Multiclassifier Systems for Sequential Recognition Based on the Concept of Meta-Bayes Classifier

MAREK KURZYNSKI
marek.kurzynski@pwr.edu.pl

MARCIN MAJAK
marcin.majak@pwr.edu.pl

KAROL PUCHALA
karol.puchala@pwr.edu.pl

Wroclaw University of Technology, Department of Systems and Computer Networks
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, POLAND

Abstract: In the sequential recognition there exist dependencies among the successive objects to be classified. In this study two original multiclassifier (MC) systems for the sequential recognition are developed. In the first MC systems base classifiers are defined for particular steps of sequential recognition independently, whereas in the second MC system base classifiers classify an object at the current step on the base of its features and features of previous objects. Both MC systems in combining procedure uses original concept of meta-Bayes classifier and produces decision according to the Bayes rule. The performance of both MC systems were evaluated experimentally and compared with six state-of-the-art sequential recognition methods using computer generated data. Results obtained in experiments imply that MC system is effective approach, which improves recognition accuracy in sequential decision scheme.

Key–Words: Multiclassifier system, Sequential recognition, Probabilistic model, Meta-Bayes classifier

1 Introduction

In many practical pattern recognition problems there exist dependencies among the objects to be recognized. Sequential medical diagnosis [7], recognition of patient's intent in the control of bioprosthetic hand [17] and words recognition [16] can be cited here as typical examples of such situation. In the first problem, which consists in multiple recognition of the patient's state based on results of successive examinations, the current state depends on the previous states and additionally on the applied treatment. In the next example, classification scheme is based on decomposition of hand movement on a sequence of elementary actions with Markov model of dependencies. In the last problem, the dependence between letters in the word results from the statistic properties of letter succession of the language. The dependence between classes of successive objects can be of a diversified nature and range. Its simplest instance can be a one-instant-backwards dependence to so complex arrangements as those in which the current class depends on classes of the all previous objects.

When we intend to support sequential recognition task using a computer, we have to take into account these sequential dependencies. In other words, when constructing an appropriate decision algorithm (classifier) we must not limit our approach to only the narrow information channel that concerns just the current features, but we have to consider all the available measurement data instead, as they may contain important information about the object at a given step. For the purpose of automatic recognition based on sequential decision scheme, several algorithms have been developed and proposed in the literature [13], [14], [22]. For the calculation of decision strategy, various mathematical models are used in algorithms, such as probabilistic model with Markov dependence [14], fuzzy relation approach [8], rough set theory [9], neural network method [7], among others.

In the last two decades, the multiclassifier (MC) systems are very strongly developed, mostly because of the fact that committee, also known as ensemble, can outperform its members [5], [6], [15], [18], [19]. For the classifier combination two main approaches used are classifiers fusion and classifiers selection. In the first method, all classifiers in the ensemble contribute to the decision of the MC system, e.g. through sum or majority voting [6]. In the second approach, a single classifier is selected from the ensemble and its decision is treated as the decision of the MC system. The selection of classifiers can be either static or dynamic. In the static selection scheme classifier is selected for all test objects, whereas dynamic classifier selection approach explores the use of different classifiers for different test objects [1].

In this study, two original multiclassifier (MC) systems for sequential recognition problem are developed. In the first MC systems base classifiers are defined for particular instants independently, whereas in the second MC system base classifiers classify an object at the current instant on the base of its feature vector and feature vectors of previous objects. In both MC systems the original concept of meta-Bayes classifier is applied for combining decisions of base classifiers. In the meta-Bayes classifier, which creates a probabilistic generalization of a set of base classifiers for particular steps of sequential recognition, first *a posteriori* probabilities for the set of classes are calculated and next decision is made according to the Bayes rule. In the proposed MC systems, combining mechanisms are constructed using the supervised learning procedure. It means, that so-called validation set must be available, which is basis for calculating in dynamic fashion class-dependent probabilities of correct classification and misclassification for base classifiers.

The paper arrangement is as follows. Section 2 presents fundamentals of sequential recognition and introduces basic notations. Two developed MC systems are described in detail in Section 3. The experiments conducted and the results with discussion are presented in Section 4. The paper is concluded in Section 5.

