

Mitotic HEp-2 Cells Recognition under Class Skew

Gennaro Percannella¹, Paolo Soda², and Mario Vento¹

¹ Dipartimento di Ingegneria dell'Informazione ed Ingegneria Elettrica,
Università di Salerno, Italy

² Facoltà di Ingegneria, Università Campus Bio-Medico di Roma, Italy
{pergen,mvento}@unisa.it, p.soda@unicampus.it

Abstract. Indirect immunofluorescence (IIF) is the recommended method to diagnose the presence of antinuclear autoantibodies in patient serum. A main step of the diagnostic procedure requires to detect mitotic cells in the well under examination. However, such cells rarely occur in comparison to other cells and, hence, traditional recognition algorithms fail in this task since they cannot cope with large differences between the number of samples in each class, resulting in a low predictive accuracy over the minority class. In this paper we present a system for mitotic cells recognition based on multiobjective optimisation, which is able to handle their low a priori probability. It chooses between the output of a classifier trained on the original skewed distribution and the output of a classifier trained according to a learning method addressing the course of imbalanced data. This choice is driven by a parameter whose value maximises, on a validation set, two objective functions, i.e. the global accuracy and the accuracies for each class. The approach has been evaluated on an annotated dataset of mitotic cells and successfully compared to five learning methods applying four different classification paradigms.

1 Introduction

Systemic autoimmune rheumatic diseases are characterised by the presence of antinuclear autoantibodies (ANAs), whose detection by indirect immunofluorescence (IIF) on HEp-2 slides [1,2] is required to confirm the diagnosis. Diagnostic procedure consists of the following four steps: image acquisition, mitosis detection, fluorescence intensity classification and staining pattern recognition (figure 1). Such tasks are very challenging for medical doctor, affecting the reliability of IIF diagnosis. Indeed, since IIF is a subjective and semi-quantitative method, physicians could be conservative or liberal during image classification on the basis of their skills and background, giving rise to classification variability [3]. Another significant reason of uncertainty is the low contrast of borderline and negative samples. In order to guarantee the correctness of the test, producers add to the slides some mitotic cells¹, which give to medical doctors the confidence

¹ Mitosis is the process by which a eukaryotic cell separates the chromosomes in its cell nucleus into two identical sets in two nuclei.

with their decisions. Indeed, first, medical doctors verify the correctness of well preparation process by detecting at least one fluorescent mitotic cell. Second, mitotic cells provide information on image staining pattern since they match with certain kinds of IIF patterns, including all stainings of antigens with different distribution throughout the cell cycle, such as midbody, CENP-F, mitotic spindle, centriole/centrosome and NuMA staining [4].

Since the demand of autoimmune laboratory tests has recently increased and ANAs detection in routine practice is far from being standardised [3], recent interests have been directed towards the development of computer-aided-diagnosis (CAD) systems supporting IIF diagnostic procedure. Investigated topics covered the areas of image acquisition [5,6], image segmentation [7,8,9] and fluorescence intensity classification [10] as well as staining pattern recognition [4,9,11,12,13].

Although mitotic cells detection plays a crucial role for CAD development being the first step in IIF diagnostic procedure, we find only one work on this topic [14]. It presented a set of features for mitotic cells recognition, testing the approach on an artificially balanced dataset composed of 126 samples. However, such cells rarely occur in a well in comparison to other cells giving rise to a skewed a priori samples distribution between mitotic and non-mitotic classes. In the following we assume that such classes correspond to minority and majority classes, respectively. Traditional classification algorithms cannot be successfully employed to recognise mitotic cells, since they are biased towards the majority class, resulting in poor predictive accuracy over the minority one. This happens because they are designed to minimize errors over training samples, ignoring classes composed of few instances [15,16].

Since the development of CAD system in IIF cannot prescind from the recognition of mitotic cells, and considering the skewed nature of the recognition problem, in the following we present a CAD system for mitotic cells recognition based on a method suited for learning under class skew, evaluating its performance on an annotated dataset of HEp-2 mitotic cells and achieving promising results. Our work differs from [14] since we consider here a strong skewed dataset reflecting the a priori distribution occurring in daily practice. We would like to address this lack found in the literature, opening the chance to develop fully automatic systems for IIF image analysis.

The paper is organised as follows: next session introduces classification approaches handling skewed distribution, Section 3 describes the dataset of HEp-2 mitotic cells and concisely recalls the set of descriptors used to represent the samples, Section 4 presents the experimental results and the discussion. Finally, Section 5 provides concluding remarks.

