Machine Vision and Applications (2009) 20:243–259
DOI 10.1007/s00138-007-0120-0

ORIGINAL PAPER

Segmentation and supervised classification of image objects in Epo doping-control

Ivan Bajla · František Rublík · Barbora Arendacká · Igor Farkaš · Klára Hornišová · Svorad Štolc · Viktor Witkovský

Received: 11 October 2006 / Revised: 15 October 2007 / Accepted: 29 November 2007 / Published online: 4 March 2008

Abstract A software system Gel Analysis System for Epo (GASepo) has been developed within an international WADA project. As recent WADA criteria of rEpo positivity are based on identification of each relevant object (band) in Epo images, development of suitable methods of image segmentation and object classification were needed for the GASepo system. In the paper we address two particular problems: segmentation of disrupted bands and classification of the segmented objects into three or two classes. A novel band projection operator is based on convenient object merging measures and their discrimination analysis using specifically generated training set of segmented objects. A weighted ranks classification method is proposed, which is new in the field of image classification. It is based on ranks of the values of a specific criterial function. The weighted ranks classifiers proposed in our paper have been evaluated on real samples of segmented objects of Epo images and compared to three selected well-known classifiers: Fisher linear classifier, Support Vector Machine, and Multilayer Perceptron.

Keywords Image segmentation · Statistical classification · Weighted ranks classifier · Epo doping-control · Epo gel image

The results presented in the paper have been obtained within the project GASepo granted by World Anti-Doping Agency and supported by the Slovak Grant Agency for Science VEGA and by the Slovak Research and Development Agency APVV.

I. Bajla
Austrian Research Centers GmbH—ARC,
Smart Systems Division, 2444 Seibersdorf, Austria
e-mail: ivan.bajla@arcs.ac.at

F. Rublík · B. Arendacká · I. Farkaš · K. Hornišová · S. Štolc (✉) · V. Witkovský
Department of Theoretical Methods,
Institute of Measurement Science, Slovak Academy of Sciences,
Dúbravská cesta 9, 84101 Bratislava, Slovak Republic
e-mail: umerstol@savba.sk

I. Farkaš
Department of Applied Informatics, Comenius University,
Mlynská dolina, 84248 Bratislava, Slovak Republic

1 Introduction

Human erythropoietin (Epo) is a glycoprotein produced mainly in kidneys whenever a tissue oxygen sensor detects oxygen depletion. It is responsible for the proliferation of erythrocytes in the human body. The recent advances in biotechnology and molecular engineering enabled the production of recombinant human erythropoietin (rEpo) by Chinese hamster ovary cells. In 1990s rEpo has been successfully applied to medical treatment of anemia attributable to renal failure. The possibility of relatively cheap production of rEpo has catalyzed its misuse as a doping substance. In endurance sports, the main performance-limiting factor is the oxygen-carrying capacity of the blood. rEpo boosts athletic performance up to 10% by increasing the number of erythrocytes. The International Olympic Committee added rEpo to its "List of prohibited substances" in 1990, though no method existed at that time to detect it in body fluids. Doping with erythropoetic proteins such as epoietin alfa, beta and darbepoietin alfa is one of the most complex issues, faced by sport authorities today.

Previous research has shown that recombinant Epo differs from human Epo in post-translational modifications [5,22,27]. This difference manifests itself in different charge ratios of sugar moieties. It was found that for detection of such small differences, electrophoretic approaches to molecule separation are suitable [8–10,21]. Especially, isoelectric focusing (IEF, [14,15,18,24]) proved to be a

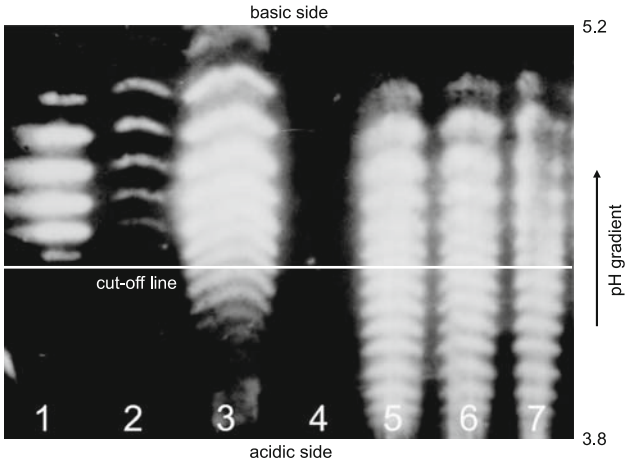

Fig. 1 An example of an Epo gel image

method of choice. However, due to very low concentrations of endogenous and recombinant Epo among all the other proteins, present in human urine, the detection of rEpo is very difficult.

A method for detecting rEpo in urine by electrophoresis was first published in 2001 [13]. Lasne introduced an IEF method coupled with a technique that reduced the non-specific binding that accompanied immunoblotting. This technique (called double-blotting—DB) led to a practical method that is used nowadays to detect epoietin alfa, beta and darbepoietin alfa in the urine of athletes. The process involves separation of Epo isoforms on a polyacrylamide gel followed by the transfer of the proteins onto a thin membrane. Detection of the isoforms is done by a chemiluminescence reaction after incubation of a second antibody, an enzyme catalyst and luminescent reagent. After analog or digital imaging, a typical pattern of lanes (vertical stripes) is finally generated. As can be seen in Fig. 1, the lanes comprise bands (spots of deposits) of individual isoforms, which have been separated by *pH* gradient. When a sample containing rEpo (upper part of the first lane in Fig. 1 with 4–6 bands at different *pH* positions) is subjected to IEF, a shift to more basic isoforms is observed compared to endogenous Epo. When urine with natural Epo is submitted to the same process (lanes 3–7 in Fig. 1), the individual bands (7–15 bands) are separated. These endogeneous bands partially overlap the region of those belonging to rEpo (lanes 3–7 in Fig. 1, in the lane 3 the positive case of doping is detected). The detection of rEpo in presence of endogenous Epo was initially based on setting the reference cut-off line (*col*) and on comparison of intensity sums above *col* in the sample and standard lanes.

Based on a discussion, having been performed in Doping Control (DC) community after the DB technique had been introduced, World Anti-Doping Agency (WADA) elaborated and approved in 2003 an advanced set of criteria for rEpo positivity. These criteria utilize *col* and identification of individual bands, performed on digital Epo images. They are based on measuring intensity characteristics of the bands, evaluation of positions and order of three most intensive bands over the *col* in a sample lane in comparison to three most intensive bands over the *col* in the standard (reference) lane. WADA formulated also a requirement to support the evaluation of the rEpo positivity criteria by a specific and suitable software tools with the following basic functionalities:

- Filtering and correction of defects and geometrical distortions encountered in Epo images,
- identification of individual bands,
- quantification of band properties,
- determination of the reference cut-off-line,
- documentation of the results of image analysis.

A verdict on rEpo positivity can ultimately be approved only by a responsible DC expert, therefore the main purpose of the required software was to enable a user-friendly cooperation in the process of Epo image analysis and to provide the expert with proper visual and quantitative information relevant for the legally significant verdict. To provide a systematic basis for computerized quantitative analysis of Epo images and to contribute to the process of standardization and harmonization of Epo DC within WADA accredited laboratories worldwide, a project granted by WADA has been carried out in the recent years in ARC Seibersdorf research GmbH. As a result of this project a software system GASepo ([4], http://www.antidoping.at/epo) has been developed.

Due to high irregularities of relevant objects and background in Epo images on one hand, and due to WADA requirements to reproducibility of the image analysis results, on the other hand, a number of specific problems of digital image processing and analysis had to be explored. A part of these problems were solved by algorithms which were suitable modifications and adaptations of the algorithms we had developed for another project of DNA gel image analysis [2]. In particular, for noise suppression we implemented a modified GDD (Geometry-Driven Diffusion) filtering [1], for correction of global geometric distortions we used an improved version of image raster rectification [2,4]. Further, we developed an algorithm for robust calculation of the reference cut-off line [3] and an algorithm for correction of background intensity inhomogeneity in each lane [4]. There were two problems that needed special research: (1) band segmentation, and (2) classification of segmented objects into band and artefact classes.

The former operation is aimed at identification of each individual band required by WADA criteria. Due to the above mentioned Epo image characteristic features, the band segmentation represents very difficult problem. Thus, every segmentation algorithm generates, besides true bands, also

artefacts. Presence of artefacts can make visual analysis very tedious. A remedy is offered by introducing an additional band classification operation that makes possible to separate and delete artefacts. So that these two operations constituted an essential part of the GASepo software development. They provide a DC expert with maximum reliable supporting information and minimum interactive load for correcting imperfect analysis results. At the same time, the GASepo is sufficiently flexible to enable incorporation of modified WADA criteria using a future feedback from the Epo doping control practice.

In this paper we outline a compound multilevel segmentation algorithm [19] and we present its improvement. The core of the paper is concentrated on the derivation of a weighted ranks (WR) classification method, novel in the field of image classification. We carry out a detailed comparative study of the proposed classifier to several most popular classifiers when applied to the testing set of segmented objects in Epo images.

The paper is organized as follows. In Sect. 2 we provide a brief review of the band segmentation algorithm and present an improvement of its part responsible for segmenting disrupted bands. In Sect. 3 we characterize statistical properties of input data for classification, and provide the information on training data generation and on evaluation measures. Section 4 is dedicated to theoretical background of the WR classification into two classes, as well as to multisample discriminant procedure. Here the basic information is also given about other classification methods involved in our comparison study. In Sect. 5 the proposed and compared classifiers are applied to two classification tasks relevant for band classification in Epo images. The misclassification errors are evaluated.

