# The problem of thresholds determination in ranking classifiers applied in medical diagnostics

DOI: 10.7365 / JHPOR.2016.1.4

**Authors:**

**Andrzej Ameljańczyk[1], Tomasz Ameljańczyk[1]**

*1- Military University of Technology, Warsaw, Poland*

**Keywords:**

*simple and complex ranking classifiers, ranking synthesis, threshold tres, Pareto filter, nondominated diagnosis, ranking diagnosis.*

# Abstract

The paper presents the opportunity for using multi-objective optimization for the development of ranking meta-classifiers being a synthesis of simple ranking domain functions. The method of threshold determination in the developed ranking classifiers applied in medical diagnostics was presented. Analysis was conducted of properties of medical diagnoses acquired with the application of thresholds for both simple and complex classifiers and for meta-classifiers.

# 1. Introduction

Diagnostic information on the patient's health contained in medical data in the area of disease symptoms, risk factors and results of specialist laboratory tests are highly diversified and in most cases of a multimedia character[8,9,13,15]. The development of classifiers deriving from such comprehensive, complex and differentiated medical data is a difficult task. The development of simple classifiers e.g. domain and in particular binary is much easier.[6,24,28]. This leads to a problem of fusion of the acquired diagnostics information, reduced in most cases to so-called simple classifier synthesis. The specific nature of medical diagnostics, due to uncertainty and incompleteness of medical data and due to the fact that the patient may suffer not from a single disease but from two or more, gives preferences to the multi-label (multi-class) classifiers[6,7,13,15]. Application of single-label (single-class) classifiers , including in particular ranking ones, requires the determination of a relevant threshold tres. As explained above, the ranking-leading diagnosis does not need to be correct, not to mention the concomitant diseases positioned further in the ranking. The tres should enable extending the diagnosis beyond the ranking-leading diagnosis in a way to ensure that the actual diseases are covered by the threshold set on one hand and that this threshold set is not excessively 'comprehensive' on the other hand. An excessively comprehensive threshold set extends and increases the costs of the diagnostic process. This set forms the basis for further diagnostics activities (iterations) consisting in the selection of a relevant set of specialist laboratory tests or consultations to make the diagnosis more precise. Due to the commonness of ranking classifier application in the medical diagnostics support algorithms, the problem of determining relevant tres value gains on importance. This paper is an attempt to present the method of tres value determination without the need to use the subjective findings of the decision-making body. Content of diagnosis set resulting from a specific tres value of the simple (component) classifier should be higher, the 'poorer' or less reliable the classifier that was applied. The rankings developed on the basis of correctly performed synthesis (fusion) of simple classifiers[2,8,10,13,28] should have relevantly higher thresholds and therefore less numerous sets of resulting diagnoses. The condition that the set of diagnoses resulting from the threshold value of a classifier being a synthesis of simple classifier should contain an intersection (product) of sets of diagnoses resulting from the thresholds of all component classifiers seems to be intuitively obvious. When introducing the term of thresholds in the rankings generating the single-class classifiers, we may convert them into multi-class classifiers. The specific nature of the medical diagnostics processes, due to common uncertainty and incompleteness of data and possibility of the presence of concomitant diseases, practically excludes the support algorithms using the single-label classifiers. On the other hand, the vast majority of medical diagnostics support algorithms are algorithms that apply the ranking functions defined on the basis of various mathematic similarity (fitting) models[24]. Thus the problem of correct determination of thresholds in medical diagnostics ranking development becomes of great importance.

# 2. Threshold in ranking applications

Let X determine the finite set of medical diagnostic data sets (observations, instances, results), called an observation space.

Let $\mathsf{L} = \{l_1,\ldots,l_m,\ldots,l_M\}$ - set of labels (objects) of disease units, numbered with the $m \in \mathsf{M} = \{1,\ldots,M\}$ index.

A single-label classifier will be the function

$$C : X \to \mathsf{L} \qquad (2.1)$$

Each observation (instance) $x \in X$ is 'associated' with the single label $l_m \in \mathsf{L}$

$$C(x) \in \mathsf{L},\ x \in X \qquad (2.2)$$

A multi-label classifier will be the function

$$C : X \to 2^{\mathsf{L}} \qquad (2.3)$$

Thus

$$C(x) \subset \mathsf{L},\ x \in X \qquad (2.4)$$

For each observation $x \in X$ we may define a relation of ranking preferences $R_x$ in a way that $(l_i, l_j) \in R_x$ when and only when for the observation $x \in X$, lablel $l_i$ is more preferred (is 'better', more 'fitted') than label $l_j$.

The symbol $r(R_x)$ will determine the ranking generated by relation $R_x$[2,5]. In practice, the models (relations) $R_x$ are frequently defined (determined) using ranking functions:

$$f_x : \mathsf{L} \to \mathsf{R}^1 \quad (2.5)$$

values of which are in general determined on the basis of different similarity ('fitting', 'distance') models, for example: Tversky, Bayes, Jaccard, Hamming, Dice, Sokal, Russel, Lance and others) observation $x$ to disease unit labeled $l \in \mathsf{L}$ [6,7,24,28].

$$R_x = \{(l_i, l_j) \in \mathsf{L}^2 | f_x(l_i) \geq f_x(l_j)\} \quad (2,6)$$

Functions $f_x(l)$, $l \in \mathsf{L}$ are sometimes called the utility functions or similarity or fitting ratios. In this paper, we will further assume that the functions $f_x(l)$ are normalized in the range of $[0,1] \subset \mathsf{R}^1$[2, 18, 25, 26, 27].
Thus, if for any $l_i, l_j \in \mathsf{L}$it is true that $f_x(l_i) \geq f_x(l_j)$, from the perspective of result (observation) $x \in X$, $l_i$ is placed in the ranking before the label $l_j$. Definition (2.6) uses purposefully the sign '≥", which results mostly from the fact that the ranking functions are generally not injective functions (this feature results usually from the properties of a model used for definition),[2,20]. Such an assumption results in that the relations of preference $R_x$. are not antisymmetric[20] which means that these determine only the so-called quasi-order[2]. The rankings acquired in this way are not permutations of set $\mathsf{L}$ (are not linear rankings). Adopting the weaker assumptions is implied, however, by the 'practice' of defining the ranking functions, which – as already mentioned – are usually not injective functions. The symbol $r(f_x(l))$ shall determine the sequence ($\mathsf{L}$ set ranking) acquired with the use of function $f_x(l)$[5]. Such ranking functions are frequently used for classifier development. Let $f_x(l)$ , $l \in \mathsf{L}$ be a certain ranking function determined on set $\mathsf{L}$.