2 Sequential Recognition

We will treat sequential recognition as a discrete dynamical process in which classes of successive objects denote its states. This process is at the n-th instant in the state $j_n \in \mathcal{M}$ (object at the n-th instant belongs to the class j_n), where $\mathcal{M}$ is an M-element set of possible states (classes) numbered with the successive natural numbers. Thus:

$$j_n \in \mathcal{M} = \{1, 2, 3, \ldots, M\}. \quad (1)$$

Obviously, the notion of *instant* has no specific temporal meaning here, as its interpretation depends on the practical character of the case under consideration. The actual measure used may be minutes, hours, days, or even weeks.

The state j_n is unknown and does not undergo our direct observation. What we can only observe is the feature vector by which a state (a class) manifests itself. We will denote a d-dimensional feature vector by $x_n \in \mathcal{X} \subseteq \Re^d$, for an object at the n-th instant (thus $\mathcal{X}$ is the symptom space).

Since class of the current object depends on history, the specificity of the investigated classification task reveals in the form of input data, which are not associated only with the features of the current object, but comprise up to an extend the historic information that regards the preceding course of recognition process. For this case we do not know how far backwards the examined input data should spread into the past; the "the more the better" rule need not necessarily be true here. As far now, there are no analytical evidence to be used in this issue, whilst any attempts to answer the question are under way of experimental research [7].

In the general case, we suppose that the decision algorithm at the n-th instant takes into account the K-instant-backwards-dependence ($K < n$). It means, that decision at the n-th instant is made on the base of vector of features

$$\bar{x}_n^{(K)} = (x_{n-K}, x_{n-K+1}, \ldots, x_{n-1}, x_n). \quad (2)$$

In consequence, the classification algorithm at the n-th instant is of the following form:

$$\Psi_n(\bar{x}_n^{(K)}) = i_n, \ i_n \in \mathcal{M}. \quad (3)$$

In this study, multiclassifier systems will be applied as classifiers (3) for the particular instances of sequential recognition. In the proposed MC systems, both the pool of base classifiers and the combining mechanism will be constructed using the supervised learning procedure, what leads to the assumption that a learning set $\mathcal{S}$ and a validation set $\mathcal{V}$ are available [6]. In the considered sequential decision problem, the learning set $\mathcal{S}$ consists of m training sequences:

$$\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_m\}, \quad (4)$$

where a single sequence

$$\mathcal{S}_k = ((x_{1,k}, j_{1,k}), (x_{2,k}, j_{2,k}), \ldots, (x_{N,k}, j_{N,k})) \quad (5)$$

denotes a sequence of features and classes of learning objects (in the sequential medical diagnosis the sequence $\mathcal{S}_k$ refers to a single patient and it contains a sequence of states that occured at the successive moments and a corresponding sequence of the observed examination results).

Similarly, the validation set $\mathcal{V}$ consists of r validation sequences $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_r\}$ and a single sequence $\mathcal{V}_k$ has the same form as in (5). The next section, describes the procedure of determining the original MC systems (3) using learning set $\mathcal{S}$ and validation set $\mathcal{V}$, in detail.

3 Multiclassifier Systems

3.1 Preliminaries

The proposed multiclassifier systems are built as a combination of the two following probabilistic paradigms:

Markov Model

We suppose that classes of objects in successive steps $j_1, j_2, \ldots, j_N$ are observed values of sequence of random variables $\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_N$ modeled by first-order Markov chain. The probabilistic formalism for such dependence is given by the vector of initial probabilities

$$p = [p_{j_1}]_{1\times M}, \text{ where } p_{j_1} = P(\mathbf{J}_1 = j_1) \tag{6}$$

and by the matrix of transition probabilities (we suppose that Markov chain is homogeneous and stationary):

$$P = [p_{j_n, j_{n-1}}]_{M\times M} \tag{7}$$

where $p_{j_n, j_{n-1}} = P(\mathbf{J}_n = j_n | \mathbf{J}_{n-1} = j_{n-1})$.