2 Classification

In this section, we first summarise recent researches on learning with class skew, and then present the approach we use.

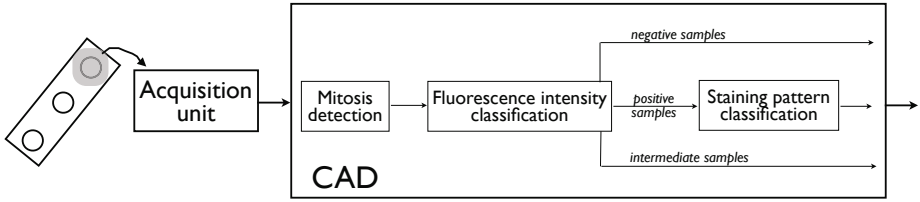


Fig. 1. Scheme of the diagnostic procedure for HEp-2 images

2.1 Classification Methods for Skewed Datasets

When the a priori samples distribution is skewed, traditional classification algorithms fail because they are designed to minimize errors over training samples, resulting in poor predictive accuracy over the minority one [15,16]. Existing learning approaches in these situations can be classified into the following five categories:

1. Under-sampling the majority class by resizing the training samples (TS), makes the class distribution more balanced [15,17]. This is done by sampling a subset N' from the majority training set N to make $|N'| = |P|$ where P is the minority class training set. The main drawback is the removal of potentially useful samples. One-sided selection is an under-sampling method that tries to overcome the above limitation removing borderline and redundant majority class samples, without touching minority class samples.
2. Over-sampling the minority class so as to match the size of the majority one, generates a set P' with $|P'| = |N|$ where P' is a combination of positive samples selected by this method, plus all positive samples in P . The main drawback of this approach is that it may increase the likelihood of over-fitting [15,17]. In order to overcome this problem, synthetic minority over-sampling technique was proposed [15]. It randomly creates synthetic samples in the feature space along the line segments to join any/all of the k minority class nearest neighbors [15].
3. Internally biasing the discrimination-based process to compensate class imbalance without altering the class distributions [16].
4. Multi-experts systems (MES), where each composing classifier C_i is trained on a subset of the majority class and on the whole minority class. After sampling several subsets N_i from N , C_i is trained on $N_i \cup P$. Then, outputs of C_i on test sample are combined to decide the final predications [18,19]. Indeed, a MES generally produces better results than those provided by its composing classifiers, avoiding drawbacks of both under and oversampling since each C_i is now trained on balanced sub-problems containing information on different aspects of N . Furthermore, in the framework of MES, in [18] the authors proposed the BalanceCascade method that supervisesly explores the majority class using a cascade of base classifiers sequentially trained. As the aforementioned learning methods, MES approach produces better results on the minority class while harms the recognition of majority class samples.

5. Multiobjective optimisation: this method chooses between the output of a classifier trained on the original skewed distribution and the output of a classifier trained according to a learning method addressing the course of imbalanced data. This choice is driven by a parameter whose value maximises, on a validation set, two objective functions, i.e. the global accuracy and the accuracies for each class [20]. This approach would balance the recognition accuracies for each class, harming as less as possible the global accuracy. Indeed, other learning methods for class imbalance increase the accuracy over the minority class and decrease the accuracy over the majority one.

Finally, observe that the global recognition accuracy (acc) is sensitive to class skew and, hence, cannot be used as the only metric for performance estimation. In such cases, the geometric mean of accuracies ($gacc$) is also used to measure the classification performance [17]. It is defined as $gacc = \sqrt{acc^+ \cdot acc^-}$, where acc^+ and acc^- denote true positive and true negative rates, respectively. $gacc$ is a non-linear measure since a change in one of the two accuracies will influence the value of $gacc$.

2.2 Recognition Approach

Preliminary, we experimentally observe that learning algorithms numbered above as 1, 2, 3 and 4 provide unsatisfactory performance since they fail on most of the minority class samples, as shown in Table 1 and discussed in section 4. In order to overcome such an issue, we utilize a multi-objective optimization technique selecting the final output of the classification system between the output of a classifier trained according to a learning method addressing the course of class imbalance, and the output of a classifier adopting a training method for non-skewed data. This choice is driven by a parameter, referred to as threshold \bar{t} in the following, whose value maximizes two objective functions, i.e. the global accuracy and the geometric mean accuracy.