This function determines the classifier

$$C : X \to 2^{\mathsf{L}}$$

according to the following formula:

$$C(x) = \arg\max_{l \in \mathsf{L}} f_x(l), \; x \in X \quad (2.7)$$

Relation (2.7) may be presented as follows:

$$C(x) = \arg\max_{l \in \mathsf{L}} f_x(l) = \left\{ \overset{*}{l} \in \mathsf{L} \,\middle|\, f_x\left(\overset{*}{l}\right) = \max_{l \in \mathsf{L}} f_x(l) \right\} \quad (2.8)$$

The relation (2.8) presents the association between the classifier $C(x)$ and the ranking function $f_x(l)$. The classifier of type (2.7) shall mean the simple ranking classifier. The specific nature of medical diagnostics makes the single-label classifiers, including in particular those developed on the basis of ranking functions, insufficient for physicians. As an initial diagnosis, they select a certain number of classes (disease units) placed in the ranking at top positions. This generates a significant problem of determining the values of the so-called threshold – tres, on which the 'content' of initial diagnosis $D_x(tres)$ (threshold diagnosis) depends. It may be written as follows:

$$D_x(tres) = \{l_m \in \mathsf{L} | f_x(l_m) \geq tres\} \quad (2.9)$$

The value of the real number of tres is of significance since it determines a certain diagnostic compromise between striving for increased reliability of initial diagnosis and the number of subsequent diagnostic iterations and, therefore, the total time and cost of the diagnostic procedure. *Tres* in diagnostic classifiers based on the ranking functions are generally the higher, the more precise and reliable are the models being the foundation of the functions.

Synthesis of simple classifiers, as a result of which more precise and reliable classifiers are obtained (with lower classification error) increases the thresholds and, therefore, narrows the set of result diagnoses. The consequence of such an approach is in general a shorter period of the diagnostic process and its lower cost. 'Component' classifier thresholds are obviously lower due to their lower accuracy and reliability.

# 3. Simple ranking synthesis – meta-rankings

Let further be $\mathsf{L} = \{l_1, \ldots, l_m, \ldots, l_M\}$ - a finite label set and ranking function of the following type:

$$f_x : \mathsf{L} \to \mathsf{R}^N$$
$$f_x(l) = \left(f_x^1(l), \ldots, f_x^n(l), \ldots, f_x^N\right) \in \mathsf{R}^N, \quad (3.1)$$

This function generates the following set (committee) of simple (ranking) classifiers):

$$C(x) = \{C_1(x), \ldots, C_n(x), \ldots, C_N(x)\} \quad (3.2)$$

where $C_n(x) = \arg\max_{l \in \mathsf{L}} f_x^n(l) \quad (3.3)$

Set $Y_x$ shall be the ranking image of set $\mathsf{L}$ for the observation $x \in X$, given by the function $f_x$ (3.1).

$$Y_x = f_x(\mathsf{L}) = \{y = f_x(l) \in \mathsf{R}^N | l \in \mathsf{L}\} \quad (3.4)$$

Element $y \in f_x(\mathsf{L})$ is an image of label l in the meaning of its assessment by all ranking functions $f_x^n(l)$, understood as multi-objective level of similarity (fitting) of the observation $x \in X$ to the disease unit labeled $l \in \mathsf{L}$.

Thus $y = (y_1, \ldots, y_n, \ldots, y_N) = (f_x^1(l), \ldots, f_x^n(l), \ldots, f_x^N(l)) \in R^N$ (3.5)
where $y_n = f_x^n(l)$ ranking value of label $l \in L$ in the meaning of n-ranking function associated with observation $x \in X$. For each $x \in X$ it is true that $Y_x \subset Y = [0,1] \times \ldots \times [0,1]$. Set Y shall mean the classifier synthesis area. The synthesis relation (or relation of preferences of classifier committee) shall be the following relation:

$$R \subset f_x(L) \times f_x(L) = Y_x \times Y_x$$

defined as follows:

$$R = \{(y, z) \in Y_x \times Y_x | \text{"committee prefers } y \text{ over } z\text{"}\} \quad (3.6)$$

The synthesis relation *R* expresses the principle of preferences of committee in the area of deciding whether label $l_k$ is 'better fitted' to observation $x \in X$ compared to label $l_m$. There are many known preferences applicable to such synthesis. The most typical principle is the Pareto principle (relation, filter). It states that label $l_k$ is more preferred (better fitted to observation x ) than label $l_m$ provided that $l_k$ is at least at the same position (or higher) as label $l_m$ in the ranking of each committee member[2,26]. This means that the following must be true:

$$f_x^n(l_k) \geq f_x^n(l_m), \; n \in N \quad (3.7)$$

The Pareto Filter (PF) is an algorithm enabling determination from any set of elements the set of elements of the highest quality in this set (in the meaning of Pareto relation)[3,4,26]. The effect (result) of applying the Pareto filter on set Y is so-called 'Pareto front' (set of nondominated (minimum)) elements in the meaning of Pareto relation $Y_N^R$ defined as follows:

$$Y_x^{RN} = \{y \in Y_x | n_{\text{does not exists with } x \text{ that } (z,y) \in R}\}$$

