Biomedical Images Classification by Universal Nearest Neighbours Classifier Using Posterior Probability

Roberto d'Ambrosio, Wafa Bel Haj Ali, Richard Nock, Paolo Soda, Franck Nielsen, Michel Barlaud

Biomedical Images Classification by Universal Nearest Neighbours Classifier Using Posterior Probability

Roberto D'Ambrosio[1,3], Wafa Bel Haj Ali[3], Richard Nock[2], Paolo Soda[1], Frank Nielsen[4], and Michel Barlaud[3,5]

[1] Universita' Campus Bio-Medico di Roma, Rome, Italy
{r.dambrosio,p.soda}@unicampus.it
[2] CEREGMIA - Université Antilles-Guyane, Martinique, France
rnock@martinique.univ-ag.fr
[3] CNRS - U.Nice, France
{belhajal,barlaud,dambrosi}@i3s.unice.fr
[4] Sony Computer Science Laboratories, Inc., Tokyo, Japan
Frank.Nielsen@acm.org
[5] Institut Universitaire de France

Abstract. Universal Nearest Neighbours (UNN) is a classifier recently proposed, which can also effectively estimates the posterior probability of each classification act. This algorithm, intrinsically binary, requires the use of a decomposition method to cope with multiclass problems, thus reducing their complexity in less complex binary subtasks. Then, a reconstruction rule provides the final classification. In this paper we show that the application of UNN algorithm in conjunction with a reconstruction rule based on the posterior probabilities provides a classification scheme robust among different biomedical image datasets. To this aim, we compare UNN performance with those achieved by Support Vector Machine with two different kernels and by a k Nearest Neighbours classifier, and applying two different reconstruction rules for each of the aforementioned classification paradigms. The results on one private and five public biomedical datasets show satisfactory performance.

1 Introduction

Humans are limited in their ability to distinguish similar objects and to diagnose diseases during image interpretation because of noise and of their non-systematic search patterns. In addition, the vast amount of image data generated by imaging devices makes the detection of potential diseases a burdensome task, may reduce the reproducibility and may cause oversight errors. In biomedical imaging, developments in computer vision, machine learning as well as artificial intelligence have shown that automatic or semi-automatic image analysis may support the physicians in different medical fields, overcoming most of the above limitations.

In this paper we focus on the challenging task of classifying biomedical images. Indeed, developing one classifier architecture with robust and satisfactory

F. Wang et al. (Eds.): MLMI 2012, LNCS 7588, pp. 119–127, 2012.

performance over different biomedical image datasets is still an open issue. Main difficulties are related to several factors, e.g. high variability of images belonging to different fields, the number of available images, the type of descriptors, etc..

Recently, some of the authors of this paper have proposed the Universal Nearest Neighbours (UNN) classifier [1]. This algorithm, intrinsically binary, requires the use of a decomposition method to cope with multiclass problems reducing these tasks into several binary subtasks. Then, a *reconstruction rule* provides the final classification [2–4]. Furthermore, it was proven that UNN classifier can effectively estimates the posterior probability of each classification act [5]. This permits us to use this information to apply reconstruction rules potentially more effective than others criteria proposed in the literature, which set the final decision using the crisp labels provided by the binary learners.

The contribution of this paper is the proposal of a new classification scheme combining UNN algorithm with a reconstruction rule based on posterior probability. To proof this claim, we have performed several tests on six different biomedical image datasets comparing UNN performance with those achieved by Support Vector Machine (SVM) with two different kernels and by a k Nearest Neighbours (kNN) classifier. Moreover, the tests were conducted applying two different reconstruction rules for each of the aforementioned classification paradigms.

2 Methods

This section first presents decomposition schemes used by UNN to address multiclass classification tasks and, second, it introduces the UNN itself.

2.1 Decomposition Methods

A classification task consists in assigning to sample $x \in \mathbb{R}^n$ a label representative of one class belonging to a set $\Omega : \{\omega_1, \omega_2, \ldots, \omega_C\}$. When we are facing with a multiclass problem, we can make use of a decomposition approach reducing problem complexity in less complex binary subtasks and recombining binary classifiers outputs through a reconstruction rule.