Meta Bayes Classifier

In the concept of Meta Bayes Classifier (MBC), which originally was introduced in [11], [12] we suppose that a base classifier ψ is given, which maps feature space into a set of class numbers, viz.

$$\psi : \mathcal{X} \longrightarrow \mathcal{M}. \tag{8}$$

The MBC ψ^{MBC} constitutes the specific probabilistic generalization of base classifier (8) which has the form of the Bayes scheme built over the classifier ψ. This means, that ψ^{MBC} takes the decision according to the maximum *a posteriori* probability rule:

$$\psi^{MBC}(\psi(x) = k) = i \longleftrightarrow$$

$$P(i|\psi = k) = \max_{l\in\mathcal{M}} P(l|\psi = k). \tag{9}$$

3.2 Multiclassifier System 1 (MC1)

Suppose first, that we have the set of N trained base classifiers:

$$\psi_1(x_1), \psi_2(x_2), \ldots, \psi_N(x_N), \tag{10}$$

which classify objects at the 1-st, 2-nd, …, N-th instant, respectively.

The MC1 systemn (3) for n-th instant is defined as the MBC classifier (9) constructed over the set of base classifiers (10) for n-th, $(n-1)$-th, …,$(n-K)$-th instants, namely:

$$\Psi_n^{MC1}(\bar{x}_n^{(K)}) = \psi^{MBC}(\psi_{n-K}(x_{n-K}) = i'_{n-K}, \ldots,$$

$$\psi_{n-1}(x_{n-1}) = i'_{n-1}, \psi_n(x_n) = i'_n). \tag{11}$$

The MC1 system (11) produces the decision about class of the object at the n-th instant according to the generalized rule (9):

$$\Psi_n^{MC1}(\bar{x}_n^{(K)}) = i_n \longleftrightarrow$$

$$P(i_n|\psi_{n-K}(x_{n-K}) = i'_{n-K}, \ldots, \psi_n(x_n) = i'_n) =$$

$$= \max_{l\in\mathcal{M}} P(l|\psi_{n-K}(x_{n-K}) = i'_{n-K}, \ldots, \psi_n(x_n) = i'_n), \tag{12}$$

where:

$$P(i_n|\psi_{n-K} = i'_{n-K}, \ldots, \psi_n = i'_n) =$$

$$= \frac{P(i_n, \psi_{n-K} = i'_{N-k}, \ldots, \psi_n = i'_n)}{P(\psi_{n-K} = i'_{N-k}, \ldots, \psi_n = i'_n)}. \tag{13}$$

Since denominator in (13) has no influence on the classification result of algorithm (12), classifying function of (12) reduces to the nominator, which – assuming that base classifiers (10) are conditionally independent – after simple calculations has the following form:

$$P(i_n, \psi_{n-K} = i'_{n-K}, \ldots, \psi_n = i'_n) =$$

$$P(\psi_n = i'_n|i_n) \sum_{j_{n-1}} P(\psi_{n-1} = i'_{n-1}|j_{n-1}) p_{i_n i j_{n-1}} \times$$

$$\times \sum_{j_{n-2}} P(\psi_{n-2} = i'_{n-2}|j_{n-2}) p_{j_{n-1} j_{n-2}} \times \ldots$$

$$\times \sum_{j_{n-K}} P(\psi_{n-K} = i'_{n-K}|j_{n-K}) p(j_{n-K}). \tag{14}$$

The key element in the algorthm (14) presented above is the calculation of probabilities $P(\psi_n = i_n|j_n)$, i.e. class-dependent probabilities of correct classification and misclassification for base classifiers (10).

The proposed method of evaluation of these probabilities is based on the original concept of a hypothetical classifier called Randomized Reference Classifier (RRC) [18], [19]. The RRC is a stochastic classifier defined by a probability distribution which is chosen in such a way, that RRC acts, on average, as an modeled base classifier. It means, that RRC can be considered equivalent to the modeled base classifier, and therefore it is justified to use the class-dependent probabilities of correct classification (misclassification) of RRC as appropriate probabilities for the evaluated base classifier. In the computational procedure, first these probabilities are calculated for validation points and then they are generalized on the whole feature space. Details of the method can be found in [18]. Furthermore, the Matlab code for calculating class-dependent probabilities of correct classification (misclassification) of RRC was developed and it is freely available for download [20].

Similarly, initial (6) and transition (7) probabilities in (14) are estimated using validation set $\mathcal{V}$.