2 Band segmentation in Epo images

Computer tests with Epo images showed that in spite of the correction of degradation effects in the preprocessing stage, the application of common edge detectors or adaptive thresholding techniques to band segmentation does not yield satisfactory results. Therefore, in [19] we developed a compound multilevel segmentation method for solving the following crucial cases of band degradation: (1) blurred bands, (2) bands which are merged into a one-blob object, and (3) bands which are represented by separate individual objects. Actual results obtained by the first release of the GASepo software revealed a necessity to improve the important part of the segmentation procedure oriented towards merging disintegrated band parts. We will briefly describe the structure and basic functions of individual elements of the band segmentation. Our attention will then be focused on the improved algorithm.

2.1 Threefold LoG filtering

In general, the bands in Epo images are objects with oblong and curved shape whose boundaries can be characterized as pixels with local maxima of intensity gradients. For segmentation of such objects the rotationally symmetric *Laplacian of Gaussian* (*LoG*) filter is convenient. Due to variable sizes of bands, different levels of their smearing and presence of various local defects in Epo images, the *LoG* operation with a selected single size of the window and a fixed value of the parameter σ yields results which are not satisfactory. It appeared that several different windows and values of σ were needed to be used simultaneously. Based on extensive computer experiments, we proposed three *LoG* filters with the following window sizes and values of σ: $w_1 = 15 \times 5$ ($\sigma_1 = 2$), $w_2 = 15 \times 15$ ($\sigma_2 = 2$), $w_3 = 25 \times 25$ ($\sigma_3 = 4$).

2.2 Thresholding of filtered data

The second step of the proposed segmentation method is a binarization of the filtered lane images $L_i(x, y)$. The experiments on a set of Epo images showed that zero crossing of the results of individual *LoG* applications, that is conventionally used for edge detection, does not provide optimum choice. As intensities inside the bands are relatively homogeneous, instead of zero crossing, we proposed to binarize the filtered results by their thresholding. The thresholded masks $O_i(x, y)$ are given as

$$O_i(x, y) = \begin{cases} 1 & \text{if } L_i(x, y) \le \beta \ , \\ 0 & \text{otherwise.} \end{cases}$$

The choice of the threshold value β, which should be fixed for all Epo images to be analyzed by the GASepo system, is not trivial task. Some threshold values result in producing all true bands accompanied by a great number of artefacts. Decreasing the threshold value reduces the number of these artefacts at the expense of missing true bands. For illustration, an example of thresholding with three different thresholds is given in Fig. 2. As will be discussed later on, for one filtered image we actually need two binary images: one for representation of normal bands, and another one for representation of segments comprising several merged bands. A special operation was therefore designed to properly segment such bands. It is based on this binary representation and uses morphological dilation operation. Based on computer experiments with hundreds of Epo images common in Epo doping-control, we have found the optimum value for the first threshold: $\alpha = -0.001$, and the optimum value for the second threshold: $\beta = -0.0001$.

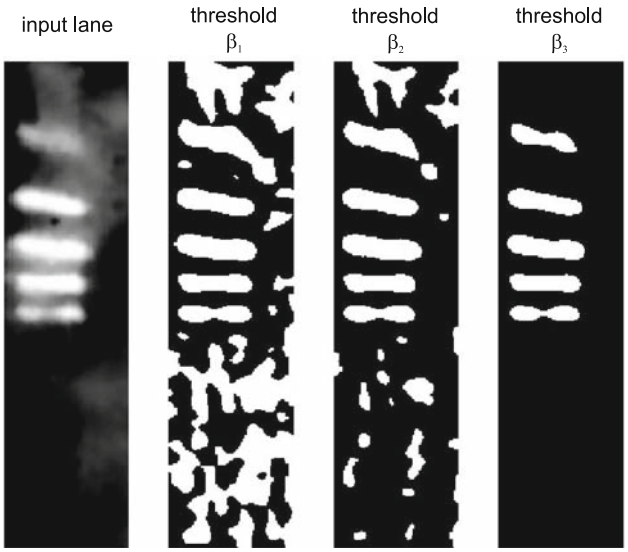

Fig. 2 An example of thresholding a *LoG*-filtered lane image by three different threshold values $\beta_{1,2,3} = 0, -0.0005, -0.002$

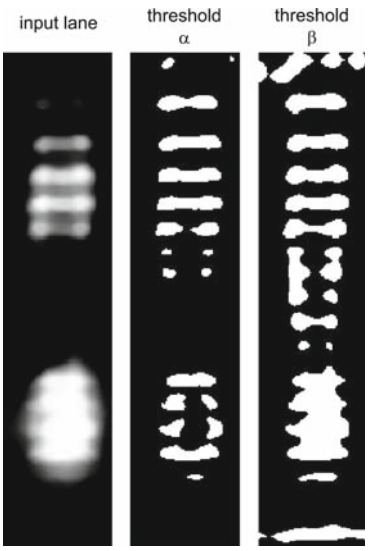

Fig. 3 An example of thresholding by two threshold values ($\alpha = -0.001$, $\beta = -0.0001$) which produces objects represented either by individual bands (threshold α) or by one saturated blob (threshold β)

2.3 Combination of the binary images obtained after thresholding the LoG filter results

Application of *LoG* filters with various values of the parameter σ and window sizes followed by thresholding results in segmented objects (masks O_i) of different area. The overall tendency observed is preferring the objects with larger area than is the area of the corresponding bands. To reflect the differences in band shape and in their intensity characteristics, we proposed to combine the results of the first two steps by the operation of a logical product of the distinct masks O_i : $O(x, y) = \bigwedge_i O_i(x, y)$.

2.4 Region growing of small band masks within larger masks

The application of the threefold *LoG* operation, thresholding, and logical mask combination, to blurred bands enables maximization of the number of true positives that is important for doping positivity criteria. For solving a problem of saturated (glued) bands, we developed another algorithm. Unfortunately, for these bands no optimum threshold value exists that could be applied to each Epo image being analyzed. Smaller values (e.g. $\alpha = -0.001$ in Fig. 3) produce a binary image in which not all constituents of saturated bands are segmented properly. On the other hand, greater threshold values (e.g. $\beta = -0.0001$ in Fig. 3) lead to undersegmentation. The solution consists in using morphological dilation for a specific region growing operation. It starts from smaller (oversegmented) objects—seeds which are embedded into a larger mask. Thus, identity of band parts is preserved and a proper area is achieved by dilation of these objects up to the outer boundary, or up to the next dilated object. The dilation with the 3×3 structure element of ones was used. The merging of separate band segments, that represents the third crucial segmentation problem, is described in the following Section.

2.5 Merging of objects which represent one band

In some cases, the degradation of the gel image may cause that bands in the segmented image are represented by several separated objects. To resolve this particular problem, i.e. to ensure that an identical index is assigned to all objects belonging to one band, we initially developed a special algorithm, called band projection.

We denote P_i the horizontal projection of every region mask R_i of a segmented object in the lane. This mask is equal to 1 if the pixel (x, y) belongs to the ith region, otherwise it is equal to zero. Then the projection P_i of the mask R_i is defined for the range $y \in Y$ of all vertical indices in the lane as the function

$$P_i(y) = \sum_x R_i(x, y). \tag{1}$$

For any two objects obj_i, obj_j, the following subsets of indices y are defined: $\Omega_i = \{y \in Y : P_i(y) \neq 0\}$ and $\Omega_j = \{y \in Y : P_j(y) \neq 0\}$. The intersection of these two index sets, $\Omega_i \cap \Omega_j$, is denoted by Ω_c. For the definition of the criteria for merging of two segmented objects obj_i, obj_j with $\Omega_c \neq \emptyset$, we consider the sum S_c of projections:

$$S_c = \sum_{y \in \Omega_c} min\{P_i(y), P_j(y)\}. \tag{2}$$

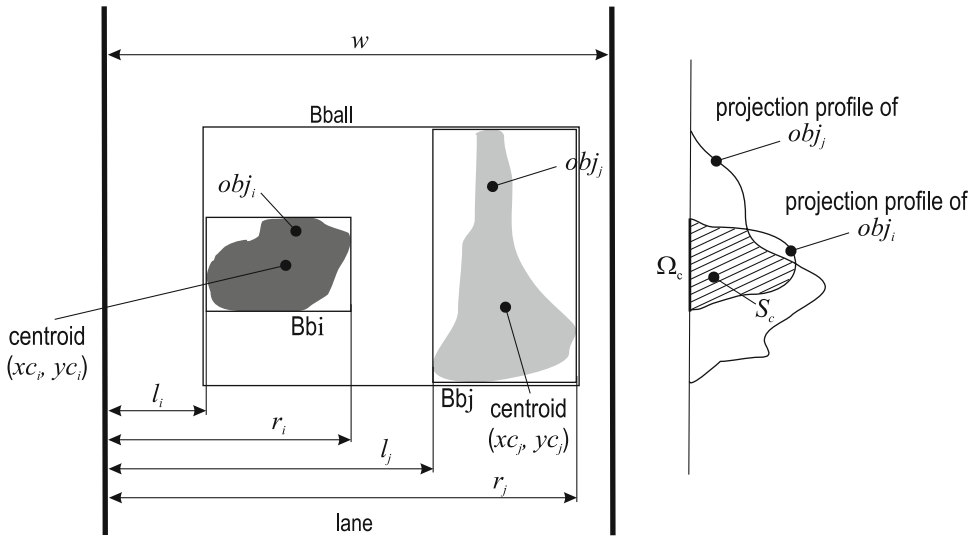

Fig. 4 Quantities defined for merging criteria

Then the coefficients *coef1* and *coef2* are defined by the formulae:

$$coef1 = \frac{S_c}{\sum_{y\in\Omega_i} P_i(y)}, \quad coef2 = \frac{S_c}{\sum_{y\in\Omega_j} P_j(y)}. \tag{3}$$

A simple merging criterion was defined in [19] as: if ($coef1 > 0.5$ & $coef2 > 0.5$) then merge obj_i & obj_j. After the first release of the GASepo software was distributed to Beta testers in DC labs, it appeared that besides properly merged band fragments, a number of artefacts were also merged together or with other band fragments. This resulted in considerable deterioration of classification of the segmented objects and increased the number of necessary user interventions. Since there is a permanent interest to minimize user intervention in the GASepo system, a need to improve the projection operation arose. As will be described in Sect. 3, for the supervised approach we developed a special procedure of training set generation that made use of the already existing GASepo system. A priori information on geometrical properties of bands and artefacts, derived from the training set of segmented objects, served us for extension of the band projection operation. In general, the artefacts can have arbitrary shapes, they can be oblong in horizontal as well as vertical direction; whereas bands have the tendency to be oblong exceptionally in horizontal direction and they differ in their shape only minimally. Their thickness is very close because they actually originated from one ideal band that had been disrupted in the course of Epo image acquisition. The definition of the additional measures for bands and artefacts should take into account these differences.