The framework of our method is based on two classifiers, one trained on the original skewed distribution and another one built according to a class imbalance learning method. Given a test sample x , its label is given by:

$$O(x) = \begin{cases} O_{nSC}(x) & \text{if } \phi(x) \geq \bar{t} \\ O_{SC}(x) & \text{otherwise} \end{cases} \quad (1)$$

where $O_{nSC}(x)$ is the label assigned to sample x by any classifier trained on the original skewed distribution that does not apply any learning methods for skewed TS, which is referred to as nSC in the following. $O_{SC}(x)$ is the label assigned to sample x by any classifier trained according to a traditional learning method addressing the course of imbalanced TS, e.g. oversampling, undersampling, etc., which is named as SC in the rest of the paper. $\phi(x)$ is the reliability assigned by nSC to sample x and \bar{t} is a real number in $[0, 1]$. The classification rule reported in 1 assigns the final label of x to the label returned by nSC when its reliability is larger than the threshold \bar{t} because, in this case, it is reasonable to assume that nSC is likely to provide a correct classification. When $\phi(x)$ is below \bar{t} , $O(x)$

is equal to the label assigned by a classification system trained according to a method specifically tailored for imbalanced TS. Indeed, in this case the value of the reliability suggests that the decision returned by nSC should be not safe.

The value of the threshold \bar{t} is set so that it maximizes both acc and $gacc$ on a validation set, providing the best global performance as well as the most balanced accuracies on set disjoint from both training and test sets. The learning algorithm works as follows:

- Divide the labeled HEp-2 dataset \mathbf{D} into training, validation and test sets, denoted by \mathbf{D}_{Tr} , \mathbf{D}_{Va} , \mathbf{D}_{Te} .
- Using \mathbf{D}_{Tr} , train both nSC and SC.
- Classify instances of \mathbf{D}_{Va} with such classifiers.
- Apply equation 1 and measure both $gacc$ and acc for each value of a threshold t ranging in $[0, 1]$. Indeed, $gacc$ measures how much the accuracies over two classes are balanced, whereas acc estimates the global performance of the classification system.
- Build a graph where $gacc$ and acc on the X and Y axis, respectively. The variation of t generates a set of points that can be used to plot a curve. Notice that curve extrema at $t = 0$ and $t = 1$ correspond to nSC and SC performance, respectively.
- The value \bar{t} is given by $\bar{t} = \arg \min_t (\|\mathbf{p}(t) - \mathbf{C}\|)$, where $\mathbf{p}(t)$ is the pair of $gacc(t)$ and $acc(t)$ values measured on the validation set when the threshold t is used. $\mathbf{C} = (1, 1)$ is the ideal point in this plot since, the nearer the curve to this point, the better the performance obtained.

Notice that it is possible to proof that \bar{t} is also an optimum solution of our problem according to multi-objective optimization theory [20].

3 Data Set and Features Extraction

Since, to our knowledge, there are not reference databases of IIF images publicly available, we populated a database of annotated mitotic cells with a resolution of 1388x1038 pixels, and a colour depth of 24 bits. The database consists of 28 positive images, showing the main six staining patterns shown in Fig. 3 [4]. Specialists manually segmented and annotated each cell from such images, reporting data on fluorescence intensity, pattern and mitosis phase. After this process, we get a mitotic dataset composed of 1527 cells, 70 mitotic cells and 1457 non mitotic cells. Non mitotic cells exhibits one of the six main staining patterns. The a priori probability of minority class is 4.8% and, hence, the dataset has a strong degree of imbalance in samples distribution.

Mitotic cells show peculiarities that can be exploited to get a set of features specifically tailored for this application. Such cells may exhibit two fluorescent patterns. The first, named as *negative mitosis*, has a fluorescent cell body, while the collapsed chromosomes mass located in the middle part of the cell does not exhibit a fluorescent pattern, or it has a weak fluorescence (panels A and B of Fig. 2). In the second pattern, referred to as *positive mitosis*, we observe

the opposite situation, i.e. the cell body is weakly or not fluorescent, while the chromosomes mass is fluorescent (panels C and D of Fig. 2). In both cases the collapsed chromosomes mass has a circular or elliptic shape, approximately. Comparing Fig. 2 and 3, we notice that such mitotic patterns are different from those of non mitotic cells.