Among the several decomposition rules proposed in literature [2, 3] we investigate the One-per-Class (OpC) decomposition method that reduces the original problem into C binary problems each one addressed by a dichotomizer M_c. We say that M_c is specialized in the cth class when it aims at recognizing if the sample x belongs either to the cth class or, alternatively, to any other class. Its output is 1 if $x \in \omega_c$, otherwise it is -1. For each sample x, the crisp outputs of dichotomizers are collected into the *binary profile*: $\boldsymbol{M}(x) = [M_1(x), M_2(x), \ldots, M_C(x)]$. Furthermore, dichotomizers may supply other information typically related to the degree that the sample belongs (or does not belong) to the corresponding class. Such information is collected in a *reliability profile*, $\boldsymbol{\psi}(x) = [\psi_1(x), \psi_2(x), \ldots, \psi_C(x)]$, whose elements measure the classification reliability on pattern x provided by each dichotomizer. Each entry varies in the interval [0, 1], and a value close to 1 indicates a very reliable classification.

Table 1. The three UNN loss functions, their corresponding solutions δ_j of eq. (4) and their corresponding weights w_i updating eq. (5)

Loss function(ψ)	W_j^+	W_j^-	δ_j in eq. (4)	w_i in eq. (5)
$\exp(-x)$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}>0} w_i$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}<0} w_i$	$\frac{1}{2}log(\frac{W_j^+}{W_j^-})$	$w_i exp(-\delta_j y_{ic} y_{jc})$
$\log_2(1+\exp(-x))$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}>0} w_i$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}<0} w_i$	$log(\frac{W_j^+}{W_j^-})$	$\frac{w_i exp(-\delta_j y_{ic} y_{jc})}{1-w_i(1+exp(-\delta_j y_{ic} y_{jc}))}$
$-x+\sqrt{1+x^2}$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}>0} w_i$	$\sum_{i:j\sim_k i,\ y_{ic}y_{jc}<0} w_i$	$\frac{2W_j^+ - 1}{2\sqrt{W_j^+ W_j^-}}$	$1 - \frac{1-w_i+\sqrt{w_i(2-w_i)}\delta_j y_{ic} y_{jc}}{\sqrt{1+\delta_j^2 w_i(2-w_i)+2(1-w_i)\sqrt{w_i(2-w_i)}\delta_{jc} y_{ic} y_{jc}}}$

Table 2. UNN loss function and posterior probability estimators ($\hat{p}_c(x)$)

Loss function name	Loss function(ψ)	$\hat{p}_c(x)$	Acronym
exponential	$\exp(-x)$	$(1+\exp(-2h_c(x)))^{-1}$	UNN(exp)
logistic	$\log_2(1+\exp(-x))$	$(1+\exp(-h_c(x)))^{-1}$	UNN(log)
Matsushita	$-x+\sqrt{1+x^2}$	$\frac{1}{2}(1+\frac{h_c(x)}{\sqrt{1+h_c(x)^2}})$	UNN(Mat)

The reconstruction rules may use both binary and reliability profiles to set the final decision. We present now two reconstruction rules for OpC: the first is a traditional implementation referred to as *Hamming decoding* (H_d), whereas the second is referred to as MDS rule and it has been introduced in [4]. In case of H_d, the final decision is given by $O(x) = \omega_s$, with:

$$s = argmin_i d_H(\mathbf{D}(\omega_i), \mathbf{M}(x)) \tag{1}$$

where

$$d_H(\mathbf{D}(\omega_i), \mathbf{M}(x)) = \sum_{c=1}^{C} (\frac{1-(D(\omega_i,c)M_c(x))}{2}) \quad ; \quad D(\omega_i, c) = \begin{cases} 1 & if\ i = c \\ -1 & if\ i \neq c \end{cases} \tag{2}$$

where s denotes the index of the dichotomizer setting the final output $O(x) \in \Omega$. In case of MDS, we have:

$$s = \begin{cases} \arg\max_c (M_c(x) \cdot \psi_c(x)), & \text{if } m \in [1, C] \\ \arg\min_c (-M_c(x) \cdot \psi_c(x)), & \text{if } m = 0 \end{cases} \tag{3}$$

where m is defined as $m = \sum_{c=1}^{C} [M_c(x) = 1]$, and square brackets denote the indicator function.