Therefore, the result of the filtration process is decisive for the adopted preferences (filtration) relation *R* (in more detail – its properties). So, such a relation is frequently called a preference filter or briefly: filter. The general reflection of the Pareto filter is a cone filter (CF), in which the filtration reaction is generated by a cone[2,3]. The CCS task – complex (integrated) classifier synthesis – may be defined as multi-objective optimization[2,3,4] of the form:

$$CCS = (L, f_x, R) \quad (3.8)$$

which may be abbreviated to[2,3,26] :

$$(Y_x, R) \quad (3.9)$$

The synthesis relation *R* may be used to develop a complex, multi-label classifier (meta-classifier) and meta-ranking (committee ranking), being a 'specific synthesis' of component rankings determined by the ranking functions $f_x^n(l)$, $n \in N$. For meta-ranking and component rankings one may determine the relevant (in the meaning of preference relation *R*) thresholds, which will enable the acquisition of justified and extended multi-label classifications upon application. The solution of the task (3.8) is thus an anti-image[4,20] of the task solution (3.9), i.e. subset of labels, from which there are no 'better' labels in the set L (better fitted) to the observation $x \in X$.

$$L_x^{RN} = f_x^{-1}(Y_x^{RN}) \quad (3.10)$$

where $Y_x^{RN} = \{y \in Y_x | \text{ does not exists with } z \in Y_x - \{y\} \text{ that } (z, y) \in R\}$ (3.11)

thus $L_x^{RN} = f_x^{-1}(Y_x^{RN}) = \{l \in L | f_x(l) \in Y_x^{RN}\}$ (3.12)

Set $L_x^{RN}$ is called a nondominated label set or Pareto set (front)[3,4,26,27]. This is a subset of these labels from the set L, from which there are no better 'fitted' labels to the observation $x \in X$. The integrated classifier in the meaning of relation R (meta-classifier) is the complex classifier

$$C_R(x) = f_x^{-1}(Y_x^{RN}) \subset L \quad (3.13)$$

This is in general the multi-label classifier, which assigns to each observation (instance) $x \in X$ the 'optimum' subset of nondominated labels $L_x^{RN}$ in the meaning of relation R. In medical diagnostics, this diagnosis is considered the 'best fitted' diagnosis corresponding to observation $x \in X$ This proposal is the most important and the most frequently applied diagnostic reference in the process of computer diagnosing support [6,7]. Having the set of simple classifiers based on the ranking functions, we may develop a meta-classifier (component classifier) based on the ranking function F (ranking meta-function), being a synthesis of the applied ranking functions $f_x^n(l), n \in N$. It is done by determining the 'resulting' ranking $r(F)$ the most similar' to the component rankings $r(f_x^n(l))$, $n \in N$. A simple method for the development of meta-ranking $r(F)$ is using the ranking function defined on the basis of the integrated classifier synthesis task (3.8) considered as a typical task of multi-objective optimization[3,4,26]. The ranking meta-function *F* may be defined using the Minkowski standard in the following manner (3.14) :

$$F(f_x(l)) = F(f_x^1(l), \ldots, f_x^n(l), \ldots, f_x^N(l)) = F(y) = \left\|\overset{*}{y}(x)\right\|_p + \left\|\overset{*}{y}(x) - y\right\|_p, \; y \in Y_x, \; p \geq 1$$

The symbol $\|y\|_p, p \geq 1$ shall apply to standard with parameter p [2,3]. Element $\overset{*}{y}(x) = \left(\overset{*}{y}_1(x), \ldots, \overset{*}{y}_n(x), \ldots, \overset{*}{y}_{N1}(x)\right) \in R^N$ is a greatest lower bound of set $Y_x$ for relation R[3,4], i.e. so-called ideal point (utopian point)[3,4,26]. The element (point) $\overset{*}{y}(x) \in R^N$ an image of utopian (virtual) label of the highest possible ranking parameters, i.e. such that

$$\overset{*}{y}_n(x) = \max_{y \in Y_x} y_n = \max_{l \in L} f_x^n(l), \; n \in N \quad (3.15)$$

Assuming for convenience that all ranking functions are normalized to the range of $[0,1] \subset R$ , we will obtain(3.14) upon normalization:

$$F(y)=1-\frac{1}{\overset{*}{y}(x)}\left\|\overset{*}{y}(x)-f_x(l)\right\|_p=1-\alpha_x\left\|\overset{*}{y}(x)-f_x(l)\right\|_p$$

i.e. in consequence a normalized ranking meta-function

$$F(f_x(l))=1-\alpha_x\left\|\overset{*}{y}(x)-f(l)\right\|_p,\ l\in L,\ p\geq 1 \qquad (3.16)$$

where $\alpha_x$ - normalization coefficient[7].
The ranking meta-classifier shall adopt the following form according to (2.7):

$$C_F(x)=\arg\max_{l\in L}F(f_x(l))\subset L \qquad (3.17)$$

Set $C_F(x)$ may be considered an initial diagnosis. In many cases this is a single-element set[3, 4,26]. The extension of initial diagnosis with regard to the specific nature of medical diagnostics may be a set of diagnoses compliant with ranking $r(F)$ above the determined tres:

$$D_F(tres)=\{l\in L|F(f(l))\geq tres\} \qquad (3.18)$$

The *tres* symbol shall be the value of threshold in the ranking acquired thanks to ranking meta-function (3.16). The problem of the method for determining an optima (relevant) value of such threshold arises here.

# 4. Method of ranking function threshold determination

Tres in diagnostic applications should meet a series of conditions referred to in Clause 2 of this paper.