We introduce the following additional measures of two segmented objects which have non-empty overlap, Ω_c (see Fig. 4):

- $BoxArea_i$, the area of the smallest circumscribed rectangle (band box Bbi) for the object obj_i; similarly $BoxArea_j$ for the object obj_j (band box Bbj),
- $BoxAreaAll$, the area of the smallest rectangle (Bball) circumscribing both objects,
- $Area_i$ and $Area_j$ as number of pixels of the objects obj_i and obj_j, respectively,
- horizontal coordinates xc_i and xc_j of the centroids of both objects,
- distances of object band boxes from the left margin of the lane: l_i, r_i, l_j, r_j (see Fig. 4).

The lane width is denoted as w. Using these measures of objects with non-empty Ω_c, we define four additional coefficients:

$$coef3 = (BoxArea_i + BoxArea_j)/BoxAreaAll,$$
$$coef4 = abs(xc_i - xc_j)/w,$$
$$coef5 = \begin{cases} Area_i/Area_j & \text{if } Area_i < Area_j, \\ Area_j/Area_i & \text{otherwise,} \end{cases}$$
$$coef6 = \begin{cases} abs(l_j - r_i)/w & \text{if } l_i < l_j, \\ abs(l_i - r_j)/w & \text{otherwise.} \end{cases}$$

The ideal goal of the operation of object merging is to allow merging only for segmented objects that are parts of one band. However, due to a great variety of geometrical properties of segmented objects with overlapping projections, this goal is not achievable. We can only try to maximize the number of such cases, and, conversely, to minimize the number of merged artefacts. As a suitable basis for searching optimum threshold values of merging coefficients we used the training set of bands and artefacts. We applied the statistical method of discrimination analysis, see, e.g. [25], to

all introduced coefficients with values calculated separately for the bands and artefacts. First, we select from the training set all object pairs which manifested non-empty overlap Ω_c. This set is then partitioned into two subsets: the set B of all pairs constituted by parts of bands, and the set A of all pairs of artefacts. The values of merging measures are calculated for the sets B and A separately. The histograms of relative frequencies of these values represent empirical probability distributions of the corresponding measures. For the choice of the threshold T_k of the given measure ($coefk, k = 1, \ldots, 6$), the sum of histogram bins corresponding to the measure values $coefk < T_k$ (or $coefk > T_k$) represents the percentage err_a of erroneously merged objects (for artefact pairs) or the percentage err_b of erroneously non-merged (for band pairs) objects. The optimum threshold value T_k is searched for as such a value of $coefk$ for which the weighted sum of both errors reaches its minimum: $w_b.err_b + w_a.err_a = min$, where w_b is the weight of non-merged bands, and w_a is the weight of erroneously merged artefacts. Based on the analysis of the results of the computer experiments with various values of the weights w_b, w_a, we found optimum values $w_b = 0.7$ and $w_a = 0.3$. Having found the optimum threshold values for the merging coefficients, we define the final criterion of merging:

if ($coef1 > 0.7$ and $coef2 > 0.5$ and $coef3 > 0.5$) or
($coef1 > 0.7$ and $coef2 > 0.5$ and
$coef4 > 0.3$ and $coef5 < 0.4$ and $coef6 < 0.2$)
then
merge obj_i & obj_j.

The application of the improved merging part of the compound multilevel band segmentation to a set of Epo images yielded approximately 30% decrease of the number of cases with erroneously merged artefacts. In Fig. 5 the flowchart of the complete algorithm of band segmentation in Epo images is displayed.

3 Characterization of input data set for classification

The improved segmentation procedure described above maximizes the number of obtained segmented true bands. Strong irregularities in Epo image background and intensity fluctuations within bands cause that a number of artefacts and mixed objects (i.e. bands merged with the artefacts) are still generated. In Fig. 6 the result of the band segmentation using GASepo is illustrated. As can be seen, all relevant bands in the lanes have been detected. However, disturbing artefacts appear in crucial areas of the lanes. These artefacts have to be excluded from the subsequent Epo image analysis. To separate them from the set of all segmented objects we have designed a post-segmentation classification procedure that needs quantitative characterization of objects to be classified.

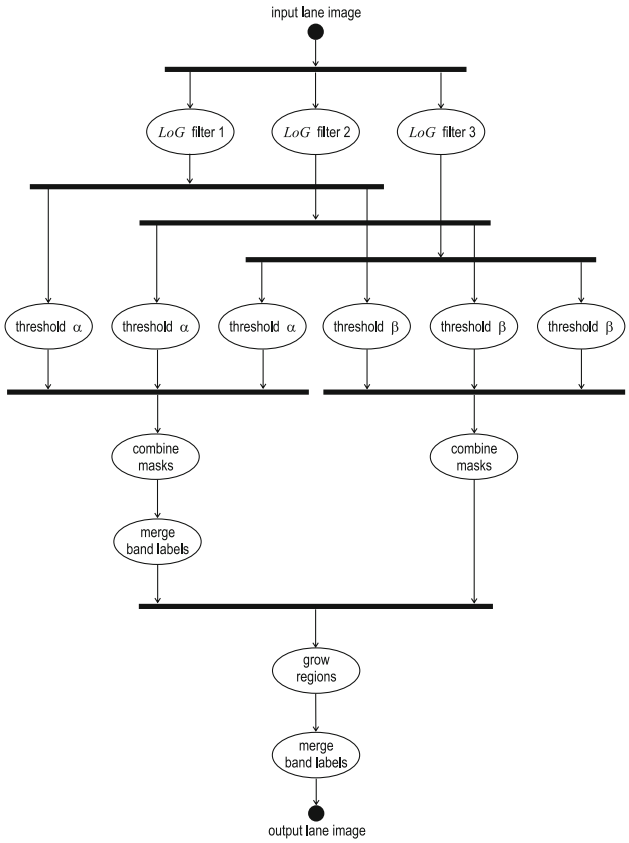

Fig. 5 The scheme of basic segmentation operations

3.1 Quantitative measures of properties of segmented objects

For characterization of geometrical and shape properties of the segmented objects, which are represented as labeled regions in each lane, we have chosen five common measures (implemented, e.g. in the MATLAB) which are used for binary images. Due to specific nature of the bands in Epo images we introduced three additional measures. They use the following quantities:

- *BoundingBox:* the smallest rectangle containing the segmented object.
- *x_width, y_width:* the sizes of the *BoundingBox*.
- *Laxis:* the distance of the central vertical axis of the given lane image from the y-axis.
- *LaneArea:* the number of all pixels in the given lane image.
- *Area:* the number of pixels belonging to the given object.
- *Centroid*$_{x,y}$*:* the coordinates of the center of mass of the given object.

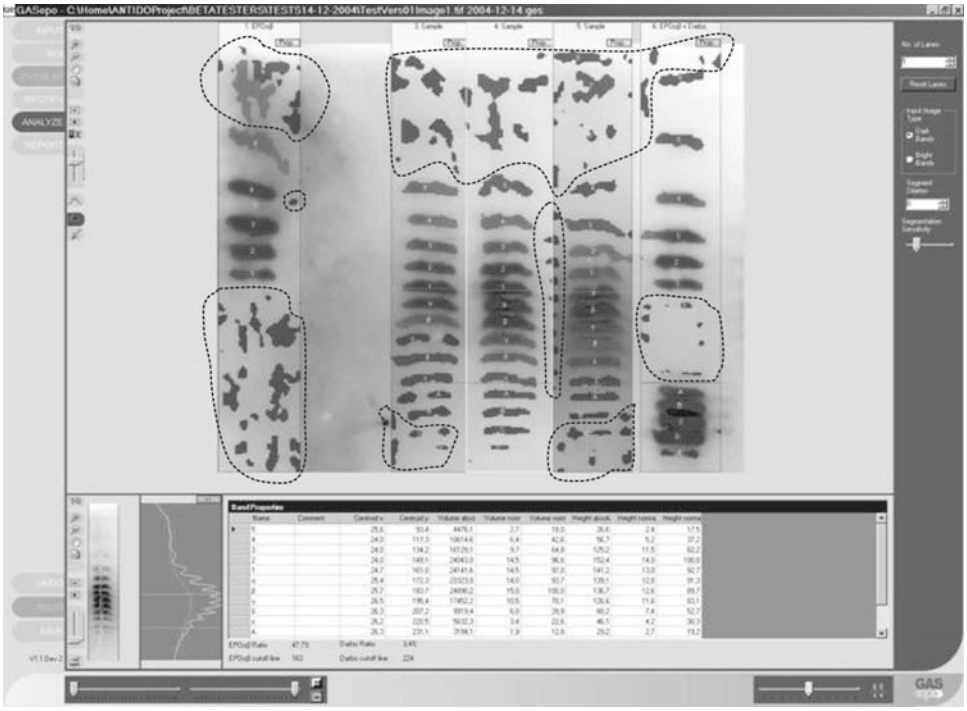

Fig. 6 Illustration of the segmentation results. Groups of the remaining artefacts are marked by *closed dashed curves*

– *Perimeter:* the number of boundary pixels of the given object.