2.2 Universal Nearest Neighbours

Universal Nearest Neighbour (UNN) is a supervised learning algorithm that induces a leveraged kNN rule by globally minimizing a surrogate loss function in a boosting framework [1, 6–8].

Algorithm 1. Algorithm UNIVERSAL NEAREST NEIGHBORS, UNN($\mathcal{S}, \psi, k$)

Input: $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, 2, ..., m,\ \boldsymbol{x_i} \in \mathbb{R}^n,\ \boldsymbol{y_i} \in \{-1, 1\}^C\}$, ψ strictly convex loss, $k \in \mathbb{N}_*$;

Let $\boldsymbol{\alpha}_j \leftarrow \mathbf{0}, \forall j = 1, 2, ..., m$;

for $c = 1, 2, ..., C$ **do**

  Let $\boldsymbol{w} \leftarrow -\nabla_\psi(0)/m$;

  for $t = 1, 2, ..., T$ **do**

    [I.0] Let $j \leftarrow$ WIC($\mathcal{S}, \boldsymbol{w}$);

    [I.1] Let $\delta_j \in \mathbb{R}$ solution of:

$$\sum_{i: j \sim_k i} y_{ic} y_{jc} \nabla_\psi \left(\delta_j y_{ic} y_{jc} + \nabla_\psi^{-1}(-w_i) \right) = 0 \ ; \tag{4}$$

    [I.2] $\forall i : j \sim_k i$, let

$$w_i \leftarrow -\nabla_\psi \left(\delta_j y_{ic} y_{jc} + \nabla_\psi^{-1}(-w_i) \right) \ , \tag{5}$$

    $\forall i = 1, 2, .., m$, let

$$w_i \leftarrow \frac{w_i}{\sum_{h=1}^m w_h} \ , \tag{6}$$

    [I.3] Let $\alpha_{jc} \leftarrow \alpha_{jc} + \delta_j$;

Output: $h_c(x_q) = \sum_{j \sim_k \boldsymbol{q}} \alpha_{jc} y_{jc}$; $\forall c = 1, 2, \cdots, C$

Let denote by $\mathcal{S} = \{(x_i, \boldsymbol{y}_i), i = 1, 2, ..., m,\ x_i \in \mathbb{R}^n,\ \boldsymbol{y_i} \in \{-1, 1\}^C\}$ the training set. According with *OpC* decomposition scheme the problem is reduced into C binary classification tasks with corresponding sets of samples $\mathcal{S}^{(c)} = \{(\boldsymbol{x}_i, y_{ic}), i = 1, 2, ..., m\}$. The vector of labels $\boldsymbol{y}_i \in \{-1, 1\}^C$ encodes class memberships, assuming $y_{ic} = 1$ iff $\boldsymbol{x}_i$ belongs to class c and $y_{ic} = -1$ otherwise. For each problems, we learn a classifier $h_c : \mathbb{R}^n \to \mathbb{R}$ by minimizing a *surrogate risk* over $\mathcal{S}^{(c)}$ [9, 7, 8]. A surrogate risk, considered as the actual missclassification rate of h_c on the training data $\mathcal{S}$, has the following general expression:

$$\varepsilon_{\mathcal{S}}^{\psi}(h_c) \doteq \frac{1}{m} \sum_{i=1}^{m} \psi(y_{ic} h_c(x_i)) \ , \tag{7}$$

for some function ψ that we call a *surrogate loss*. Quantity $y_{ic} h_c(x_i) \in \mathbb{R}$ is called the multiclass *edge* of classifier h_c on example (x_i, y_{ic}).

Let $NN_k(x)$ be the set of the k nearest neighbours ($k \in \mathbb{N}^*$) of an example x. Then, the UNN classification rule , introduced in [1], is expressed as the following *weighted* kNN voting rule:

$$h_c(x) = \sum_{j \in NN_k(x)} \alpha_{jc} y_{jc} \tag{8}$$

where $\alpha_{jc} \in \mathbb{R}$ is the leveraging coefficient for example j in class c, with $j = \{1, 2, \ldots, m\}$ and $c = \{1, 2, \ldots, C\}$. Note that those coefficients are the solution of minimising the *surrogate risk* in eq. (7). Hence, eq. (8) linearly combines class labels of the k nearest with their leveraging coefficients. Eventually, one leverage coefficient (α_{jc}) per class is learned for each weak classifier (Alg.1)

We report in this paper three versions of UNN with the following losses: exponential, logistic and Matsushita, detailed in Table 1 and Table 2. For each of those functions we give the corresponding expression of δ_j, approximation to the solution of eq. (4), and w_i in eq. (5) in Table 1.