One of the key conditions to be met by tres is a request that the threshold label set (3.18) contains the set of all labels, of which there are no better (more fitted or similar) labels in the label set $L$ in the meaning of the adopted preference relation $R$, i.e. the set of nondominated labels $L_N^R$. Set of threshold labels cut off by the determined value of tres is specified by the relation (3.18). To meet this condition, the following must be true:

$$L_x^{RN}=f_x^{-1}(Y_x^{RN})\subset D_x(tres) \qquad (4.1)$$

The correlation between the preference relation $R$ and the ranking function $F(y)$ is specified by the following lemma.

**Lemma 1**
Let R be the preference relation (3.6) and function F be a ranking meta-function (3.14). For each pair $(y,z)\in R$ it is true that $F(y)\geq F(z)$. This means that y is higher in the ranking than z (i.e. if y is better from z in the meaning of the adopted preference relation $R$, y must be placed nearer the ideal point than element z) .

Proof:
The condition $F(y)\geq F(z)$ may be written as follows:

$$\left\|\overset{*}{y}(x)\right\|_p-\left\|\overset{*}{y}(x)-y\right\|_p\geq\left\|\overset{*}{y}(x)\right\|_p-\left\|\overset{*}{y}(x)-z\right\|_p,\ p\geq 1$$

Subtracting by sides $\left\|\overset{*}{y}(x)\right\|_p$ we obtain

$$-\left\|\overset{*}{y}(x)-y\right\|_p\geq-\left\|\overset{*}{y}(x)-z\right\|_p \text{ and further}$$

$$\left\|\overset{*}{y}(x)-y\right\|_p\leq\left\|\overset{*}{y}(x)-z\right\|_p \qquad (4.2)$$

For Minkowski metric[2,26]:

$$\left\|\overset{*}{y}(x)-y\right\|_p=\left(\sum_{n\in N}\left(\left|\overset{*}{y}(x)-y_n\right|^p\right)\right)^{1/p},\ p\geq 1 \qquad (4.3)$$

To prove the lemma, we must demonstrate that for any pair $(y,z)\in R$ is is true that (4.2).
if $(y,z)\in R,\ y_n\geq z_n,\ n\in N$ (4.4)

From the definition of greatest lower bound $\{\overset{*}{y}(x)\}$ of set $Y_x$:

$$\overset{*}{y}(x)\geq y \quad \text{for each } y\in Y_x \qquad (4.5)$$

Therefore it is true that:

$$a)\ \overset{*}{y}_n(x)\geq y_n,\ n\in N \qquad b)\ \overset{*}{y}_n(x)\geq z_n,\ n\in N \qquad (4.6)$$

With regard to (4.6), the relation (4.3) shall be as follows:

$$\left(\sum_{n\in N}\left(\overset{*}{y}_n(x)-y_n\right)^p\right)^{1/p}\leq\left(\sum_{n\in N}\left(\overset{*}{y}_n(x)-z_n\right)^p\right)^{1/p} \qquad (4.7)$$

The relation (4.7) will be met if we demonstrate that:

$$\left(\overset{*}{y}_n(x)-y_n\right)\leq\left(\overset{*}{y}_n(x)-z_n\right) \text{ for each } n\in N$$

Subtracting from both sides $\overset{*}{y}_n(x)$, we obtain

$$-y_n\leq -z_n \text{ that is } y_n\geq z_n,\ n\in N$$

This is true since $(y,z)\in R$ and this means that $y_n\geq z_n$ for each $n\in N$.

---

Lemma 1 may be also read as follows: if element y precedes element z in the meaning of preference relation $R$ (is better in terms of $R$ the element y is in the meta-ranking $r(F)$ at higher position (at east at the same position) than element z, which further means that if label $l_m$ such that $f_x(l_m)=y$ is preferred (in terms of $R$ more than label $l_k$ such that $f_x(l_k)=z$, then in the ranking $r(F)$ determined by ranking meta-function $F$, label $l_m$ precedes label $l_k$ . To meet the general condition (4.1), value of *tres* should be such that even the most distant from the ideal point (in terms of the adopted metric) element $y \in Y_x^{RN}$ should be placed in a set cut-off by tres.

Let's determine with $\overline{tres}$ symbol the number equal to $\max_{y \in Y_x^{RN}} \left\| \overset{*}{y}(x) - y \right\|_p$ (4.8)

Thus the *tres* value meeting the condition (4.1) may be defined as follows:

$$tres = \left\| \overset{*}{y}(x) \right\|_p - \overline{tres} \quad (4.9)$$

The set of 'cut-off elements' by the threshold of function $F(y)$ shall be as follows:

$$Y_x(F) = \{ y \in Y_x | F(y) \geq tres \} \quad (4.10)$$

The anti-image of this set may be considered the value (indication) of the new ranking meta-classifier $C_F^{\overset{*}{y}}(x)$:

$$C_F^{\overset{*}{y}}(x) = f_x^{-1}(Y_x(F)) = D_x(F) \quad (4.11)$$

This set may be considered the extender initial diagnosis.

The key property of this new extension is the fact that the content (indication) of the integrated classifier $C_R(x)$ is a subset of values (indications) of the ranking meta-classifier $C_F^{\overset{*}{y}}(x)$ i.e. $C_R(x) \subset C_F^{\overset{*}{y}}(x)$. This is true since $Y_x^{RN} \subset Y_x(F)$ i.e. for each $y \in Y_x^{RN}$ it must be true that: $F(y) \geq tres$

According to relation (4.8) for each $y \in Y_x^{RN}$ it is true that

$$\left\| \overset{*}{y}(x) - y \right\|_p \leq \max_{y \in Y_x^{RN}} \left\| \overset{*}{y}(x) - y \right\|_p$$

thus $$-\left\| \overset{*}{y}(x) - y \right\|_p \geq -\max_{y \in Y_x^{RN}} \left\| \overset{*}{y}(x) - y \right\|_p$$

addiing to both sides $\left\| \overset{*}{y}(x) \right\|_p$ we obtain

$$\left\| \overset{*}{y}(x) \right\|_p - \left\| \overset{*}{y} - y \right\|_p = F(y) \geq \left\| \overset{*}{y}(x) \right\|_p - \max_{y \in Y_x^{RN}} \left\| \overset{*}{y}(x) - y \right\|_p = tres$$

thus for each $y \in Y_N^R$ we obtain $F(y) \geq tres$

Therefore $$f^{-1}(Y_x^{RN}) = L_x^{RN} \subset f_x^{-1}(Y_x(F) = D_x(F))$$
thus $$L_x^{RN} \subset D_x(F)$$

A set of labels $D_x(F)$ may contain, apart from labels from set $L_x^{RN}$, the labels ('diagnosis extensions') which do not belong to this set. These labels are of unique properties, specified (among others) by the two following lemmas.

