Performance Analysis of Classifiers with Feature Selection and Optimization in CBIR System for Biological Images

B. S. Shajee Mohan^{a,*}, K. S. Shanthini^b

^aAssociate Professor and Head, A E & I Dept., Govt. Engineering College Kozhikode, Kerala, India, 673005.

^bAssistant Professor, A E & I Dept., Govt. Engineering College Kozhikode, Kerala, India, 673005.

*Corresponding Author: shajeemohan@yahoo.com

Abstract

The core objective of this paper is to improve the performance of Content Based Image Retrieval (CBIR) system for biological images by intelligent selection of most discriminative set of features from the canonical set of features. The performance of the CBIR system can be further enhanced by proper selection of a classifier and fine tuning of model parameters to obtain improved classification accuracy. We extracted canonical set of feature vectors from biological images using a popular tool called Weighted Neighbor Distance using Compound Hierarchy of algorithms Representing Morphology (**Wndchrm**). We adopted a two step approach for the selection of features. The first step is to partition the canonical set of feature vectors into four distinct set of feature vectors based on the methodology and algorithms for extraction of features. The second step is to perform Principal Component Analysis (PCA) and Fisher Score based selection of features from the partitioned set of feature vectors. The optimum set of feature vectors thus obtained is applied as training patterns to different classifier implementations such as Bayesian and Support Vector Machine (SVM) Classifiers. We used IICBU-2008 benchmark dataset [5, 7] of biological images for our experiments. We also compared the results with the results available for the wndchrm classifier. The results show that careful selection of the features and optimization of the classifier performance will lead to efficient implementation of CBIR systems.

Keywords: feature selection, canonical feature set, Fisher score, Bayesian classifier, support vector machine classifier, classification accuracy.

1. Introduction

Content based image retrieval (CBIR) systems are becoming popular in these days. These systems help in retrieving the desired images from a database containing fairly large number of multiple category images. The schematic of a typical CBIR system is shown in Figure 1. Most of the CBIR systems make use of relevance feedback to improve the system performance. We consider the design of CBIR system for biological images. An exclusive CBIR system for biological images will really help the scientific community in great deal in analyzing, storing and retrieving biological images. We consider a two-step process for the optimization of the classifier performance in CBIR system. The first step is to select features from a canonical set of features to obtain partitioned sets of feature vectors. The selection is based on the methodology and algorithms for extraction of features. The second step is to select optimum features using principal component analysis (PCA) and feature selection methodology that make use of Fisher score. After performing the above mentioned feature extraction and selection processes, we applied these patterns to different classifier implementations such as Bayesian classifier and support vector machine (SVM) based classifier. We analyzed the performance of these classifiers by proper selection of the model parameters.

The remaining part of this paper is organized as follows: Section 2 describes various methodologies adopted for feature extraction and selection. In section 3 we explain the basic theory associated with the implementation of Bayesian, SVM based classifiers. We also give an overview

of wndchrm [2, 3] classifier. In section 4 we present the details of the experiments and the subsequent observations. We make a conclusion of our discussions in section 5.

2. Feature Extraction and Selection

2.1 Canonical Set of Feature Vectors

Wndchrm [2, 3] is an open source utility used for biological image analysis. The software enables the user to extract the features from an image. Wndchrm have two options that enable the user to select between extracting a smaller set of features containing 1025 features and a larger set containing 2659 features from an image.

Figure 1. Schematic of a general CBIR system

The canonical set of feature vectors extracted by wndchrm contain Radon transform features [4], Chebyshev transform features, Gabor filter features, features extracted from multi scale histogram, first four moments, Tamura texture features of contrast, directionality and coarseness, edge and object statistics features, Zernike, Haralick features and Chebyshev Fourier features.

2.2 Selection of Features

In this paper we select 4 sets of feature vectors for each image from the original canonical feature vector set. They are:

- 1) **Set 1:** Basic feature set include Radon transform, Chebyshev transforms, Zernike features, Haralick features etc.
- 2) **Set 2:** Transform of features such as Wavelet transform of Chebyshev features, Fourier of Haralick features etc.
- 3) **Set 3:** Transform of transform of features such as Fourier transform (Wavelet transform (Haralick features)) etc.
- 4) **Set 4:** Edge features and transform of Edge features.

