201
BSA	79.50	78.60
RRs	0.0372	0.0138
RRd	0.9052	0.9293
EF%	0.5846	0.5288
AspAt	0.1154	0.1243
Family history of CAD	0.7253	8.71
Diabetes mellitus	–47.31	–46.61
AO	–40.02	–34.61
Hyperthyroidism	–20.73	–17.08
Previous cardiac surgery	–40.01	–28.20
Left main stenosis	18.76	19.03
Priority of operation	–54.09	–44.46

Source: author's calculations.

5. CONCLUSIONS

The following conclusions may be drawn from the Tab. 1 and Fig. 1–4.

The data set is not easy to classify. The set of case records describing patients undergoing CABG has two typical for medical diagnosis tasks properties – a lot of missing values and a great disproportion in the number of patients in classes.

Some pattern recognition algorithms, i.e. NN, α-NN and linear classifying functions are very good at prediction the "Good outcome" class but the classification of the deaths is incorrect.

The distance – based algorithm improves the recognition of objects from the class with complications and death, but the best results we obtained using the Mahalanobis distance rules.

Classification trees building procedures were used assuming unequal misclassification costs. The higher misclassification cost for the class of the deaths makes the pattern recognition task more realistic. Decision rules based on classification trees are easy to interpret and the frequencies of incorrect predictions – not too high. Tree constructed classification provides the researcher with understanding and insight of the data.

It is proper to add that (for trees with univariate splits) we can classify a new patient knowing the values only for a few risk factors.

REFERENCES

Breiman L., Friedman J., Olshen R., Stone C. (1984), *Classification and Regression Trees*, CRC Press, London.

Cessie S., Houwelingen H. C. (1995), *Testing the Fit of a Regression Model via Score Tests in Random Effects Models*, "Biometrics", **51**, 2, 600–614.

Gatnar E. (2001), *Nieparametryczna metoda dyskryminacji i regresji*, PWN, Warszawa.

Kim H., Loh W.-Y. (2000), *CRUISE User Manual, version 1.05*, http://www.wpi.edu/hkim/cruise/.

Krzyśko M. (1990), *Analiza dyskryminacyjna*, WNT, Warszawa.

Kurzyński M. (1997), *Rozpoznawanie obiektów. Metody statystyczne*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.

Loh W.-Y., Shih Y.-S. (1997), *Split Selection Methods for Classification Trees*, "Statistica Sinica", 7, 815–840.

Shih Y.-S. (2000), *QUEST User Manual version 1.8.8*, http://www.stat.wisc.edu/~loh/quest.html.

Małgorzata Misztal

ZASTOSOWANIE WYBRANYCH ALGORYTMÓW ROZPOZNAWANIA OBRAZÓW DO KLASYFIKACJI PACJENTÓW Z CHOROBĄ WIEŃCOWĄ LECZONYCH OPERACYJNIE

Przedmiotem badań w rozpoznawaniu obrazów są metody wspomagania procesów podejmowania decyzji, przy czym przez obraz rozumiemy ilościowy opis przedmiotu, zdarzenia lub zjawiska.

Ogólnie zadanie teorii rozpoznawania polega na określeniu przynależności rozmaitego typu obiektów do pewnych klas. Jeżeli mamy do czynienia z zadaniem rozpoznawania, w którym występuje k klas: $K = 1, 2, ..., k$ – to celem klasyfikacji jest przypisanie rozpoznawanemu obiektowi numeru klasy $i \in K$ na podstawie wartości p wybranych cech obiektu.

W referacie przedstawiono przykłady zastosowań wybranych algorytmów rozpoznawania w diagnostyce medycznej. Obiektami podlegającymi klasyfikacji są pacjenci z chorobą niedokrwienną serca, zakwalifikowani do leczenia operacyjnego, opisani za pomocą wektora cech oceniających ich stan przed i w trakcie zabiegu, a także przebieg leczenia około- i pooperacyjnego.

Klasyfikacji pacjentów do wyodrębnionych grup ryzyka operacyjnego dokonano za pomocą reguł decyzyjnych bazujących na pojęciu minimalnej odległości (algorytmy najbliższego sąsiada i α najbliższych sąsiadów oraz funkcje klasyfikujące oparte na odległościach), liniowych i kwadratowych funkcji klasyfikacyjnych oraz algorytmów tworzących drzewa klasyfikacyjne (CART, QUEST, CRUISE).