## Lemma 2

If in the set $D_x(F) \setminus L_x^{RN}$ (in the set of the remaining initial diagnosis labels) there is a label better than any other label in set $L_x^{RN}$ in the meaning of ranking meta-function *F*, in set $L_x^{RN}$ there is at least one label better than such label in the meaning of preference relation $R$.

Proof:

Let $l_k \in D_x(F) \setminus L_x^{RN}$ such that
$F(l_k) \geq F(l_m)$ for a certain label $l_m \in L_x^{RN}$

Since $l_k \notin L_x^{RN}$, according to (3.10) and (3.11) in set $L_x^{RN}$ there must be such label $l_r$, that $(f_x(l_r), f_x(l_k)) \in R$

## Lemma 3

For each label $l_k \in D_x(F) \setminus L_x^{RN}$ in set $L_x^{RN}$ there is such label $l_m$, which is higher from this label (at least at the same position) in the meta-ranking $r(F)$.

Proof:

In set $L_x^{RN}$ there is such label $l_m$, which is better than label $l_k$ in the meaning of relation $R$ (see (3.10),(3.11) ) i.e. such that $(f_x(l_m), f_x(l_k)) \in R$. If so, under lemma 1 the following is true:

$$F(f_x(l_m)) \geq F(f_x(l_k))$$

which means that $l_m$ is positioned higher in the meta-ranking that label $l_k$. Thus, the tres defined in (4.9) determines the nontrivial 'cut-off diagnoses set' containing all non-dominated diagnoses (3.12) and the additional ones meeting lemmas 2 and 3.

The ranking meta-function $F(y)$ may be considered as the effect of synthesis of ranking functions $f_x^n(l)$ $n \in N$ and ranking $r(F)$, respectively, as the synthesis of component rankings $r(f_x^n(l))$, $n \in N$.

For the component ranking functions $f_x^n(l)$ we may determine the value of $tres(n)$, $n \in N$. These numbers, identically as the *tres* value, must meet the following conditions (among others, be a part of the 'cut-off' label set, labels from which there are no better fitted labels $L_x^{RN}$ ). Thus we may determine them as follows (see (3.11),(3.12)):

$$tres(n) = \min_{y \in Y_x^{RN}} y_n, \quad n \in N \quad (4.12)$$
$$Y_x(tres(n)) = \{ y \in Y_x | (y_n \geq tres(n)) \} \quad (4.13)$$

It is easy to demonstrate that

$$Y_x^{RN} \subset Y_x(tres(n)) \text{ for each } n \in N \quad (4.14)$$

| m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $f_x^1$ | 0.3 | 0.4 | 0.5 | 0.6 | 0.6 | 0.5 | 0.5 | 0.3 | 0.2 | 0.5 | 0 | 0.1 | 0.2 | 0.4 | 0.5 | 0.4 | 0.3 | 0.2 | 0.3 | 0.4 |
| $f_x^2$ | 0.6 | 0.7 | 0.6 | 0.4 | 0.3 | 0.2 | 0.1 | 0.1 | 0.1 | 0.5 | 0.4 | 0.6 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.4 | 0.5 | 0.6 |

Table 1.

Set of labels (diagnoses) determined by *tres*(n) is as follows:

$$D_x(tres(n)) = f_x^{-1}(Y_x(tres(n))) \quad (4.15)$$

For each $n \in \mathrm{N}$ it is obviously true that:

$$\mathrm{L}_x^{RN} \subset D_x(tres(n)) \quad (4.16)$$

# 5. Ranking synthesis – diagnostic example

The example concerns determination of initial diagnosis for the patient on the basis of diagnosed symptoms of disease and risk factors with regard to their intensification. The applied diagnostic model, described in the papers[5,6,7] also considers the significance levels for the applicable symptoms and risk factors in diagnosing individual diseases. Data on the patient's health condition (observation $x \in X$ ) were divided by domains into two areas: data on symptoms presence and data on risk factors and their intensification. These data formed a basis for developing two classifiers.

$f_x^1(l)$ - similarity rate for symptoms[6,7]
$f_x^2(l)$ - similarity rate for risk factors[6,7] (5.1)

Set $\mathrm{L} = \{l_1, ..., l_m, ..., l_{20}\}$ in the analyzed example is a set of twenty disease units (labels) indexed with $m \in \mathrm{M}$, presented in Table 1. This table also contains (for the adopted observation of medical results $x \in X$ ) the values of both ranking functions (5.1). As we see, these values are not injective, therefore the rankings $r(f_x^1)$ and $r(f_x^2)$ developed on the basis thereof will be not linear.