Based on these terms, the following measures of the objects are defined:

– *Relative Area:* $RelArea = Area/LaneArea$.
– *Band Box Ratio:* $BBRatio = y\_width/x\_width$.
– *Eccentricity:* $Eccentr$, the eccentricity of the ellipse that has the same second moments as the object; it is the ratio of the distance between the the foci of the ellipse and its major axis length.
– *Orientation:* $Orientn$, the angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second moments as the object.
– *Solidity:* $Solidit$, the proportion of the pixels in the *BoundingBox*.
– *Centroid Eccentricity:* $CentrEcc = abs(Laxis - Centroid_x)$.
– *Boundary Complexity:* $ObjBndComx = 2 * (x\_width + y\_width)/Perimeter$.
– *Band Box Fullness:* $BBFuls{=}Area/(x\_width * y\_width)$.

3.2 Normality test

Since optimality properties of some classical discrimination procedures are related to the assumption that the observations have normal distribution, we tested the hypothesis that the data we process are normally distributed. Testing the multivariate normality is usually carried out by the Mardia test.

Table 1 Critical constants for the Mardia test and for the Lilieforth test

α	0.1	0.05	0.01
Mardia test	140.2326	146.5674	158.9502
Lilieforth test	0.819	0.895	1.035

Its critical values are given in Table 1. When applied to 4091 artefacts included in the considered training dataset, the Mardia test statistic gives the value $A = 63658$, and when applied to 1789 bands, the value $A = 64383$ is obtained. Thus, in both cases the critical values from Table 1 are clearly exceeded and hence the hypothesis, that the data (either the artefacts or the bands) form a Gaussian population, is rejected.

Even though the data do not come from the multivariate normal distribution, it is still possible that they have some properties of the normal distribution. An important property of the normal distribution is that its coordinates have univariate normal distribution. To find out whether this is true for our observed data, we tested whether single coordinates of the observations follow normal distribution. The test was carried out by means of the Lilieforth test. The critical values of this test are given in Table 1. The observed values exceeded the critical constants from Table 1 and hence the hypothesis that the single quantity has normal distribution is rejected (with the exception of *BBFuls* of bands where $0.6491 < 0.819$). Consequently, employing the above mentioned tests of normality we can reject the hypothesis that the considered training data come from population with multivariate normal distribution. Consequently, it could be expected that

the classical discrimination procedures which have the optimum properties under normality assumptions, like, e.g. the standard Fisher linear classifier, will not reach the optimum properties for the considered data set.

3.3 Generation of training and testing sets

Supervised classification utilizes a priori information represented by the values of features of known objects. As in our case the objects to be classified are generated by the GASepo system, an expert classification should be related to the segmented objects already provided by this system (i.e. using no ideal segmentation made by an expert). We prepared an auxiliary program environment in which the tools available in the GASepo system itself could be used. More than 200 lanes originating from Epo images provided by several DC labs were used for generation of segmented objects. An expert classified obtained objects into three classes: artefacts, bands, and mixed objects. We obtained altogether: 4091 artefacts (labeled by A), 1789 bands (labeled by B), and 97 mixed objects (labeled by C).

To perform correct evaluation of all classifiers on a unified basis (i.e. not admitting any apparent error rate), each of the input data sets of known artefacts, bands, and mixed objects was randomly partitioned into two subsets: 60% of objects in each subset served as a training sample set, and 40% as a testing set. In the evaluation of classifier performance, mean misclassification errors averaged over 100 simulations for each individual classifier were used.

3.4 Misclassification measures

As will be described in the following section we have focused on two basic classification problems: (1) classification of the segmented objects into three classes: A-artefacts, B-bands, and C-mixed objects, and (2) classification of the segmented objects into two classes: artefacts and bands. For characterization of classification success we have used relative and overall misclassification errors and their standard deviations (std) (as error $\pm$ std). The following denotation for relative misclassification errors is used:

- $P(A|A)$—proportion of the artefacts correctly classified as artefacts,
- $P(B|A)$—proportion of the artefacts misclassified as bands,
- $P(C|A)$—proportion of the artefacts misclassified as mixed objects,
- $P(A|B)$—proportion of the bands misclassified as artefacts,
- $P(B|B)$—proportion of the bands correctly classified as bands,
- $P(C|B)$—proportion of the bands misclassified as mixed objects,
- $P(A|C)$—proportion of the mixed objects misclassified as artefacts,
- $P(B|C)$—proportion of the mixed objects misclassified as bands,
- $P(C|C)$—proportion of the mixed objects correctly classified as mixed objects.

We denote the overall misclassification errors as follows:

- $P(error|A)$—overall error of misclassified artefacts,
- $P(error|B)$—overall error of misclassified bands,
- $P(error|C)$—overall error of misclassified mixed objects.
- $P(error)$—overall misclassification error. It is given as

$$P(error) = P(A)P(error|A) + P(B)P(error|B) + P(C)P(error|C),$$

where $P(A)$, $P(B)$, and $P(C)$ represent the probabilities that a randomly chosen object is of type A, B or C, respectively. In practice we estimate these probabilities by relative frequencies of objects of the given class present in the input data set.

4 Classification of segmented objects in Epo images—Theory

We have accomplished a number of tests with segmentation of Epo images generated in practice of DC labs worldwide. Based on the analysis of types of segmented objects obtained we concluded that besides two basic types, i.e. bands and artefacts, there is a certain number of objects which are constituted by artefacts merged with band parts or whole bands. These mixed objects can influence throughput of Epo image analysis by the GASepo system and therefore we were interested if it is possible to design a classifier that could perform an acceptable classification into three given classes of segmented objects. A novel adaptation of the WR statistical approach suitable for application to our task has been developed. Since the original method is formulated for two-class problem of classification, we first provide the basic information on the design of WR classifier for two classes. Then the mathematical results are extended to the design of a multisample discriminant procedure. The performance of the novel classifier applied to the task of band classification in Epo images should be compared to the performance of other most popular classifiers. We have chosen Fisher linear classifier (FLC), the classifier based on support vector machine (SVM), and the neural net classifier based on multilayer perceptron (MLP). The basic information on application of

these classifiers to our problem is also provided in this section. Its structure will preserve the concept of the WR theory development, i.e. after considering the standard cases of two classes, the methods are extended to more classes.

4.1 Weighted ranks classification

The WR classification is a non-parametric procedure of discriminant analysis and the only assumption hidden in its design is that the observed vector data come from the populations with continuous distribution (which can be interpreted as the property that no two observed objects are identical, i.e. they are different in some respect). This procedure has been implemented in accordance with [23].

4.1.1 The weighted ranks procedure for the classification problem into two classes

This procedure is a modification of the method presented in [6], when a recorded observation has to be allocated to one of two populations, in our case either to the class A of artefacts or to the class B of bands. Suppose that n_1 denotes the number of observed artefacts and n_2 the number of observed bands. The procedure consists in adding the tested segmented object to the set of the artefacts, computing values of the discriminant function for all observations from this extended set and then finding the rank R_1 of the value corresponding to the classified object; in the second step the tested segmented object is added to the set of the bands and the rank R_2 is computed. The object is allocated to the population $i = 1$ (artefacts) or $i = 2$ (bands) for which the ratio $w(i)R_i/(1 + n_i)$ is larger; here $w(1)$ is the weight of the first class (artefacts) and $w(2)$ is the weight of the second class (bands). It follows from the comments in [6] that if $w(1) = w(2) = 1$, then procedure of such a type handles the classes in an identical way and therefore the resulting misclassification errors are approximately equal. In order to obtain a smaller misclassification error, we included into the decision process the use of weights. The weights $w(1) = 1$, $w(2) = n_2/(n_1 + n_2)$ were chosen to emphasize the role of larger training set of artefacts. Since $w(2) < 1$, the artefacts will be classified correctly more often and, as $n_1 > n_2$, this will result in a smaller total error. We remark that this procedure is a special $q = 2$ sample case of the general q-sample WR rule, described in detail in the next Section.

4.1.2 The weighted ranks method for multisample discriminant procedure

As mentioned already, within the design of the GASepo system, a classification of segmented objects into three classes was of our interest. To apply the WR classifier to this initial task, it is necessary to extend the principle of the two-sample WR rule to a multisample discriminant version.