After partitioning the canonical feature vector set into 4 distinct feature vector sets, we further optimize the number

of features using principal components to analysis (PCA) and Fisher score based methods.

Principal Component Analysis

Principal Component Analysis describe a methodology for finding low dimensional representation (or projections) of high dimensional data that still maintain and capture most of the variability and discriminant information in the original representation. We summarize the steps for finding principal components.

- 1) Find the mean (μ) of the feature set $(\{\mathbf{x}_i\}_{i=1}^N)$ where N is the total examples in the training set.
- 2) Find $\mathbf{x}_i - \mu$ for $i = 1, \dots, N$.
- 3) Find the covariance matrix of the entire training data.
- 4) Find the Eigen values and vectors corresponding to the covariance matrix.
- 5) Sort the Eigen value in descending order.
- 6) Set a threshold for the Eigen value and take the Eigen vector above this threshold value. Project the data in the direction of the Eigenvectors to get reduced dimensional representation.

Fisher Score based Feature Selection

The objective of finding Fisher score is to obtain a subset of features, such that in the data space spanned by the selected features, the distance between data points in different classes is maximized, while the distances between the data points in the same class is maintained as small as possible [10]. Fisher score is evaluated using the equation given below [10].

$$F_j = \frac{\sum_{k=1}^M n_k (\mu_k^j - \mu^j)^2}{(\sigma^j)^2} \quad (1)$$

Where μ_k^j is the mean of j^{th} feature for the examples of the k^{th} class, and n_k is the number of examples in k^{th} class. μ^j and σ^j denote the mean and standard deviation of the j^{th} feature over the whole data set. The total variance for the j^{th} feature is given by $(\sigma^j)^2 = \sum_{k=1}^M n_k (\sigma_k^j)^2$, where σ_k is the standard deviation of the j^{th} feature for the examples of the k^{th} class. The total number of classes is denoted by M .

3. Classifiers for Biological images

3.1 Bayesian Classifiers

Approaches based on statistical pattern classification make use of Bayesian decision theory [6]. This approach is

based on making decisions by using probability and by minimizing the risk involved. Bayes formula can be expressed informally as:

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Evidence}} \quad (2)$$

Pattern classifiers can be implemented using different methodologies. An efficient way of building a classifier is by using a set of discriminant functions $g_i(\mathbf{x})$ for $i = 1, \dots, C$. The classifier is said to assign a feature vector $\mathbf{x}$ to class $\mathbf{w}_i$ if $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for all $i \neq j$ [6].

We assume that the density distribution of the data set is bivariate Gaussian. Under this assumption, the use of the discriminant functions $g_i(\mathbf{x}) = \ln \mathbf{p}(\mathbf{x}/\mathbf{w}_i) + \ln P(\mathbf{w}_i)$ can yield minimum-error-rate classification. The discriminant function can be evaluated by substituting $\mathbf{p}(\mathbf{x}/\mathbf{w}_i)$ by the multivariate normal distribution. Then $g_i(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i) - \ln |\Sigma_i| + \ln P(\mathbf{w}_i)$ where $\mathbf{x}$ is a d dimensional column vector μ is a d dimensional mean vector Σ is the $d \times d$ covariance matrix, $|\Sigma_i|$ is its determinant, and $(\mathbf{x} - \mu_i)^T$ is the transpose of $(\mathbf{x} - \mu_i)$.

Case 1: $\Sigma_i = \sigma^2 \mathbf{I}$

In case 1 the features are assumed to be statistically independent, under this assumption, each feature has the same variance, σ^2 . In this case the covariance matrix is diagonal, being merely σ^2 times the identity matrix $\mathbf{I}$. Geometrically, this situation corresponds to the condition that the samples fall in equal-size hyper-spherical clusters, the cluster for the i^{th} class being centered about the mean vector μ_i . Since both $|\Sigma_i|$ and the $\frac{d}{2} \ln 2\pi$ term are not dependent on i , they can be ignored. We can deduce the simple linear discriminant function as:

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + \mathbf{w}_{i0}, \text{ where}$$

$$\mathbf{w}_i = \left(\frac{1}{\sigma^2}\right) \mu_i \quad \text{and} \quad \mathbf{w}_{i0} = -\left(\frac{1}{2\sigma^2}\right) \mu_i^T \mu_i + \ln P(\mathbf{w}_i)$$

If prior probabilities $P(\mathbf{w}_i)$ are the same for all 'C' classes, then the $\ln P(\mathbf{w}_i)$ term can be ignored. Then, the optimum decision rule can be stated as: to classify a feature vector $\mathbf{x}$, measure the Euclidean distance $\|\mathbf{x} - \mu_i\|$ from each of the C mean vectors, and assign class label to $\mathbf{x}$ to the class of the nearest mean. This type of classifier is called a minimum distance classifier [6].

Case 2: $\Sigma_i = \Sigma$

Case 2 deals with the condition that all the classes are having the same covariance matrix that is otherwise arbitrary. This leads to the geometrical situation that the samples fall in hyper ellipsoidal clusters of equal size and shape, the cluster for the i^{th} being centered about the mean vector μ_i . Since both $|\Sigma_i|$ and the $\frac{d}{2} \ln 2\pi$ term are no dependent on i , and can be ignored, thus leads to the discriminant functions.

$$g_i(\mathbf{x}) = \left(\frac{1}{2}\right) (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) + \ln P(\mathbf{w}_i) \quad (3)$$

If the prior probabilities $P(\mathbf{w}_i)$ are the same for all 'C' classes, then the $\ln P(\mathbf{w}_i)$ term can be ignored. In this case, the optimal decision rule can once again be stated as: to classify a pattern $\mathbf{x}$, measure the squared Mahalanobis distance $(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)$ from $\mathbf{x}$ to each of the 'c' mean vectors, and assign $\mathbf{x}$ to the category of the nearest mean [6]. The discriminant functions are again linear and given as follows:

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + \mathbf{w}_{i0}, \text{ where}$$

$$\mathbf{w}_i = \Sigma^{-1} \mu_i \quad \text{and} \quad \mathbf{w}_{i0} = -\left(\frac{1}{2}\right) \mu_i^T \Sigma_i^{-1} \mu_i + \ln P(\mathbf{w}_i)$$

Case 2: $\Sigma = \text{Arbitrary}$

The most general case for the assumption of multivariate normal density, the covariance matrices are different for each category. In this situation, only $\frac{d}{2} \ln 2\pi$ term that can be dropped which will lead to quadratic discriminant functions as follows:

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + \mathbf{w}_{i0}, \text{ where}$$

$$\mathbf{W}_i = -\left(\frac{1}{2}\right) \Sigma_i^{-1}; \quad \mathbf{w}_i = \Sigma_i^{-1} \mu_i \quad \text{and}$$

$$\mathbf{w}_{i0} = -\left(\frac{1}{2}\right) \mu_i^T \Sigma_i^{-1} \mu_i + \ln |\Sigma_i| + \ln P(\mathbf{w}_i)$$

The decision surfaces are of complex shapes which include hyper quadrics, hyper spheres, hyper ellipsoids, hyper paraboloids, and hyper-hyperboloids of various types [6].

3.2 Support Vector Machine based Classifiers

SVM is a popular binary learning machine with some

highly desirable properties. The concept of SVM based on two mathematical operations [9]: (1) The mapping of the training patterns using a non-linear mapping into a high dimensional feature space that is hidden from both input and output. (2) Construction of a maximum margin hyper-plane for separating the patterns obtained using the non-linear transformation in step 1. The separating optimal hyper-plane is completely determined by the support vectors which lie on the marginal hyper-planes. Thus, SVM gives an analytical approach for finding the optimum separating hyper-plane [9]. For a given training set $\mathcal{T} = \{(\mathbf{x}_i, d_i)\}_{i=1}^N$ where $\mathbf{x}_i$ is the input pattern for the i^{th} example and d_i is the corresponding desired response. SVM construct the decision surface such that the margin of separation between positive labeled and negative labeled examples is maximized. The equation for optimum hyper-plane for linearly separable patterns is given by $\mathbf{w}_0^T \mathbf{x} + \mathbf{b}_0 = 0$, where $\mathbf{w}_0$ is the optimum weight vector, $\mathbf{x}$ is the input pattern and $\mathbf{b}_0$ is the optimum bias [9].