UNN may supply information related to the degree that the test sample x belongs (or not) to the corresponding class. Thus, we estimate the posterior probability ($\hat{p}_c(x)$) of decision taken by UNN on a query sample x. We report estimator's formal definition for each loss function of UNN in Table 2. The theorical approach for deriving $\hat{p}_c(x)$ from $h_c(x)$ is fully given in [5].

3 Datasets

Table 3. Summary of the used datasets

Dataset	Samples	Classes	Majority class	Minority class	features	Availability
BIOCELLS	489	2	79.6%	20.5%	64	Private
DERM	366	6	30.6%	5.5%	33	UCI
IIFI	600	3	36.0%	31.5%	14	UCBM
YEAST	1479	9	31.3%	1.6%	8	UCI
$ICPR_{BOF}$	721	6	28.9%	8.0%	1024	ICPR2012
$ICPR_{BIF}$	721	6	28.9%	8.0%	1024	ICPR2012

We used one private and five public datasets, belonging to images classification problems of different biomedical domains. They are characterized by a large variability with respect to the number and type of features, classes and samples, allowing the assessment of the performance of classifiers in different conditions. Synthetic data about the used datasets are reported in Table 3, while a more detailed description is reported in the following:

Bio Cells (BIOCELLS)[10]: The images were acquired by means of a fully fluorescence microscope. In biological experiments different NIS proteins mutated are expressed for putative sites of phosphorylation. The effect on the protein localization of each mutation is studied after immunostaining using anti-NIS antibodies. Immunocytolocalization analysis on 489 cells revealed 2 cell types with different subcellular distributions of NIS.

Dermatology (DERM)[11]: This is a dataset with 366 instances represented by 12 clinical features and 21 histopathological features taken from skin samples.

Indirect Immunofluorescence Intensity (IIFI) [12]: Connective tissue disease is autoimmune disorder identified by a chronic inflammatory process diagnosed by Indirect Immunofluorescence on HEp-2 substrate. The dataset consists of 14 statistical features extracted from 600 samples distributed over 3 classes.

Yeast (YEAST)[11]: This database contains information about 10 localization sites of Yeast cells. It is composed of 1484 instances represented by 8 features. We

remove the endoplasmic reticulum lumen class that makes impossible perform ten-fold cross validation since it has only 5 sample.

International Conference on Pattern Recognition HEp2 Cells (ICPR): HEp2 images were acquired by means of a fluorescence microscope coupled with a 50W mercury vapor lamp. This is a dataset has 791 instances distributed over six classes. We generated two version of this dataset, ICPR$_{BOF}$ and ICPR$_{BIF}$ using two kind of descriptors: Bag of Features and BIF respectively.

4 Experiments

To proof that UNN in conjunction with MDS provides robust performance, we performed several tests on six datasets comparing UNN performance with those achieved by SVM with a linear (SVM_l) and a gaussian (SVM_{RBF}) kernels, and by a kNN classifier. According to section 2.1, compared decomposition schemes are MDS and *Hamming decoding* (H_d).

As measure of classifier performance, we compute the accuracy and the F-*measure*. The latter is defined as F-$measure = 2((Recall)^{-1} \times (Precision)^{-1})^{-1}$. *Recall* is the fraction of samples labelled as belonging to the considered class that are correctly classified, whereas *Precision* is the fraction of samples in the considered class that are correctly classified.

Experiments are performed using a 10-fold cross validation scheme. Each fold is randomly generated maintaining the a-priori distribution of the original dataset. For each fold, classifiers parameters are optimized running a 5-fold cross validation. Reported results are computed averaging out the results obtained for each fold.