On the basis of the a/m functions, the following two classifiers were developed:

$$C_1(x) = \arg\max_{l \in \mathrm{L}} f_x^1(l)$$
$$C_2(x) = \arg\max_{l \in \mathrm{L}} f_x^2(l) \quad (5.2)$$

Classifications acquired with these classifiers (initial diagnoses) are as follows:

$$D_1(x) = C_1(x) = \{l_4, l_5\},\ D_2(x) = C_2(x) = \{l_2\} \quad (5.3)$$

Diagnostic concluding on the basis of these results is most probably hindered and doubtful, for example due to the fact of their divergence:

$$C_1(x) Ç C_2(x) = Ć \quad (5.4)$$

Determining the applicable thresholds - *tres* (1) and *tres* (2) – we could 'extend' the sets of indications (5.3), increasing the safety of further diagnostic process. The rankings of set $\mathrm{L}$ generated by the ranking functions (6.1) are 'highly diversified' and have the following form:

$$r(f_x^1) = \langle 4,5,3,10,15,6,7,2,20,14,16,1,19,17,8,13,18,9,12,11 \rangle$$
$$r(f_x^2) = \langle 2,3,20,1,13,12,10,14,19,4,15,18,11,5,16,6,17,7,8,9 \rangle$$

Assuming on arbitrary basis that $tres(1) = 0{,}5$ and $tres(2) = 0{,}6$ we obtain:

$$C_1(tres(1) = 0{,}5) = D_x(tres(1)) = \{l_4, l_5, l_3, l_{10}, l_{15}, l_6, l_7\}$$
$$C_2(tres(2) = 0{,}6) = D_x(tres(2)) = \{l_2, l_3, l_{20}, l_1, l_{13}, l_{12}\}$$

Referring to ((5.3) and (5.4)) we obtain the initial indication:

$$D_x^1(0) = D_x(tres(1)) \cap D_x(tres(2)) = \{3\} \quad (5.6)$$

Adopting the proposal $D_x^1(0)$ as an initial diagnosis is risky – only one element $l_3$, whereas adopting the proposal $D_x^2(0)$ is safer, however, expensive due to time and costs of further diagnostic iterations making the diagnosis more precise.

Figure 1 presents the area of synthesis $Y$ as well as ranking image $Y_x$ of set for observation $x \in X$ (3.5) and ideal point $\overset{*}{y}(x) = (0{,}6; 0{,}7)$ (see (3.15). This point is a ranking image of virtual label (utopian label) of such a disease unit, which would have the highest similarity rate in terms of symptoms and risk factors under observation $x \in X$ [6,7,24]. As the synthesis relation the Pareto relation was adopted – the most common in such cases[2,3,4,6,26,27]:

| m | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $F(l_m)$ | 0.06 | 0.722 | 0.781 | 0.622 | 0.522 | 0.412 | 0.314 | 0.251 | 0.201 | 0.699 | 0.251 | 0.412 | 0.510 | 0.639 | 0.606 | 0.475 | 0.339 | 0.422 | 0.562 | 0.699 |
| $r(F)$ | 3 | 2 | 10 | 20 | 14 | 4 | 15 | 1 | 19 | 5 | 13 | 16 | 18 | 6 | 12 | 17 | 7 | 8 | 11 | 9 |
| $r(f_x^1)$ | 4 | 5 | 3 | 10 | 15 | 6 | 7 | 2 | 20 | 14 | 16 | 1 | 19 | 17 | 8 | 13 | 18 | 9 | 12 | 11 |
| $r(f_x^2)$ | 2 | 3 | 20 | 1 | 13 | 12 | 10 | 14 | 19 | 4 | 15 | 18 | 11 | 5 | 16 | 6 | 17 | 7 | 8 | 9 |

**Table 2.**

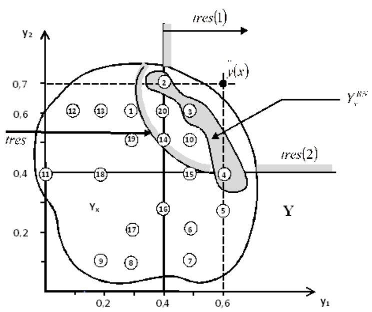


**Figure 1. Area of synthesis Y and set $Y_x$ along with ranking reference point $\overset{*}{y}(x)$**

According to (3.11) and (3.13), an integrated classifier (generated in result of synthesis) shall be the classifier:

$$C_R(x) = f_x^{-1}\left(Y_x^{RN}\right) \subset L \qquad (5.8)$$

Set $Y_x^{RN} = \{2,3,4\}$, present in (5.8) was marked in Figure 1

Thus

$$C_R(x) = f_x^{-1}\left(\{(0{,}4;0{,}7),(0{,}5;0{,}6),(0{,}6;0{,}4)\}\right) = \left\{l_m \in L \,\middle|\, f_x(l_m) \in Y_x^{RN}\right\} = \{l_2,l_3,l_4\}$$

that is $C_R(x) = L_x^{RN} = \{l_2,l_3,l_4\} \subset L$

This is a set of disease units (labels), from which there are no 'more fitted' units in set L with regard to observation $x \in X$ in the area of diagnosed symptoms and risk factors. This is the effect of operation of the integrated classifier (3.13).

Let's notice that

$$C_R(x) \cap D_x(tres(1)) = \{3,4\}$$
$$C_R(x) \cap D_x(tres(2)) = \{2,3\} \qquad (5.10)$$

An obligatory condition (4.1) is therefore not met which means that the arbitrary values *tres*(1) and *tres*(2) were not selected properly.
Further part of this paper presents the results of classification acquired by application of an additional complex classifier developed based on ranking meta-function (3.16). Values of this function and applicable rankings are presented in Table 2.