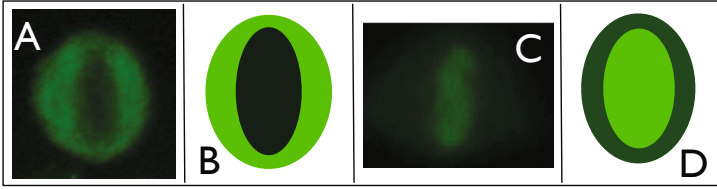


Fig. 2. Examples and stylised representations of positive (panels A-B) and negative (panels C-D) mitosis. Light and dark green represent high and low fluorescence.

In order to catch such peculiarities, we compute an heterogeneous set of features belonging to morphological descriptors, texture measures, and local binary pattern (LBPs) [14].

Morphological descriptors look for elliptic shape and analyse the fluorescence intensity inside the cells. Information on the elliptic shape of chromosome mass, being either a positive or negative mitosis, is computed fitting a bivariate Gaussian distribution inside the cell body and deriving features based on the analysis of consecutive fittings. The second set of features consists of texture measures related to statistical and spectral descriptors. The third set of features is based on LBPs, which assign to each pixel of the image a label obtained comparing it with its neighbourhood matrix. The interested reader can refer to [14] for further details on used features.

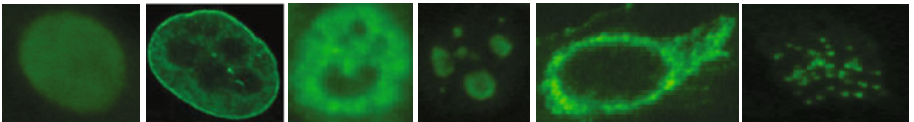


Fig. 3. Examples of the homogenous, peripheral nuclear, speckled, nucleolar, cytoplasmic and centromere staining patterns (left to right).

4 Results

The classification approach chooses the final output between the predictions of nSC and SC. As reported above, nSC is a classifier trained on the original skewed distribution that does not apply any learning method for imbalanced TS, whereas SC is a classifier trained according to a traditional learning method addressing the course of skewed TS, e.g. oversampling, undersampling, etc.

In order to analyse the classification performance, we test the following nine different classification configurations:

1. nSC: a classifier trained on the original skewed distribution which does not apply any learning methods for skewed TS;
2. SMOTE: a classifier trained according to synthetic minority over-sampling technique. In its implementation, we considered the three nearest neighbours, as suggested in [15];
3. OSS: a classifier trained according to one-sided selection [17];
4. MES-DCS: a multi-experts system where each classifier is trained on a subset $N_i \cup P$. According to results reported in [19], we apply random selection to sample N and dynamic classifier selection with local accuracy [21] to combine the outputs of base classifiers;
5. BaCa: a multi-experts system trained according to the BalanceCascade serial scheme [18];
6. nSC + SMOTE: a multi-objective optimisation technique choosing the final output between the predictions of nSC and SMOTE;
7. nSC+OSS: a multi-objective optimisation technique choosing the final output between the predictions of nSC and OSS;
8. nSC+MES-DCS: a multi-objective optimisation technique choosing the final output between the predictions of nSC and MES-DCS;
9. nSC+BaCa: a multi-objective optimisation technique RbB method choosing the final output between the predictions of nSC and BaCa.

Testing different configurations allows us to investigate different aspects of our recognition task. Indeed, configuration 1 permits us to estimate the performance achievable using a classifier trained on the original skewed distribution. Configurations 2-5 permit us to measure the performance achievable using “traditional” learning methods for skewed TS, whereas classifiers 6-9 are different applications of the proposed method. On this basis, we can compare the results of different classification methods on a strong skewed dataset.

We test four popular classifiers belonging to different paradigms. They are a Multi-Layer Perceptron (MLP) as a neural network, a k -Nearest Neighbour as a statistical classifier (k NN), a Support Vector Machine (SVM) as a kernel machine, and AdaBoost as an ensemble of classifiers. We report the results achieved using the best configuration of each classifier, set by preliminary tests.