Suppose that $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$ is a sample from some distribution Π_i on the p-dimensional Euclidean space R^p, $i = 1, \ldots, q$, and $\mathbf{Z}$ denotes the p-dimensional vector which is subjected to classification, i.e. it has to be assigned to one of the populations $\Pi_1, \ldots, \Pi_q$. Let

$$w = [w(1), \ldots, w(q)]$$

be some suitably chosen positive numbers, which will be referred to as weights. For all $i = 1, \ldots, q$ we construct the criterial function $D_i(\mathbf{U})$ by attaching the classified vector $\mathbf{Z}$ to the ith sample $\mathbf{X}_i$ in the following way. We put

$$\begin{aligned}\tilde{\mathbf{X}}_\mathbf{i} &= \frac{1}{n_i+1}\left(\sum_{j=1}^{n_i}\mathbf{X}_{ij}+\mathbf{Z}\right),\\ \tilde{\mathbf{S}}_\mathbf{i} &= \frac{1}{n_i+1}\left(\sum_{j=1}^{n_i}(\mathbf{X}_{ij}-\tilde{\mathbf{X}}_\mathbf{i})(\mathbf{X}_{ij}-\tilde{\mathbf{X}}_\mathbf{i})'\right.\\ &\quad\left.+(\mathbf{Z}-\tilde{\mathbf{X}}_\mathbf{i})(\mathbf{Z}-\tilde{\mathbf{X}}_\mathbf{i})'\right),\end{aligned} \tag{4}$$

and for all $r = 1, \ldots, q$, $r \neq i$

$$\bar{\mathbf{X}}_\mathbf{r}=\frac{1}{n_r}\sum_{j=1}^{n_r}\mathbf{X}_{rj},\quad \mathbf{S}_r=\frac{1}{n_r}\sum_{j=1}^{n_r}(\mathbf{X}_{rj}-\bar{\mathbf{X}}_\mathbf{r})(\mathbf{X}_{rj}-\bar{\mathbf{X}}_\mathbf{r})'. \tag{5}$$

Further, let

$$\begin{aligned}\tilde{t}_i(\mathbf{U}) &= (\mathbf{U}-\tilde{\mathbf{X}}_\mathbf{i})'(\tilde{\mathbf{S}}_i)^{-1}(\mathbf{U}-\tilde{\mathbf{X}}_\mathbf{i}),\\ M(\mathbf{U}) &= \min_{r=1,\ldots,q,\,r\neq i}(\mathbf{U}-\bar{\mathbf{X}}_\mathbf{r})'(\mathbf{S}_\mathbf{r})^{-1}(\mathbf{U}-\bar{\mathbf{X}}_\mathbf{r}).\end{aligned}$$

Form the criterial function by the formula

$$D_i(\mathbf{U}) = \begin{cases}\log\left(\dfrac{M(\mathbf{U})}{\tilde{t}_i(\mathbf{U})}\right) & \text{if } M(\mathbf{U}) \le \tilde{t}_i(\mathbf{U}),\\ M(\mathbf{U})-\tilde{t}_i(\mathbf{U}) & \text{if } M(\mathbf{U}) > \tilde{t}_i(\mathbf{U}).\end{cases} \tag{6}$$

By means of this function we compute the ranks $R_i(\mathbf{Z})$, $R_i(\mathbf{X}_{i1})$, $R_i(\mathbf{X}_{i2})$, …, $R_i(\mathbf{X}_{in_i})$ of the numbers $D_i(\mathbf{Z})$, $D_i(\mathbf{X}_{i1})$, $D_i(\mathbf{X}_{i2})$, …, $D_i(\mathbf{X}_{in_i})$ in their increasing ordering. The resulting rule is

$$\text{if } w(i)\frac{R_i(\mathbf{Z})}{n_i+1}=\max_{j=1,\ldots,q}\left(w(j)\frac{R_j(\mathbf{Z})}{n_j+1}\right) \text{ decide } \mathbf{Z}\in\Pi_i. \tag{7}$$

In case this is true for more than one i, choose the index i for which the number n_i of observations is the largest. The simulations have been performed for a suitable choice of weights, that is

$$w(i) = \begin{cases}1 & \text{if } \hat{p}_i = \max\limits_{k=1,\ldots,q}\hat{p}_k,\\ \hat{p}_i & \text{otherwise,}\end{cases} \tag{8}$$

which in the 2-sample case coincide with the weights described in the previous section, here $\hat{p}_i = n_i/n$ and $n = \sum_{j=1}^{k} n_j$.

We remark, that for $q = 2$ in [6] the target discriminant function is the Fisher linear discriminant function, which corresponds to the case of equal covariance matrices. In [20] the possible occurrence of different covariance matrices is taken into account by using the target discriminant function which is a convex mixture of the Fisher linear discriminant function and the quadratic discriminant function. In this method we use Mahalanobis distances of the classified vector $\mathbf{Z}$ from the underlying training sets; this is in line with the remark in [6], where it is observed that any continuous target function D can be used as the discriminant function provided that the samples are treated symmetrically and $D(\mathbf{Z}_1), \ldots, D(\mathbf{Z}_n)$ are different with probability 1 for a sample from the continuous distribution on R^p.

For particular GASepo data, the derived multisample discriminat decision rule utilizes three training sets consisting of n_1 objects A, n_2 objects B and n_3 objects C. It consists in carrying out three steps analogous to the previous ones. In the jth step the recorded observation (tested segmented object) is added to the corresponding training set (artefacts, bands, or mixed objects) consisting of n_j objects, then values of the discriminant function are computed for all $n_j + 1$ observations from this extended jth set and the rank R_j of the value corresponding to the classified object is found. The object is allocated to the ith category, if

$$w(i)\frac{R_i}{n_i + 1} = \max_{j=1,2,3} w(j)\frac{R_j}{n_j + 1}.$$

In this decision rule, $w(j)$ is the weight assigned to the jth category, and it should be chosen by the user. In accordance with simulations from various theoretical distributions, we employed weights computed from the relative sample sizes

$$\hat{p}_1 = \frac{n_1}{n}, \quad \hat{p}_2 = \frac{n_2}{n}, \quad \hat{p}_3 = \frac{n_3}{n},$$

where $n = n_1 + n_2 + n_3$ is the total sample size. The weights we have used are defined by the formula

$$w(j) = \begin{cases} 1 & \text{if } \hat{p}_j = \max\{\hat{p}_1, \hat{p}_2, \hat{p}_3\}, \\ \hat{p}_j & \text{otherwise.} \end{cases}$$

This method (with $q = 3$) is a particular case of the general q-sample version of the WR procedure.

4.2 Fisher linear classification

4.2.1 Classification into two classes

The Fisher linear classifier (FLC) is a well-known method based on separating the classes (Ω_0, Ω_1) by a hyperplane $c^T y + d$. We just note here that we use it in a form that does not require the assumption of normality and/or a common covariance matrix in the classes. The Fisher criterion leads to the following discrimination rule, see [25],

$$\begin{aligned} y \in \Omega_0 \text{ if } g(y) &= (m_0 - m_1)^T (1/2(\Sigma_0 + \Sigma_1))^{-1} y > T, \\ y \in \Omega_1 \text{ if } g(y) &= (m_0 - m_1)^T (1/2(\Sigma_0 + \Sigma_1))^{-1} y < T, \end{aligned} \tag{9}$$

where m_0, m_1 and Σ_0, Σ_1 are the means and the covariance matrices of the populations of the two classes, respectively, and the threshold T is appointed to control the misclassification rate of Ω_1 at level α. In practice, $m_0, m_1, \Sigma_0, \Sigma_1$ are replaced by their sample counterparts.

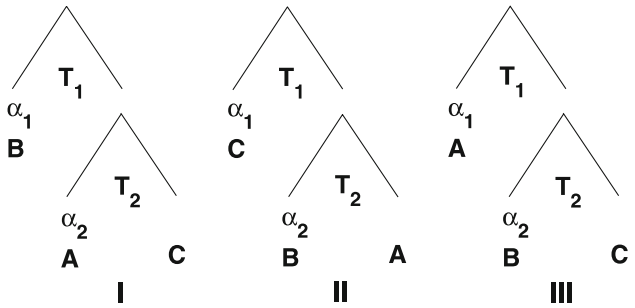

Fig. 7 Three different schemes for a 2-stage classification into three classes using an FLC at each stage

4.2.2 Classification into three classes

The generalization of FLC to a multiclass case (without assuming normality and a common covariance matrix for all classes) can be done through employing a multistage classification that involves distinguishing between only two classes at each stage. Namely, in a 3-class problem with classes denoted A, B, C, 2-stage classification schemes, as depicted in Fig. 7, may be used.

4.3 Classification using the support vector machine

4.3.1 Classification into two classes

The principle of SVM [7,26] for binary classification is to find a hyperplane in the sample space that maximizes the margin, i.e. the distance of the nearest training sample to the hyperplane—equal for both classes. This corresponds to the structural minimization principle that minimizes a trade-off between the empirical error and complexity of class of approximating functions to avoid overfitting. This leads to a quadratic programming problem where the minimized function is an upper bound of the generalization error of a decision hyperplane.

4.3.2 Classification into three classes

The one-against-one method of generalization of binary classification to a multiclass case is used.

4.4 Classification based on a multilayer perceptron

4.4.1 Classification into two classes

The multilayer perceptrons (MLPs) are standard neural network models used for classification tasks. When properly trained, an MLP has a potential to separate classes with non-linear decision boundaries. Unlike classical statistical methods, learning methods for MLPs do not directly rest on assumptions about data (such as normality of input data), [11]. In this application we found it sufficient to employ a perceptron with a single layer of hidden neurons (a two-layer perceptron). For binary classification, we used one logistic output neuron, whose high activation indicates predicting one class and low activation the other class. Learning with logistic outputs is based on minimization of the squared output error.

4.4.2 Classification into three classes

For classification into three classes, we chose three output neurons, each representing one class. The output neurons had the softmax activation function and the rest of the perceptron model remained unchanged. The error to be minimized was based on the cross-entropy between the target distribution and the actual output distribution.