For simplified form, instead of $\|y\|_p$, we will use $\|y\|$ (assuming that p=2). Threshold *tres*, in line with (4.9) has a value of:

$$tres = \left\|\overset{*}{y}(x)\right\| - \overline{tres}\ \text{, where } \left\|\overset{*}{y}(x)\right\| = \|(6{,}7)\| \cong 0{,}922$$

wheras $\overline{tres} = \max_{y \in Y_N^R} \left\|\overset{*}{y}(x) - y\right\| = 0{,}3 \qquad (5.11)$

Thus $tres = 0{,}922 - 0{,}3 \cong 0{,}62$

Figure 1 presents the cut-off area according to threshold tres=0,62 and cut-off set $Y_x(F)$

$$Y_x(F) = \{y \in Y_x \mid F(y) \geq 0{,}62\} = \{2,3,4,10,14,20\}$$

Thus according to (4.11) we obtain:

$$C_F^{\overset{*}{y}}(x) = D_x(F) = \{l_2,l_3,l_4,l_{10},l_{14},l_{20}\} \qquad (5.12)$$

It is true that:

$$C_R(x) \subset C_F^{\overset{*}{y}}(x) \text{ that is } \{l_2,l_3,l_4\} \subset \{l_2,l_3,l_4,l_{10},l_{14},l_{20}\}$$

As we see, $C_F^{\overset{*}{y}}(x)$ is the extension of diagnosis of the integrated classifier $C_R(x)$ by certain new labels (disease units):

$$C_F^{\overset{*}{y}}(x) \setminus C_R(x) = \{l_{10},l_{14},l_{20}\} = C_R^{\oplus}(x) \qquad (5.13)$$

The labels from this set are better in terms of meta-function *F* from label no. 4 (see Lemma 2). In line with Lemma 2, if there is a label in this set, which is better than any other label from set $C_R(x)$ in terms of meta-function *F*, there is a label in set $C_R(x)$ better than such label in the meaning of *R*. Therefore:

for $l_{10} \in C_R^{\oplus}(x)$ such label is $l_3$,
for $l_{14}$ such label is $l_2$ and $l_3$,
for $l_{20}$ such label is also $l_2$ and $l_3$.

The ranking function $F(y)$ generates the ranking meta-classifier (3.17):

$$C_F(x) = \arg\max_{l \in L} F(f_x(l)) = \{3\} \qquad (5.14)$$

For analytical purposes is is worth to determine the values of *tres*(1) i *tres*(2) for the rankings leading to simple classifiers $C_1(x)$ and $C_2(x)$ (see 4.12). Pursuant to (4.12) we will obtain the following values *tres*(1)=0,4 and *tres*(2)=0,4 and therefore the applicable cut-off sets (4.13):

$$Y_x(tres(1) = 0{,}4) = \{4,5,3,10,15,6,7,2,20,14,16\}$$
$$Y_x(tres(2) = 0{,}4) = \{2,3,20,1,13,12,10,14,19,4,15,18,11\}$$

The obtained results and their properties are presented in the table below.

Table 3 presents the list of indications of the individual classifiers on the basis of observation $x \in X$ with reference to set $\mathsf{L}$.

As we see, a significant discrepancy of the component rankings $r(f_1)$ and $r(f_2)$ resulted in lowering the *tres*(1) and *tres*(2) and thus led to highly 'extensive' (blurred, non-expressive) diagnostic proposals.

The last two columns of the table present information on conformity of the conformity rate of the indication of the given classifier with 'baseline indication' $\mathsf{L}_x^{RN}$ , concerning the set of labels, from which there are no other better fitted (see (3.15)). The last column of the table contains the values of Jaccard's conformity (similarity) index[24,28] of indication of a given classifier with a set of diagnosis, from which there are no better fitted (see (3.11) and (3.12)). The next to last column contains information on 'intersection' of the indication of a given classifier with the set of nondominated labels $C_R(x) = \mathsf{L}_x^{RN} = \{l_2, l_3, l_4\} \subset \mathsf{L}$ and coverage index with regard to intersection (proportion of a number of nondominated labels contained in the classifier indication to the total number of nondominated labels).

# 7. Summary

The paper presents the method of synthesis of simple classifiers based on the ranking function and using the multi-objective optimization methodology. The synthesis applies the simpliest relation used in the multi-objective optimization (3.7), the so-called Pareto relation. Using the Pareto classifier (3.13) also enables additional determination of the tres values for ranking classifiers (simple) and any complex classifiers. Among others, the paper defines the ranking meta-classifier (3.17) and applicable tres. Also, the specific properties of the additionally obtained diagnoses present in the cut-off set $D_x(F)$ were discussed. In practice, determination of tres (4.9) must not require determination of Pareto set, which may be complicated. Useful estimation of the tres value may be a distance of the lexicographic element of set $Y_x$ the most distant from the ideal point, which is much easier to determine[3,4]. The method of threshold determination may be used for any ranking classifiers, among others, those developed on the basis of weighted totals of component rankings, averaged rankings, voting rankings, etc. Even a brief analysis of the results obtained in the example (including analysis of Figure 1) demonstrates obvious benefits resulting from synthesis of classifiers leading to increased value of many indexes used for the assessment of quality of classifiers, such as: ranking function injectivity index, ambiguity index, expressiveness index and reliability of indications [1,6,7,28].

| NO. | Classifier | CLASSIFIER INDICATIONS INITIAL DIAGNOSIS | INTERSECTION WITH SET $\mathsf{L}_N^R$ | JACCAR'S CONFORMITY TO SET $\mathsf{L}_N^R$ |
|---|---|---|---|---|
| 1 | $C_1(x)$ | $\{l_4, l_5\}$ | $\{l_4\}$ , ⅓ | ¼ |
| 2 | $C_2(x)$ | $\{l_2\}$ | $\{l_2\}$ ⅓ | ⅓ |
| 3 | $C_1(tres(1)) = 0{,}5$ | $\{l_4, l_5, l_3, l_{10}, l_{15}, l_6, l_7\}$ | $\{l_3, l_4\}$ ⅔ | ¼ |
| 4 | $C_2(tres(2)) = 0{,}6$ | $\{l_2, l_3, l_{20}, l_{11}, l_{13}, l_{12}\}$ | $\{l_2, l_3\}$ ⅔ | 2/7 |
| 5 | $C_1(tres(1)) = 0{,}4$ | $\{l_4, l_5, l_3, l_{10}, l_{15}, l_{16}, l_6, l_7, l_2, l_{20}, l_{14}, l_{16}\}$ | $\{l_2, l_3, l_4\}$ 1 | 1/11 |
| 6 | $C_2(tres(1)) = 0{,}4$ | $\{l_2, l_3, l_{20}, l_1, l_{13}, l_{12}, l_{10}, l_{14}, l_{19}\}$ | $\{l_2, l_3, l_4\}$ 1 | 3/13 |
| 7 | $C_F^{\overset{*}{y}}(x)$ | $\{l_2, l_3, l_4, l_{10}, l_{14}, l_{20}\}$ | $\{l_2, l_3, l_4\}$ 1 | ½ |
| 8 | $C_F(x)$ | $\{l_3\}$ | $\{l_3\}$ , ⅓ | ⅓ |
| 9 | $C_R(x)$ | $\{l_2, l_3, l_4\}$ | $\{l_2, l_3, l_4\}$ 1 | 1 |