Recognition performance, evaluated according to a 3-fold cross validation averaging out the final values of $gacc$ and acc , are reported in Table 1. The first part of the table reports the mean values of global accuracy (acc), which measures the global recognition rate. The second part of the table shows the mean values of geometric mean of accuracies ($gacc$), reflecting how much classifier performance are balanced. The third part of the table provides a synthetic comparison between the performance of nine classification schemes, considering together two previous performance metrics. To this aim, we compute for each classifier the L^2 distance between the average performance of each classification configuration and the ideal point \mathbf{C} with coordinates $(1, 1)$ (see section 2.2). In order to facilitate the comparison between such distances, these values are normalised with

Table 1. Performance of tested classification configurations

Classifier	Classification Performance (<i>acc%</i>)								
	nSC	SMOTE	OSS	MES-DCS	BaCa	nSC+SMOTE	nSC+OSS	nSC+MES-DCS	nSC+BaCa
SVM	95.4	95.4	95.4	52.4	92.2	95.4	95.4	65.4	94.0
AdaBoost	95.4	94.8	94.2	57.4	94.6	95.1	94.7	74.8	95.3
MLP	95.4	94.5	93.3	55.3	93.1	94.5	94.9	74.7	94.4
3NN	95.2	92.9	84.0	54.6	92.4	92.7	84.5	74.6	92.5
Classification Performance (<i>gacc%</i>)									
SVM	0.0	0.0	0.0	52.8	23.5	0.0	0.0	52.9	25.4
AdaBoost	0.0	14.2	6.7	52.1	21.3	21.3	7.1	54.7	22.0
MLP	0.0	9.5	6.5	55.3	4.3	20.1	9.0	54.4	20.7
3NN	7.4	23.5	27.0	50.2	20.3	27.0	30.8	54.3	27.8
Normalised distance with ideal point C (<i>d</i>)									
SVM	70.8	70.8	70.8	47.4	54.4	70.8	70.8	41.3	52.9
AdaBoost	70.8	60.8	66.1	45.3	55.8	55.8	65.8	36.7	55.2
MLP	70.8	64.1	66.3	44.7	67.8	56.6	64.4	36.9	56.2
3NN	65.6	1591.0	52.8	47.7	56.6	51.9	50.1	37.0	51.3

respect to $\sqrt{2}$, which is the maximum distance from point **C**. Hence, such values range in $[0,1]$: the smaller the value, the more balanced the performance of the classifier.

Reported performance can be analysed from two points of view. First, we can compare the different learning methods on a high skewed dataset. Second, we can analyse the results considering the biomedical application at hand.

With reference to a machine learning perspective, we observe that nSC returns a value of *acc* equal to the a priori probability of majority class, thus misclassifying all minority class samples. This observation is confirmed by the value of *gacc*, which is equal to zero. This happens for all classifiers, other than 3NN. Introducing traditional learning methods for class imbalance, i.e. SMOTE, OSS, MES-DCS and BaCa, the situation improves since, now, *gacc* is different from zero. However, while *acc* keeps large, the value of *gacc* is still small, showing that such learning methods suffer with a so strong degree of imbalance. This does not happen using a SVM classifier in conjunction with learning methods for skewed data. Reason should lie in observing that such learning methods do not improve the linear separability of the samples induced by the chosen kernel in the features space. Let us now focus on the four applications of the proposed methods, i.e. nSC+SMOTE, nSC+OSS, nSC+MES-DCS and nSC+BaCa. With respect to the corresponding composing classifier, we observe that values of *acc* are approximately the same except for nSC+MES-DCS, whereas values of *gacc* increase. In case of nSC+MES-DCS, improvement of *gacc* counterbalances the decrease of *acc*. Turning our attention to the normalised distance from the ideal point, representing a perfect classification of both positive and negative samples, we notice that nSC+MES-DCS provides the best balanced performance, i.e. the best balance between the recognition rates of positive and negative samples. This happens for all tested classifiers, suggesting that such a scheme is the one best suited for this application. This observation introduces us to look at the results from a biomedical point of view. Results achieved by any nSC are clearly useless

in a real scenario, since they fail on all positive samples which are the mitotic cells we would like to detect. The introduction of learning methods for skewed data improves a little the situation in comparison with nSC, since now *gacc* is larger than zero. However, in several cases, the values of *gacc* are still small, suggesting that such methods misclassify several positive samples. In contrast, the four applications of the proposed learning rule for the four different classifiers increase the value of *gacc*, implying that more mitotic cells are now correctly classified.

5 Conclusions

In this paper we have presented a classification approach for mitotic cells since their detection is a fundamental issue in developing a comprehensive CAD system in IIF. We have taken into consideration their low a priori probability, applying a learning rule which provides more balanced recognition rate with respect both to traditional classification methods as well as to other learning methods for class imbalance.

Future works are directed towards the integration of the system for mitotic cells recognition with systems for IIF image acquisition, fluorescence intensity classification and staining pattern recognition, according to figure 1.