5 Classification of segmented objects in Epo images: evaluation

5.1 Classification into three classes

As outlined in the introduction to the previous section, our primary interest is to explore possibilities of the proposed WR classifier to classify the data of Epo image segmentation into three classes. We are also interested in comparison of its performance to the performance of selected three classifiers. For this purpose a set of real Epo images (see Sect. 3.3) was analyzed by an expert using a preliminary version of the GASepo system. The obtained input data set has the following structure:

$$\begin{array}{lll} \text{population 1:} & \text{artefacts } A & m_1 = 4091 \text{ objects,} \\ \text{population 2:} & \text{bands } B & m_2 = 1789 \text{ objects,} \\ \text{population 3:} & \text{mixed objects } C & m_3 = 97 \text{ objects.} \end{array} \tag{10}$$

Table 2 Average misclassification rates (±standard deviation) and their decomposition using WR method for three classes

$P(error)$	0.054 ± 0.0039	$P(A\|B)$	0.060 ± 0.0087
$P(error\|A)$	0.028 ± 0.0053	$P(B\|B)$	0.934 ± 0.0097
$P(error\|B)$	0.066 ± 0.0097	$P(C\|B)$	0.006 ± 0.0031
$P(error\|C)$	0.932 ± 0.0399	$P(A\|C)$	0.050 ± 0.0351
$P(A\|A)$	0.972 ± 0.0053	$P(B\|C)$	0.882 ± 0.0453
$P(B\|A)$	0.027 ± 0.0051	$P(C\|C)$	0.068 ± 0.0395
$P(C\|A)$	0.001 ± 0.0008		

5.1.1 Weighted ranks approach

Table 2 shows various misclassification rates based on the WR method. While the misclassification rates for classes A and B may be considered acceptable, the misclassification rate for the class C is too high. Practically none of the C-class objects was correctly recognized (the most of the mixed objects have been classified as bands).

5.1.2 Fisher linear classifier

Following Sect. 4.2.2 and the schemes in Fig. 7, in case of the scheme I (the other two cases are similar), for $\Omega_1 = B$ and $\Omega_0 = A \cup C$, function g_1 was constructed with T_1 appointed at level α_1. Next, taking into consideration the data in the training set belonging only to the classes A or C, g_2 was constructed with $\Omega_1 = A$ and $\Omega_0 = C$ and T_2 appointed at level α_2.

Table 3 displays the results. For the scheme I the best result (with regard to the total error) was obtained for $\alpha_1 = 0.08, \alpha_2 = 0.01$. However, while $P(error|A)$ and $P(error|B)$ may be considered acceptable, $P(error|C)$ is too high. In fact, none of the objects from C was correctly recognized. Looking at $P(A|C)$ and $P(B|C)$ we see that the main problem lies in distinguishing the class C from B. In this view, the scheme I is not optimal for managing the proportions of misclassification rates of the classes B and C. This is done easier using the schemes II or III. Table 3 shows that total errors for the scheme II with $\alpha_1 = 0.95, \alpha_2 = 0.08$, and for the scheme III with $\alpha_1 = 0.03, \alpha_2 = 0.01$ are comparable to the ones obtained for the scheme I, $\alpha_1 = 0.08, \alpha_2 = 0.01$. Although $P(error|C)$ is less than 1 in these situations, it is still unacceptably high. As the scheme II allows for a direct control over the proportion of misclassified objects from C, we used it to further illustrate the problems in distinguishing between B and C. Pressing $P(error|C)$ down to 30 % ($\alpha_1 = 0.30$) resulted in a considerable increase of $P(error|B)$ (naturally leading to increase of the total error). From the results obtained we may conclude that the classes B and C are not well discernible.

Table 3 Average misclassification rates(±standard deviation) and their decomposition obtained using the FLC and the 2-stage classification schemes

Scheme/α_1/α_2	I/0.08/0.01	II/0.95/0.08	II/0.30/0.08	III/0.03/0.01
$P(error)$	0.056 ± 0.0032	0.058 ± 0.0035	0.113 ± 0.0080	0.058 ± 0.0038
$P(error\|A)$	0.023 ± 0.0034	0.026 ± 0.0041	0.030 ± 0.0043	0.030 ± 0.0055
$P(error\|B)$	0.081 ± 0.0119	0.082 ± 0.0117	0.289 ± 0.0305	0.081 ± 0.0092
$P(error\|C)$	1.0 ± 0.0	0.952 ± 0.0454	0.338 ± 0.0962	0.835 ± 0.0621
$P(A\|A)$	0.977 ± 0.0034	0.974 ± 0.0041	0.970 ± 0.0043	0.970 ± 0.0055
$P(B\|A)$	0.023 ± 0.0034	0.023 ± 0.0040	0.014 ± 0.0030	0.025 ± 0.0053
$P(C\|A)$	0.0 ± 0.0	0.003 ± 0.0014	0.017 ± 0.0031	0.005 ± 0.0026
$P(A\|B)$	0.081 ± 0.0119	0.081 ± 0.0118	0.080 ± 0.0117	0.072 ± 0.0085
$P(B\|B)$	0.919 ± 0.0119	0.918 ± 0.0117	0.711 ± 0.0305	0.919 ± 0.0092
$P(C\|B)$	0.0 ± 0.0	0.001 ± 0.0012	0.209 ± 0.0291	0.009 ± 0.0041
$P(A\|C)$	0.024 ± 0.0192	0.024 ± 0.0192	0.024 ± 0.0192	0.024 ± 0.0192
$P(B\|C)$	0.976 ± 0.0192	0.928 ± 0.0486	0.315 ± 0.0936	0.811 ± 0.0668
$P(C\|C)$	0.0 ± 0.0	0.048 ± 0.0454	0.662 ± 0.0962	0.165 ± 0.0621

Table 4 Average misclassification rates ($\pm$ standard deviation) and their decomposition using SVM for three classes

$P(error)$	0.045 ± 0.0032	$P(A\|B)$	0.054 ± 0.0078
$P(error\|A)$	0.018 ± 0.0036	$P(B\|B)$	0.946 ± 0.0079
$P(error\|B)$	0.054 ± 0.0079	$P(C\|B)$	0.000 ± 0.0001
$P(error\|C)$	0.985 ± 0.0163	$P(A\|C)$	0.098 ± 0.0444
$P(A\|A)$	0.982 ± 0.0036	$P(B\|C)$	0.886 ± 0.0418
$P(B\|A)$	0.018 ± 0.0034	$P(C\|C)$	0.015 ± 0.0163
$P(C\|A)$	0.001 ± 0.0007		

Table 5 Average misclassification rates ($\pm$ standard deviation) and their decomposition using MLP for three classes

$P(error)$	0.050 ± 0.0036	$P(A\|B)$	0.054 ± 0.0082
$P(error\|A)$	0.026 ± 0.0048	$P(B\|B)$	0.946 ± 0.0082
$P(error\|B)$	0.054 ± 0.0082	$P(C\|B)$	0.000 ± 0.0001
$P(error\|C)$	0.999 ± 0.0036	$P(A\|C)$	0.063 ± 0.0393
$P(A\|A)$	0.974 ± 0.0048	$P(B\|C)$	0.937 ± 0.0397
$P(B\|A)$	0.026 ± 0.0048	$P(C\|C)$	0.001 ± 0.0037
$P(C\|A)$	0.000 ± 0.0000		

5.1.3 Support vector machine classifier

We used the LIBSVM software [7] and its Matlab version. As in [12], the SVM was applied in all cases to data linearly scaled to a common interval. We used the Gaussian kernel. The averaged results of 100 runs of SVM for three classes are displayed in Table 4. The class A, as the most frequent label, is the best recognizable class by SVM method. The objects from the class C with the smallest number of samples are almost unrecognizable by the method and they are predominantly classified as being from the class B.

5.1.4 Multilayer perceptron classifier

We trained the MLP with the scaled conjugate gradients (SCG) algorithm [16] implemented within the Netlab toolbox [17]. All parameters were kept their default values as specified in the toolbox. Polak–Ribiére formula was used for update of (conjugate) search directions. The SCG algorithm usually converged within 100 iterations.

Table 5 shows different misclassification rates for the best case (regarding the total error), i.e. the 8-6-3 perceptron. While the misclassification rates for classes A and B may be considered acceptable, the misclassification rate for the class C is too high, practically none of the C objects was correctly recognized. This is likely due to a very low relative size of class C. Looking at $P(A|C)$ and $P(B|C)$ reveals that most of C patterns (mixed objects) are classified as bands. Since other models suffer from the same problem, this raised the question whether C class should be considered separately.

5.1.5 Discussion

Based on the above results, it is clear that the overall probability of misclassification, $P(error)$, is quite low and acceptable for the given application with its minimum 0.045 ± 0.0032 for the classifier based on the SVM method and its maximum value 0.113 ± 0.008 for the FLC calculated under the scheme II/0.30/0.08. However, all of the considered classifiers failed to correctly classify the objects from the class C (mixed objects). The probability of misclassification of such objects, i.e. $P(error|C)$, is unacceptably high (approaching 100% of all considered cases) with its minimum 0.338 ± 0.04 for the FLC calculated under the scheme II/0.30/0.08 and its maximum 1.0 ± 0.0 for the FLC calculated under the scheme I/0.08/0.01. This result corresponds to our expectation based on preliminary analysis of the conditional distributions of the considered measures. This analysis led us to

Table 6 Average misclassification rates ($\pm$ standard deviation) and their decomposition obtained using WR, FLC with $\alpha = 0.08$, SVM and MLP method with mixed objects eliminated from the training set

	WR	FLC	SVM	MLP
$P(error)$	0.034 ± 0.0033	0.043 ± 0.0040	0.028 ± 0.0030	0.034 ± 0.0038
$P(error\|A)$	0.020 ± 0.0036	0.027 ± 0.0041	0.017 ± 0.0033	0.026 ± 0.0051
$P(error\|B)$	0.066 ± 0.0106	0.078 ± 0.0145	0.052 ± 0.0084	0.053 ± 0.0087
$P(A\|A)$	0.980 ± 0.0036	0.973 ± 0.0041	0.982 ± 0.0033	0.973 ± 0.0051
$P(B\|A)$	0.020 ± 0.0036	0.027 ± 0.0041	0.018 ± 0.0033	0.026 ± 0.0051
$P(A\|B)$	0.066 ± 0.0106	0.078 ± 0.0145	0.052 ± 0.0084	0.053 ± 0.0087
$P(B\|B)$	0.934 ± 0.0106	0.922 ± 0.0145	0.948 ± 0.0084	0.947 ± 0.0087
$P(A\|C)$	0.089 ± 0.0066	0.021 ± 0.0000	0.101 ± 0.0094	0.062 ± 0.0233
$P(B\|C)$	0.911 ± 0.0066	0.979 ± 0.0000	0.899 ± 0.0094	0.938 ± 0.0233

the conclusion that the classification into three classes does not provide desired results and the idea to implement a suitable classifier of this type into the GASepo system has to be abandoned. Consequently, it was necessary to restrict the classification methods to two classes, in particular the artefacts A and the bands B. There are two possibilities how to do this job in accordance with Epo image analysis priority: (1) to simply eliminate the class C from the input data set, or (2) to add the mixed objects to the class B of bands.