4.1 Classifier Reliability

We have tested three implementations of UNN, i.e. UNN(exp), UNN(log) and UNN(Mat), using three different loss functions for the learn of leveraging kNN. Reliabilities of UNN implementations tested measured in terms of posterior probabilities $\hat{p}_c(x)$ are computed as reported in Table 2. For further details, the interested reader may refer to [5]. To estimate the posterior probabilities for SVM we use the method presented in [13]. Given a SVM decision value f for class c, we compute $\psi_c(x)$ as:

$$\psi_c(x) = \frac{1}{1 + \exp(af(x) + b)} \ , \tag{9}$$

where a and b are estimated by maximizing the log-likelihood on training samples using a five-fold cross validation.

kNN is a statistical classifier where its classification reliability is computed using the definition reported in [14]:

$$\psi_c(x) = \min[\ \max\left(1 - \tfrac{O_{min}}{O_{max}}, 0\right)\ , 1 - \tfrac{O_{min}}{O_{min2}}\ ] \tag{10}$$

Table 4. Average values (%) of accuracy and F-measure of the different classifiers. We mark highest value (red) and the second one (green) in each row.

	Metrics (%)	Classifiers											
		UNN(exp)		UNN(log)		UNN(Mat)		SVM_l		$SVMRBF$		kNN	
		MDS	H_d	MDS	H_d	MDS	H_d	MDS	H_d	MDS	H_d	MDS	H_d
BioCells	*Accuracy*	87.1	87.1	86.5	86.5	**87.3**	**87.3**	74.3	74.3	**87.7**	**87.7**	85.2	85.2
	F-*measure*	77.7	77.7	76.9	76.9	**78.2**	**78.9**	66.9	66.9	76.9	76.9	75.2	75.2
DERM	*Accuracy*	97.5	**97.6**	96.5	96.4	**97.7**	97.1	97.1	87.7	96.9	95.5	95.9	95.5
	F-*measure*	**97.3**	**97.3**	96.1	95.7	**97.3**	**96.6**	96.5	81.2	**96.6**	95.1	95.4	95.2
IIFI	*Accuracy*	69.5	69.3	68.8	69.1	**70.8**	68.8	67.2	66.7	**71.5**	67.4	70.3	68.7
	F-*measure*	**69.0**	68.5	68.4	68.4	**70.3**	68.0	66.8	64.8	**70.3**	65.5	69.6	67.7
Yeast	*Accuracy*	**59.1**	58.0	57.1	55.5	53.9	53.5	52.8	48.3	**58.4**	54.5	54.1	54.3
	F-*measure*	**50.7**	46.3	47.5	45.4	41.5	40.9	41.2	24.2	**47.8**	41.7	46.1	44.5
ICPR$_{BOF}$	*Accuracy*	**88.1**	85.3	**87.1**	84.6	85.9	80.5	65.4	66.0	86.3	81.6	25.1	26.6
	F-*measure*	**87.4**	84.9	86.2	83.3	**85.8**	81.1	72.3	55.1	85.2	79.8	21.5	21.2
ICPR$_{BIF}$	*Accuracy*	**95.7**	**95.6**	94.9	95.5	95.4	94.9	91.8	89.8	95.3	94.4	95.1	93.91
	F-*measure*	**95.6**	**95.4**	94.4	95.4	**95.6**	95.1	90.7	85.5	95.2	94.0	94.8	93.7

O_{min} is the distance between x and the nearest sample of the validation set, i.e. the sample determining the class, O_{max} is the highest among the values of O_{min} obtained from all samples of the output class belonging to the test set, and O_{min2} is the distance between x and the nearest sample in the validation set belonging to a class other than the output one.

4.2 Results on Biomedical Images Datasets

We report in Table 4 the classification performance provided by UNN, SVM and kNN classifiers on the six datasets. For each classification task, we report the results obtained using both MDS and H_d reconstruction rules. In order to provide a global comparison among the results, we calculate the relative performance of each experimental configurations with respect to the others (Fig. 1). For each dataset, the twelve columns with the accuracy values are sorted individually, and each classification method is assigned a rank with respect to its place among the others. The largest rank is twelve (assigned to the best method) and the lowest is one (assigned to the worst method). The six ranks for each classification method are then summed up to give a measure of the overall dominance among the methods in terms of accuracy. An analogous procedure has been carried out in case of F-measure. The analysis of data reported both in Table 4 and Fig. 1 permits us to derive the following three considerations. The first one concerns the comparison between MDS and H_d reconstruction rules. Independently of the classifier and of performance metric considered, the former improves classification results in comparison with the latter over 90%.We deem that such performance improvement is mainly due to the fact that MDS rule uses not only predicted crisps labels, as H_d does, but also the corresponding classification reliability. The second consideration focuses on UNN, observing that its performance improve using