Acknowledgments. This work has been supported by “Regione Campania” within the project “Classification of Immunofluorescence Images for the Diagnosis of Autoimmune Diseases” and by “Regione Lazio” under the programme “Intese con i Parchi scientifici e tecnologici del Lazio per favorire l’innovazione e il trasferimento tecnologico alle PMI laziali”. The authors thank DAS s.r.l. for acquisition equipment and A. Rigon and L. Onofri for IIF images acquisition and annotation.

References

1. Kavanaugh, A., Tomar, R., et al.: Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. American College of Pathologists, Archives of Pathology and Laboratory Medicine 124(1), 71–81 (2000)
2. Rigon, A., Soda, P., et al.: Indirect immunofluorescence in autoimmune diseases: Assessment of digital images for diagnostic purpose. Cytometry B (Clinical Cytometry) 72, 472–477 (2007)
3. Bizzaro, N., Tozzoli, R., et al.: Variability between methods to determine ANA, anti-dsDNA and anti-ENA autoantibodies: a collaborative study with the biomedical industry. Journal of Immunological Methods 219, 99–107 (1998)
4. Sack, U., Knoechner, S., et al.: Computer-assisted classification of HEp-2 immunofluorescence patterns in autoimmune diagnostics. Autoimmunity Reviews 2, 298–304 (2003)
5. Hiemann, R., Hilger, N., et al.: Objective quality evaluation of fluorescence images to optimize automatic image acquisition. Cytometry Part A 69, 182–184 (2006)

6. Soda, P., Rigon, A., et al.: Automatic acquisition of immunofluorescence images: Algorithms and evaluation. In: *Computer Based Medical Systems*, pp. 386–390. IEEE Computer Society, Los Alamitos (2006)
7. Huang, Y.L., Jao, Y.L., et al.: Adaptive automatic segmentation of HEP-2 cells in indirect immunofluorescence images. In: *IEEE Int. Conf. on Sensor Networks, Ubiquitous and Trustworthy Computing*, pp. 418–422 (2008)
8. Huang, Y.L., Chung, C.W., et al.: Outline detection for the HEP-2 cells in indirect immunofluorescence images using watershed segmentation. In: *IEEE Int. Conf. on Sensor Networks, Ubiquitous and Trustworthy Computing*, pp. 423–427 (2008)
9. Perner, P., Perner, H., Muller, B.: Mining knowledge for HEP-2 cell image classification. *Journal Artificial Intelligence in Medicine* 26, 161–173 (2002)
10. Soda, P., Iannello, G., Vento, M.: A multiple experts system for classifying fluorescence intensity in antinuclear autoantibodies analysis. *Pattern Analysis & Applications* 12(3), 215–226 (2009)
11. Hiemann, R., Büttner, T., et al.: Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEP-2 cells. *Autoimmunity Reviews* 9(1), 17–22 (2009)
12. Hiemann, R., Hilger, N., et al.: Automatic analysis of immunofluorescence patterns of HEP-2 cells. *Annals of the New York Academy of Sciences* 1109(1), 358–371 (2007)
13. Soda, P., Iannello, G.: Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis. *IEEE Transactions on Information Technology in Biomedicine* 13(3), 322–329 (2009)
14. Foggia, P., Percannella, G., et al.: Early experiences in mitotic cells recognition on hep-2 slides. In: *23rd IEEE Int. Symp. on Computer-Based Medical Systems, CBMS 2010*, pp. 38–43 (2010)
15. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16(3), 321–357 (2002)
16. Barandela, R., Sanchez, J.S., Garca, V., Rangel, E.: Strategies for learning in class imbalance problems. *Pattern Recognition* 36(3), 849–851 (2003)
17. Kubat, M., Matwin, S.: Addressing the curse of imbalanced training sets: One-sided selection. In: *Machine Learning-International Workshop Then Conference*, pp. 179–186. Morgan Kaufmann Publishers, Inc., San Francisco (1997)
18. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 39(2), 539–550 (2009)
19. Soda, P.: An experimental comparison of MES aggregation rules in case of imbalanced datasets. In: *22nd IEEE Int. Symp. on Computer-Based Medical Systems, 2009*, pp. 1–6 (2009)
20. Soda, P.: A multi-objective optimisation approach for class-imbalance learning. *Pattern Recognition* (2010) (in press)
21. Woods, K., Kegelmeyer, W.P., Bowyer, K.: Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 405–410 (1997)