3.3 Multiclassifier System 2 (MC2)

In the MC2 system we assume, that for the n-th instant we have a set of (K+1) trained base classifiers:

$$\psi_n(x_n), \psi_{n-1}(x_{n-1}), \ldots, \psi_{n-K}(x_{n-K}), \quad (15)$$

which classify the state at the n-th instant on the base of different feature vectors.

The ensemble (15) creates the algorithm (3) at the n-th instant using – as previously – the MBC for combining of base classifiers, namely:

$$\Psi_n^{MC2}(\bar{x}_n^{(K)}) = \psi^{MBC}(\psi_n(x_n) = i_n^{(0)},$$

$$\psi_{n-1}(x_{n-1}) = i_n^{(1)}, \ldots, \psi_n(x_K) = i_n^{(K)}). \quad (16)$$

This means, that classifier (16), i.e. the MC2 system classifies the object at the n-th instant according to the maximum of *a posteriori* probability rule:

$$\Psi_n^{MC2}(\bar{x}_n^{(K)}) = i_n \longleftrightarrow$$

$$P(i_n|\psi_n(x_n) = i_n^{(0)}, \ldots, \psi_{n-K}(x_{n-k}) = i_n^{(K)}) =$$

$$= \max_{l \in \mathcal{M}} P(l|\psi_n(x_n) = i_n^{(0)}, \ldots, \psi_{n-K}(x_{n-k}) = i_n^{(K)}). \quad (17)$$

From the Bayes rule and assuming that base classifiers make decisions independently, we get:

$$\Psi_n^{MC2}(\bar{x}_n^{(K)}) = i_n \longleftrightarrow$$

$$p(i_n) \prod_{k=0}^{K} P(\psi_{n-k}(x_{n-k}) = i_n^{(k)}|i_n) =$$

$$= \max_{l \in \mathcal{M}} p(l) \prod_{k=0}^{K} P(\psi_{n-k}(x_{n-k}) = i_n^{(k)}|l). \quad (18)$$

Probabilities $P(\psi_n = i_n|j_n)$, i.e. class-dependent probabilities of correct classification and misclassification for base classifiers (15) can be calculated as in the MC1 system using validation set $\mathcal{V}$ and the concept of RRC classifier. Similarly, *a priori* probabilities $p(i_n)$ in (18) are estimated using validation set.

4 Experimental Investigations

4.1 Experimental Setup

In order to evaluate the performance of the proposed multiclassifier systems and to compare with state-of-the-art sequential recognition methods, several experiments were made on the computer generated data. The experiments were conducted in MATLAB using PRTools 4.1 [4]. In the recognition process two-class problem was considered with the following Gaussian class-dependent probability density functions of scalar feature x:

$$f_1(x) = N(0,1), f_2(x) = N(1,1). \quad (19)$$

In experiments, three different matrices of transition probabilities of the first-order Markov chain were adopted:

Experiment A

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \quad (20)$$

Experiment B

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix} \quad (21)$$

Experiment C

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad (22)$$

Our choice was deliberate one and results from the fact, that for two-state (two-class) Markov chain the parameter $\alpha = p_{1,1} + p_{2,2}$ represents the strength of the dependencies between Markov chain states (classes): $\alpha = 0(2)$ denotes the deterministic dependence between states (classes), whereas $\alpha = 1$ denotes the case where states (classes) are totally independent [14]. It means, that in the Experiment A classes are strongly dependent, in the Experiment B this dependence is moderate, and in the Experiment C classes are independent.

In each experiment the set of 1000 objects was generated according to the adopted probability distributions (19) – (22). The training and testing sets were extracted from each dataset using two-fold cross-validation method. For combining the MC systems, a two-fold stacked generalization technique [21] was used.

The experiments were conducted using three different recognition algorithms as base classifiers (10) and (15):

1. (L) Linear classifier based on normal distribution with the same covariance matrix for each class [3];
2. (3) 3-nearest neighbours classifier;
3. (N) feed-forward back-propagation neural network with 1 hidden layer.