Table 3.

1. Acid S., Campos L.M., "A comparison of learning algorithms for Bayesian Networks: a case study based on data from an emergency medical service", Artificial Intelligence in Medicine, 30, 215–232 (2004).
2. Ameljańczyk A., „Metryki Minkowskiego w tworzeniu uniwersalnych algorytmów rankingowych", Biuletyn WAT, Vol. LXIII, Nr 2, 324–336 (2014).
3. Ameljańczyk A., Multipe optimization, WAT, Warszawa, 1986.
4. Ameljańczyk A., Optymalizacja wielo-kryterialna w problemach sterowania i zarządzania, Ossolineum, Wrocław, 1984.
5. Ameljańczyk A., „Metoda podziału zbioru obiektów na wielokryterialne klastry jakościowe", Biuletyn Instytutu Systemów Informatycznych, Nr 12, 1–7 (2013).
6. Ameljańczyk A., „Multicriteria similarity models in medical diagnostics support algorithms", Bio-Algorithms and Med.-Systems, Vol. 21, No.1, 33–39 (2013).
7. Ameljańczyk A., „Wiarygodność komputerowych systemów wspomagania diagnostyki medycznej", w Problemy modelowania i projektowania opartych na wiedzy systemów informatycznych na potrzeby bezpieczeństwa narodowego, 23–39, WAT, Warszawa, 2014.
8. Balcan M.F., Bansal N., Beygelzimer A., Coppersmith D., Langford J., and Sorkin G.B., "Robust reductions from ranking to classification", Machine Learning, 72(1–2):139–153 (2008).
9. Bouyssou D., Marchant T., An axiomatic approach to noncompensatory sorting methods in MCDM, I: The case of two categories, EJOR, 178(1):217–245 (2007).
10. Brans J.P.,Vincke Ph., "A preference ranking organization method: The PROMETHEE method for Multiple Criteria Decision-Making", Management Science, Vol. 31, No. 6, 647–656 (1985).
11. Courtney H., Kirkland J., Viguerie P., "Strategia w warunkach niepewności", w: Zarządzanie w warunkach niepewności, Harvard Busines Review, 316–324, Helion, Gliwice, 2006.
12. Dietterich G.T., "Machine learning research: four current directions", AI Magazine, 18:97–136 (1997).
13. Furnkranz J., Hullermeier E., Mencia E., and Brinker K., "Multilabel classification via calibrated label ranking", Machine Learning, 73:133–153 (2008).
14. Kahneman D., Tversky A., "Prospect theory: An Analysis of Decision Under Risk", Econometrica, Vol. 47, No. 2, 67–72 (1978).
15. Kuncheva L.I., Whitaker C.J., "Measures of diversity in classifier ensembles", Machine Learning, Vol. 51, No. 2, 181–207 (2003).
16. Larose D., Metody i modele eksploracji danych, Wydawnictwo Naukowe PWN, Warszawa, 2008.
17. Pawlak Z., "Rough Sets", International Journal of Computer and Information Sciences, Vol. 11, 341–356 (1965).
18. Pawlak Z., Systemy informacyjne – podstawy teoretyczne, WNT, Warszawa, 1983.
19. Perl J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, Los Angeles, California, 1988.
20. Rasiowa H., Wstęp do matematyki współczesnej, PWN, Warszawa, 2005.
21. Ruiz C., Illustration of the K2 Algorithm for Learning Bayes Net Structures, Department of Computer Science WPI, Bayesian Network Power Constructor, Worcester, MA, 2009.
22. Saaty T.L., Rank from comparisons and from ratings in the analytic hierarchy/network processes, EJOR, 168(2):557–570 (2006).
23. Seo F., Sakawa M., Multiple Criteria Decision Analysis in Regional Planning,
24. D. Reidel-Kluwer, Dordrecht-Boston-Lancaster-Tokyo, 1988.
25. Seung-Seok Choi, Sung-Hyuk Cha, Tappert Charles C., A Survey of Binary Similarity and Distance Measures, Pace University, New York, 2006.
26. Shipp C.A., Kuncheva L.I., "Relationships between combination methods and measures of diversity in combining classifiers", Information Fusion, Vol. 3, No. 2, 135–148 (2002).
27. Yu P.L., Leitmann G., "Compromise solutions, domination structures and Salukwadze's solution", JOTA, Vol. 13, 14–21 (1974).
28. Yu P.L., Leitmann G., "Nondominated decision and cone convexity in dynamic multicriteria decision problems", JOTA Vol.14, 195–203 (1974).
29. Walczak A., Bieniek B., Różyk-Jahnz K., Paluchowska E., "Fuzja klasyfikatorów w diagnostyce chorób skóry", w: Problemy modelowania i projektowania opartych na wiedzy systemów informatycznych na potrzeby bezpieczeństwa narodowego (red. T. Nowicki i Z. Tarapata), 143–152, WAT, Warszawa, 2014.