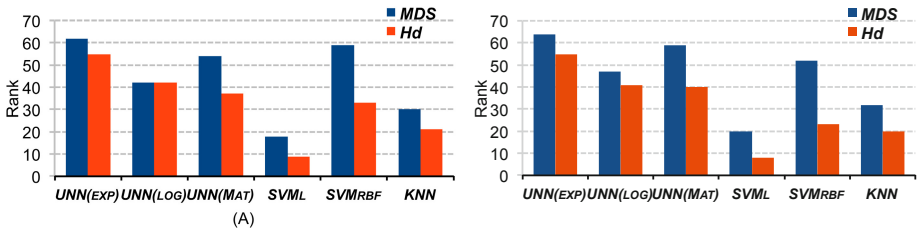

Fig. 1. Panel A and B show the rank for accuracy and F-measure respectively

posterior based reconstruction rule. Indeed, MDS scheme equals or improves UNN performance with H_d scheme in 85% of the cases, at least. For instance, focusing on ICPR$_{BOF}$ dataset, MDS improves UNN performance for all the three configurations of 2%, at least, in terms of both accuracy and F-measure. The third observation concerns how UNN performance compares with those provided by other classifiers. From a general point of view, turn our attention to Fig. 1 where we notice that the value of UNN(exp) rank is larger than the ones of other classifiers. Focusing now on recognition performance we note that UNN classifiers with MDS scheme always overcome performance of SVM with linear kernel. UNN also overcome kNN results with at least one configuration among the three tested. Comparing performance of UNN with those of SVM_{RBF} we note that results are quite similar.

5 Conclusion

In this paper we have proofed that the UNN algorithm in conjunction with a reconstruction rule based on the posterior probabilities provides a classification scheme robust among various biomedical image datasets. Indeed, this classification scheme outperforms other statistical and kernel-based classifiers. Furthermore, this reconstruction rule based on the posterior probabilities has shown larger recognition performance than a reconstruction rule based on crisp labels only.

References

1. Piro, P., Nock, R., Nielsen, F., Barlaud, M.: Leveraging k-nn for generic classification boosting. Neurocomputing 80, 3–9 (2012)
2. Allwein, E.L., et al.: Reducing multiclass to binary: a unifying approach for margin classifiers. Journal of Machine Learning Research 1, 113–141 (2001)
3. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research 2, 263 (1995)
4. Iannelo, G., et al.: On the use of classification reliability for improving performance of the one-per-class decomposition method. DKE 68, 1398–1410 (2009)
5. D'Ambrosio, R., Nock, R., Bel Haj Ali, W., Nielsen, F., Barlaud, M.: Boosting Nearest Neighbors for the Efficient Estimation of Posteriors (April 2012)

6. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning Journal 37, 297–336 (1999)
7. Nock, R., Nielsen, F.: On the efficient minimization of classification-calibrated surrogates. In: NIPS*21, pp. 1201–1208 (2008)
8. Nock, R., Nielsen, F.: Bregman divergences and surrogates for learning. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(11), 2048–2059 (2009)
9. Bartlett, P., Jordan, M., McAuliffe, J.D.: Convexity, classification, and risk bounds. Journal of the Am. Stat. Assoc. 101, 138–156 (2006)
10. Bel Haj Ali, W., et al.: A bio-inspired learning and classification method for subcellular localization of a plasma membrane protein. In: VISAPP 2012 (2012)
11. Frank, A., Asuncion, A.: UCI machine learning repository (2010)
12. Rigon, A., et al.: Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose. Cytometry Part B: Clin. Cytometry 72B(6), 472–477
13. Platt, J.-C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in Large Margin Classifiers, pp. 61–74. MIT Press (1999)
14. Cordella, L.P., et al.: Reliability parameters to improve combination strategies in multi-expert systems. Pattern Analysis & Applications 2(3), 205–214 (1999)