The performace of the proposed MC systems for K=1 and K=2 in the sequential scheme was compared against the following six state-of-the-art sequential classifiers:

- The probabilistic algorithm based on the first (second) order Markov dependence (MV1, MV2) [10];
- The fuzzy algorithm based on the Mamdani inference scheme with 1- (2-) instant-backward-dependence (MM1, MM2) [17];
- The fuzzy algorithm based on the fuzzy relation with 1- (2-)instant-backward-dependence (FR1, FR2) [8].

4.2 Results and Discussion

Classification accuracies (i.e. the percentage of correctly classified objects) for methods tested are listed in Table 1. The accuracies are average values obtained over 10 runs (5 replications of two-fold cross validation). Statistical differences between the performances of the MC systems and the six sequential classification methods were evaluated using F test [2]. The level of $p < 0.05$ was considered statistically significant. In Table 1, statistically significant differences are given under the classification accuracies as indices of the method evaluated, e.g. for the Experiment A, MC1(L)1 system produced statistically better classification accuracies from the MM1, FR1 and FR2 methods.

These results imply the following conclusions:

1. The MC systems produced statistically significant higher scores in 119 out of 216 pairwise tests (3 experiments $\times$ 6 classifiers compared $\times$ 12 MC systems);
2. The MC1 system with two-instant-backwards-dependence and with ANN base classifiers (MC1(N)2) achieved the highest overall classification accuracy averaged over all experiments - it outperformed the MV1, MV2, MM1, MM2, FR1, FR2 systems by 1.6%, 0.7%, 6.5% , 2.2%, 13.1%, 12.4%, respectively. This results confirm the effectiveness of the use the multiclassifier system in the sequential recognition;
3. There occurs a common effect within each classifier (MC system) type: one-instant-backwards-dependence is always worse than two-instant-backwards-dependence.
4. When the strength of the dependencies between Markov chain states (classes) increases then the accuracy of all methods investigated also increases.

Table 1: Classification accuracies of classifiers compared in the experiment (description in the text). The best score for each dataset is highlighted.

		Experiment/ Mean accuracy [%]			
No	Classifier	A	B	C	Mean
1	MC1(L)1	93.5 15,17,18	93.2 15,17,18	91.5 15,17,18	92.7
2	MC1(3)1	93.9 15,16,17,18	92.8 15,17,18	91.1 15,17,18	92.6
3	MC1(N)1	94.3 15,16,17,18	**93.7** 13,15,16,17,18	92.1 15,16,17,18	93.4
4	MC1(L)2	93.9 15,17,18	93.5 15,16,17,18	91.6 15,17,18	93.0
5	MC1(3)2	94.5 15,16,17,18	93.2 15,17,18	91.0 15,17,18	92.9
6	MC1(N)2	**95.9** 15,16,17,18	93.4 15,16,17,18	**92.3** 15,16,17,18	**93.9**
7	MC2(L)1	92.7 15,17,18	91.8 15,17,18	91.2 15,17,18	91.9
8	MC2(3)1	92.9 15,17,18	92.0 15,17,18	91.1 15,17,18	92.0
9	MC2(N)1	93.2 15,17,18	92.5 15,17,18	90.7 15,17,18	92.1
10	MC2(L)2	93.1 15,17,18	92.2 15,17,18	91.1 15,17,18	92.1
11	MC2(3)2	93.5 15,17,18	92.8 15,17,18	91.0 15,17,18	92.4
12	MC2(N)2	94.1 15,17,18	93.2 15,16,17,18	90.9 15,17,18	92.7
13	MV1	93.7	92.4	90.7	92.3
14	MV2	94.5	93.6	91.5	93.2
15	MM1	88.9	87.1	86.2	87,4
16	MM2	92.8	91.7	90.5	91.7
17	FR1	82.5	80.8	79.1	80.8
18	FR2	83.7	81.2	79.6	81.5

5 Conclusion

Nowadays, many researchers have been focused on MC systems and consequently, many new solutions have been dedicated to different recognition approaches. In this study we have focused on sequential recognition for which two original MC systems have been proposed. The proposed methods differ with the form of base classifiers, whereas in both MC systems responses of base classifiers are combined using probabilistic model and meta-Bayes classifier scheme. Experimental results clearly showed that the idea of meta-Bayes classifier and a common probabilistic model of base classifiers is correct and leads to the accurate and efficient multiclassifier systems for sequential recognition.