5.2 Classification into two classes. CASE 1: the mixed objects are eliminated from the training set

In this particular case the input data set has been modified to have the following structure:

population 1: artefacts A $m_1 = 4091$ objects,
population 2: bands B $m_2 = 1789$ objects.

5.2.1 Weighted ranks approach

In this case a training set consisting of 60% of artefacts and training set consisting of 60% of bands were randomly chosen and all the remaining observations were classified by means of the WR rule. The obtained results are presented in Table 6. We can see that this classification leads to a total error 20% smaller when compared to Table 2.

5.2.2 Fisher linear classifier

Following the previous notation, the FLC was constructed with $\Omega_0 = A$, $\Omega_1 = B$ and different values of α. Table 6 shows the results for the best case (with regard to the total error) achieved, $\alpha = 0.08$. Analogously to the WR method, the total error of classification into two classes decreased.

5.2.3 Support vector machine classifier

The averaged results of 100 runs of SVM for two classes with mixed objects eliminated from the training set are listed in Table 6. Elimination of C objects from the training set leads to decrease of overall error in comparison to the classification into three classes. The percentage of tested C objects classified as B slightly increased.

5.2.4 Multilayer perceptron classifier

Table 6 shows the proportions of different misclassification rates for the best, 8-6-1 case (regarding the total error). It can be seen that, as with other models, the exclusion of mixed objects from training decreased the total error by 1.6%. As with three classes, most of the C objects are classified as bands.

5.3 Classification into two classes. CASE 2: the mixed objects are added to the class of bands in the training set

Adding the mixed objects to the class of bands leads to the following modification of the input data set:

population 1: artefacts A $m_1 = 4091$ objects,
population 2: bands B $m_2 = 1886$ objects.

5.3.1 Weighted ranks approach

The obtained misclassification errors are presented in Table 7. We can see that this yields a slight increase in the percentage of the wrong decision when compared with Table 6.

5.3.2 Fisher linear classifier

Following the previous notation, the FLC was constructed with $\Omega_0 = A$ and $\Omega_1 = B$. Table 7 shows the results for the best case (as to the total error) achieved. Comparison of this Table to Table 6 shows that the inclusion of the objects

Table 7 Average misclassification rates ($\pm$ standard deviation) and their decomposition obtained using WR, FLC with $\alpha = 0.08$, SVM, and MLP method, mixed objects were added to the class of bands in the training set

	WR	FLC	SVM	MLP
$P(error)$	0.035 ± 0.0031	0.043 ± 0.0034	0.030 ± 0.0034	0.034 ± 0.0031
$P(error\|A)$	0.021 ± 0.0039	0.024 ± 0.0038	0.019 ± 0.0037	0.026 ± 0.0049
$P(error\|B)$	0.066 ± 0.0095	0.085 ± 0.0121	0.053 ± 0.0080	0.053 ± 0.0083
$P(A\|A)$	0.979 ± 0.0039	0.976 ± 0.0038	0.981 ± 0.0037	0.974 ± 0.0049
$P(B\|A)$	0.021 ± 0.0039	0.024 ± 0.0038	0.019 ± 0.0037	0.026 ± 0.0049
$P(A\|B)$	0.066 ± 0.0095	0.085 ± 0.0121	0.053 ± 0.0080	0.053 ± 0.0083
$P(B\|B)$	0.934 ± 0.0095	0.915 ± 0.0121	0.947 ± 0.0080	0.947 ± 0.0083
$P(A\|C)$	0.080 ± 0.0431	0.024 ± 0.0192	0.086 ± 0.0437	0.056 ± 0.0389
$P(B\|C)$	0.920 ± 0.0439	0.976 ± 0.0192	0.914 ± 0.0437	0.944 ± 0.0389

from C into B in the training set did not much influence the results.

5.3.3 Support vector machine classifier

The averaged results of 100 runs of SVM for two classes with mixed objects added to the class of bands are given in Table 7. There are no significant differences in comparison to Table 6.

5.3.4 Classifier based on multilayer perceptron

Table 7 shows the proportions of different misclassification rates for the best, 8-6-1 case. These results show that the alternative training strategy with two classes leads to practically identical results as in the CASE 1.

5.3.5 Discussion

Based on the misclassification errors of all compared classifiers related to the CASE 1 (Table 6) and CASE 2 (Table 7), the following conclusions can be drawn:

- The modification of the training procedure, either by elimination of the mixed objects or by adding them to the class B, yields comparable misclassification errors; the first way is recommended for its simplicity.
- Solving the modified classification task for two classes (artefacts and bands) by the novel WR classifier yields the second least value of the overall misclassification error (0.034 for the CASE 1).
- In both cases the mixed objects are mostly classified as band objects, so that the final interactive separation of band parts from the artefacts can be confined to the class of the bands.

5.4 An illustration of the segmentation and classification results in the final documentation of rEpo positivity tests performed by the GASepo system

In Fig. 8 two copies of an Epo image are documented together with the intensity profiles and tables of numerical characteristics of the incriminated lanes (N.4 of positive case, N.2 of negative case). The result of band segmentation operation is apparent from labeled bands. The result of classification operation is articulated implicitly, i.e. all segmented objects classified as artefacts have been automatically removed from the visual presentation. Two cut-off-lines are depicted (in each copy):—the upper one for epoietin alfa,—the lower one for darbepoietin alfa. In the positive case, detected in the lane N.4, the crucial bands which are located above the *col* (marked by an ellipse) represent the most intensive bands above the *col* within this lane (see a dashed rectangle in the top table). On the contrary, in the negative case, detected in the lane N.2, the intensity characteristics of the bands, located above the *col* (marked by another ellipse) are dispersed throughout the ordered values given in the bottom table.

6 Conclusion

In 2004 an international project of the Gel Analysis System for Epo doping-control (GASepo) was approved by World Anti-Doping Agency. The goal of this project is to develop a software system that could be used as a standardized objective tool for Epo image analysis carried out in everyday practice of WADA-accredited doping control laboratories. During the development of this system [4] we encountered two main problems: object segmentation in digital Epo images, and classification of segmented objects.

In the paper the basic approach to the first problem is described that represents a compound multilevel procedure including several image processing algorithms. We have

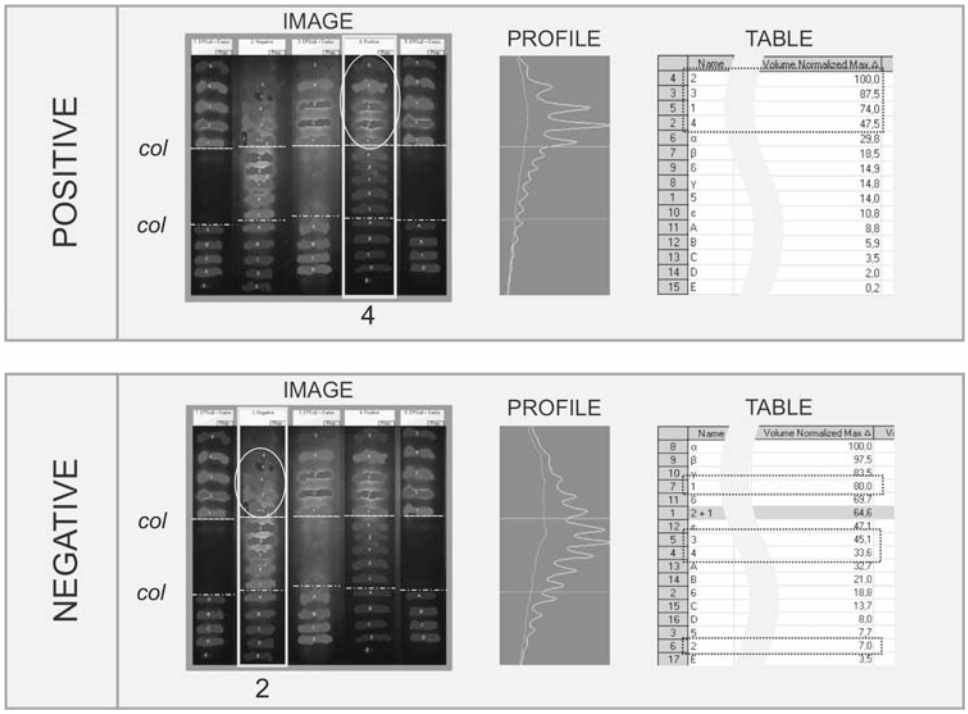

Fig. 8 An illustration of the positive and negative results of the rEpo positivity test performed via the GASepo system in which band segmentation and classification operations have been incorporated

concentrated on an improvement of the important band projection operator that is dedicated to resolve frequent problems with disrupted bands. The novelty of this approach can be characterized by the following points.

- Four additional measures of disrupted band parts and artefacts with the goal to increase the number of correctly merged band parts and to decrease the number of improperly merged artefact parts.
- The tools incorporated in the GASepo system for user interactive analysis were utilized for expert classification of segmented objects into several object classes.
- The obtained expert knowledge enabled us to carry out discrimination analysis for each of six measures using the histograms of their values calculated separately for the artefacts and bands.
- Optimum thresholds of the object merging measures have been found and incorporated into the final decision rule; the results for the given set of Epo images which served for the generation of the input dataset showed 30% improvement in merging separate parts of the individual bands.