Acknowledgements: This work was supported by the statutory funds no S50020 of the Department of Systems and Computer Networks, Wroclaw University of Technology.

References:

[1] A. Britto, R.Sabourin, et al., Dynamic selection of classifiers a comprehensive review. *Pattern Recognition*, 47, 2014, pp. 3665–3680.

[2] T. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation* 10, 1988, pp. 1895–1923.

[3] R. Duda, P. Hart, D. Stork, *Pattern Classification*, John Wiley and Sons, New York, 2000.

[4] R. Duin, P. Juszczak et al., PRTools4. A Matlab Toolbox for Pattern Recognition, Delft University of Technology, 2007.

[5] H. Ko, R. Sabourin, A. Britto, From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, 41, 2008, pp. 1718-1731.

[6] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.

[7] M. Kurzynski, Benchmark of approaches to sequential diagnosis, *Artificial neural networks in biomedicine*, Lisboa P, Ifeachor E, eds. Springer, London, 2000, pp. 129–140.

[8] M. Kurzynski, A. Zolnierek, Computer-aided sequential diagnosis using fuzzy relations - comparative analysis of methods, *Lecture Notes in Bioinformatics* 3745, 2005, pp. 242–251.

[9] M. Kurzynski, A. Zolnierek, A. Wolczowski, Control of bio-prosthetic hand via sequential recognition of EMG signals using rough sets, (in) *Computer Recognition Systems 3*, Springer Verlag, Berlin Heidelberg, 2009, pp. 455–462.

[10] M. Kurzynski, E. Puchala, Recognition of EMG signals based on the first-order Markov: model applied to the control of bioprosthetic hand, *Proceedings of the 5th WSEAS International Conference on Sensors and Signals* (SENSIG '12), Malta, Sept. 2012, pp. 113–118.

[11] M. Kurzynski, M. Majak, A. Zolnierek, Multiclassifier systems applied to the computer-aided sequential medical diagnosis, *Biocybernetics and Biomedical Engineering* (to appear).

[12] M. Kurzynski, M. Majak, Meta-Bayes classifier with Markov model applied to the control of bioprosthetic hand, Czarnowski I., Mateos A.,et al.(eds.), *Intelligent Decision Technologies, Smart Innovation, Systems and Technologies*, vol. 56, 2016, pp. 107–117

[13] U. Marti, H. Bunke, Using a statistical language model to improve the performance of an HMM-based cursive handwritting recognition system. *Int. Journ. of Pattern Recognition and Artificial Intelligence*, 15, 2001, pp. 65–90.

[14] J. Raviv, Decision making in Markov chains applied to the problem of pattern recognition, *IEEE Trans. on Inf. Theory*, 13, 1967, pp. 536–551.

[15] L. Rokach, Ensemble-based classifiers, *Artificial Intelligence Review* 33, 2010, pp. 1–39.

[16] J. Sas, M. Kurzynski, Application of statistic properties of letter succession in Polish language to handprint recognition, *Computer Recognition Systems*, Proc. IV Int. Conference, Springer Verlag, 2005, pp.206–212

[17] A. Wolczowski, M. Kurzynski, Human – machine interface in bio-prosthesis control using EMG signal classification, *Expert Systems* 27, 2010, pp. 53–70.

[18] T. Woloszynski, M. Kurzynski, A probabilistic model of classifier competence for dynamic ensemble selection, *Pattern Recognition* 44, 2011, pp. 2656–2668.

[19] T. Woloszynski, M. Kurzynski et al., A measure of competence based on random classification for dynamic ensemble selection, *Information Fusion* 13, 2012, pp. 207–213.

[20] T. Woloszynski, Classifier competence based on probabilistic modeling (ccprmod.m) at Matlab central file exchange. URL: http://www.mathworks.com/matlabcentral/fileexchange/28391- a-probabilistic-model-of-classifier-competence.

[21] D. Wolpert, Stacked generalization, *Neural Networks* 5, 1992, pp. 214–259.

[22] M. Wozniak, Markov chains pattern recognition approach applied to the medical diagnosis tasks, (in) *Biological and Medical Data Analysis*, LNBI 3745, Springer Verlag, 2005, pp. 231–241.