The proposed improved algorithm has been implemented in the GASepo software.

In Sect. 4.1 a WR classification method, novel in the field of image classification, is introduced. It is based on ranks of the values of criterial function that involves Mahalanobis distances generated by the given data populations. The weights are included into the decision process with the aim to achieve a smaller total error. In the case of two classes the rule is a modification of the Broffitt–Randless–Hogg rule. The generalization to its multiclass counterpart is proposed. The weights should be chosen by the user. The particular choice yielding good results is found. The both WR approaches are suitable for any classification tasks beyond the application area discussed in the paper, especially if the normality assumptions are critically violated.

The WR classifiers proposed in our paper have been evaluated on real samples of segmented objects of Epo images. They have been compared to three selected well known classifiers: Fisher linear classifier, Support Vector Machine, and Multilayer Perceptron.

For the initial problem of three classes (artefacts, bands, and mixed objects in segmented Epo images), it turned out that no classifier is able to discriminate the class of mixed objects satisfactorily. Therefore, we have reduced the initial problem to classification into two basic classes of artefacts and bands. Two possibilities of doing so have been explored. We have shown that for all classifiers, no significant differences between these two approaches were obtained. The WR classifier reached misclassification errors smaller than in the case of other classifiers, except vector support machine. However, it should be emphasized, that the WR classifier does not require incorporation of any additional program package in the GASepo software. As majority of operations in GASepo have to be performed in real time, such a software solution would cause undesired extension of the Epo image analysis.

References

1. Bajla, I., Holländer, I.: Nonlinear filtering of magnetic resonance tomograms by geometry-driven diffusion. Mach. Vis. Appl. **10**, 243–255 (2001)
2. Bajla, I., Holländer, I., Fluch, S., Burg, K., Kollár, M.: An alternative method for electrophoretic gel image analysis in the GelMaster software. Comput. Methods Programs Biomed. **77**, 209–231 (2005)
3. Bajla, I., Holländer, I., Gmeiner, G., Reichel, C.: Quantitative analysis of images in erythropoietin doping-control. Med. Biol. Eng. Comput. **43**, 403–409 (2005)
4. Bajla, I., Holländer, I., Minichmayr, M., Gmeiner, G., Reichel, C.: GASepo—a software solution for quantitative analysis of digital images in Epo doping-control. Comput. Methods Programs Biomed. **80**(3), 246–270 (2005). doi:10.1016/j.cmpb.2005.09.005
5. Berglund, B., Ekblom, B.: Effect of recombinant human erythropoietin treatment on blood pressure and some haematological parameters in healthy men. J. Intern. Med. **229**, 125–130 (1991)
6. Broffitt, J., Randles, R., Hogg, R.: Distribution-free partial discriminant analysis. J. Am. Stat. Assoc. **71**, 934–939 (1976)
7. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. National Taiwan University, Taipei. http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.ps.gz (2001). Software: http://www.csie.ntu.edu.tw/~cjlin/libsvm
8. Görg, A., Postel, W., Günther, S.: The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis **9**(9), 531–552 (1988)
9. Häder, D.: Image Analysis, Methods and Applications, 2nd edn. CRC Press, Boca Raton, London, New York (2001)
10. Horgan, G., Glasbey, C.: Uses of digital image analysis in electrophoresis. Electrophoresis **16**, 298–309 (1995)
11. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**, 359–366 (1989)
12. Hsu, C.W., Chang, C.C., Lin, C.J.: A Practical Guide to Support Vector Classification. National Taiwan University, Taipei (2001)
13. Lasne, F.: Double-blotting: a solution to the problem of non-specific binding of secondary antibodies in immunoblotting procedures. J. Immunol. Methods **253**, 125–131 (2001)
14. Lasne, F., de Ceaurriz, J.: Recombinant erythropoietin in urine. Nature **405**, 635 (2000)
15. Lasne, F., Martin, L., Crepin, N., de Ceaurizz, J.: Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones. Anal. Biochem. **311**, 119–126 (2002)
16. Møller, M.: A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. **6**, 525–533 (1993)
17. Nabney, I., Bishop, C.: Netlab toolbox (release 3). Available at http://www.ncrg.aston.ac.uk/netlab/index.php. Neural Computing Research Group, Aston University, Birmingham (2003)
18. Parisotto, R., Wu, M., Ashenden, M., Emslie, K., Gore, C., Howe, C.: Detection of recombinant human erythropoietin abuse in athletes utilizing markers of altered erythropoiesis.. Haematologica **86**(2), 128–137 (2001)
19. Ramoser, H., Biber, J., Bajla, I., Holländer, I.: Segmentation of electrophoretic images in doping-control. In: Proc. of the Intern. Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, METMBS'04, pp. 467–470. CSREA Press, Las Vegas, USA (2004)
20. Randles, R., Broffitt, J., Ramberg, J., Hogg, R.: Discriminant analysis based on ranks. J. Am. Stat. Assoc. **73**, 379–384 (1978)
21. Richards, P.: Protein electrophoresis. In: Rapley, R., Walker, J. (eds.) Handbook of Molecular Biomethods, pp. 413–462. Humana Press, Totowa, NJ (1998)
22. Roberts, D., Smith, D.: Erythropoietin: induction of synthesis to signal transduction, protein electrophoresis. J. Mol. Endocrinol. **12**, 131–148 (1994)
23. Rublík, F.: Modification of the two-sample discrimination rule of Broffitt, Randless and Hogg and its extension to the multisample case. Technical Report. Institute of Measurement Science of the Slovak Academy of Sciences, Bratislava (2006)
24. Stübiger, G., Marchetti, M., Reichel, C., Gmeiner, G., Allmaier, G.: Towards the identification/verification of recombinant Epo applied in doping using matrix-assisted laser desorption/ionization mass spectrometry. In: Proc. 22nd Manfred Donike Workshop on Dope Analysis 2004 (2004)
25. Therrien, C.: Decision, Estimation and Classification: An Introduction to Pattern Recognition and Related Topics, 1st edn. Wiley, New York (1989)
26. Vapnik, V.: Statistical learning theory, 1st edn. Wiley, New York (1998)
27. Wide, L., Bengtsson, C., Berglund, B., Ekblom, B.: Detection in blood and urine of recombinant erythropoietin administered to healthy men. Med. Sci. Sports Exerc. **27**, 1569–1576 (1995)

Author biographies

Ivan Bajla received Ph.D. degree in 1977 in measurement science (biomeasurement and image processing) from the Institute of Measurement Theory of the Slovak Academy of Sciences in Bratislava, Slovak Republic. His research interests include gray-scale and local transformations of digital images, geometry-driven diffusion methods of image filtering, image segmentation methods in the field of MRI and electrophoretic gel image analysis. Since 1998 he has been with the High Performance Image Processing Department of the ARC Seibersdorf research GmbH. He teaches image processing courses at the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava.

František Rublík works at the Institute of Measurement Science of the Slovak Academy of Sciences in Bratislava. Graduated at the Comenius University in Bratislava in 1970, degree RNDr received at the Comenius University in Bratislava in 1973, Ph.D. in 1977 at the mathematical Institute of the Academy of Sciences of the Czech republic in Prague. Research interests: Asymptotic problems of mathematical statistics. Published works include papers on asymptotic properties of likelihood ratio tests and of rank tests, and also papers on goodness-of-fit testing and on characteristic roots and eigenvectors of covariance matrices without the assumption of normality of distribution.

Barbora Arendacká graduated at Comenius University in Bratislava (specialization: Mathematical Statistics and Theory of Probability) and became a Ph.D. student at the Institute of Measurement Science, Slovak Academy of Sciences in Bratislava. Her research interests comprise mainly mixed linear models and generalized inference.

Igor Farkaš received his MSc (Technical Cybernetics, 1991) and his Ph.D. (Applied Informatics, 1995) from Slovak University of Technology in Bratislava, Slovakia. He is affiliated in part with the Institute of Measurement Science, Slovak Academy of Sciences in Bratislava. Since 2003 he has also been with the Department of Applied Informatics, Comenius University in Bratislava, where he became an Associate Professor in 2007. In 2000-2003 he worked as a research fellow at the Department of Psychology, University of Richmond, VA. He was a Fulbright fellow (in 1998) at the Department of Computer Sciences, University of Texas in Austin, and a Humboldt fellow (in 2005) at the Department of Computational Linguistics and Phonetics, Saarland University, Germany. His research areas include neural networks, self-organization and cognitive science with focus on connectionist natural language modeling.

Klára Hornišová received Ph.D. degree in 2005 in mathematical statistics from the Faculty of Mathematics, Physics, and Informatics, Comenius University, Bratislava. Her research interest is non-linear regression. Since 2003 she has been with the Institute of Measurement Science of Slovak Academy of Sciences.

Svorad Štolc is now pursuing Ph.D. degree at the Institute of Measurement Science of Slovak Academy of Sciences. He studied artificial intelligence and parallel computing at Faculty of Mathematics, Physics and Informatics of Comenius University in Bratislava, Slovakia and he obtained his master's degree in 2001. His research interest areas are image processing, feature selection and object classification.

Viktor Witkovský works as the researcher at the Institute of Measurement Science of the Slovak Academy of Sciences in Bratislava, and the head of the Department of Theoretical Methods. Graduated at the Comenius University in Bratislava in 1986, he received his Ph.D. degree in 1993 at the Mathematical Institute of the Slovak Academy of Sciences in probability and mathematical statistics. Research interests: mixed linear models, variance components estimation, probability distributions for small sample inference, statistical methods and algorithms for evaluation of measurements in engineering and bioengineering.