In the case of non-linearly separable classes, the margin of separation is considered to be soft if data pattern $\mathbf{x} \in (\mathbf{x}_i, d_i)_{i=1}^N$ violates the condition $\mathbf{w}_0^T \mathbf{x} + \mathbf{b}_0 = 1$. The violations are of two types. The data point $\mathbf{x} \in (\mathbf{x}_i, d_i)_{i=1}^N$ falls inside the margin of separation and is on the correct side of the decision surface. Here, there is no misclassification. If the data point falls inside the margin of separation and is on the wrong side of the decision surface, then this leads to misclassification of the pattern. So, a new set of non negative scalar variable known as slack variables $\{\xi_i\}_{i=1}^N$ are introduced. Now the hyper plain will be $\mathbf{w}_0^T \mathbf{x} + \mathbf{b}_0 = 1 - \xi_i$, where ξ_i measures the deviation of the data point from the ideal condition of pattern separation [9]. SVMs can also be thought as kernel machines that can use different Mercer kernels. These include linear $(\mathbf{x}^T \mathbf{y})$, Polynomial $((1 + \mathbf{x}^T \mathbf{y})^q)$ where q is the degree of the polynomial and Gaussian kernel $e^{-(a\|\mathbf{x}-\mathbf{y}\|^2)}$, where the feature vector $\mathbf{x}$ is of d dimension.

Let $\mathcal{T} = \{(\mathbf{x}_i, d_i)\}_{i=1}^N$ be the training set, where $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^d$ and $d_i \in \{+1, -1\}$. The kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ and $n \times n$ kernel gram matrix be $\mathbf{K}$ where the ij^{th} element is denoted as $k(\mathbf{x}_i, \mathbf{x}_j)$. SVM make use of the kernel trick so that explicit computation of the transformation is not required for the computation of the kernel gram matrix for the training data. The quadratic optimization problem for the soft margin SVM can entirely be expressed in terms of kernel gram matrix as follows [9]:

$$\min_{\alpha_1, \alpha_2, \dots, \alpha_N} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j d_i d_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

Subject to the constraints:

1. $\sum_{i=1}^N \alpha_i d_i = 0$,
2. $0 \leq \alpha_i \leq C$ for $i = 1, 2, \dots, N$.

Where C is a user defined parameter and $\alpha_1, \alpha_2, \dots, \alpha_N$ are the Lagrangian multipliers that correspond to the inputs $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$. Solving the above optimization problem yield α_{oi} s. The optimum weight vector $\mathbf{w}_0$ and optimum bias $\mathbf{b}_0$ can be computed so as to obtain the equation of decision surface with maximum margin

$$\mathbf{w}_0^T \mathbf{x} + \mathbf{b}_0 = 0 \text{ for } i = 1, 2, \dots, N.$$

An arbitrary test example $\mathbf{x}_t$ is classified with the help of the decision rule

$$h(\mathbf{x}_t) = \text{sgn}(\sum_{j=1}^N \alpha_j d_j k(\mathbf{x}_t, \mathbf{x}_j) + \mathbf{b}_0) \quad (5)$$

SVM Torch [8]: This is a freely available tool for SVM classification. It can also be used for regression problems. With this algorithm one can now efficiently solve large scale problems containing tens of thousands of patterns and even for input dimension higher than 100. It is also suitable for multiclass problems. SVM torch provides several options for the user to select the type of kernels, parameters of kernels, etc., SVM torch is the training machine and SVM test is the testing machine. The command line used for training data is:

`./svm_torch [options] Example_file model_file`

The command line used for testing the data is

`./svm_test [options] model_file Test_file`

3.3 Wndchrm Classifier [2, 3, 5]

The open source utility called Wndchrm [2, 3] is a powerful tool for biological image analysis that can be used on a wide range of biological datasets. Wndchrm has an in-built image classifier based on the variation of nearest neighbour classification. In this classifier, the distance of a pattern $\mathbf{x}$ extracted from a test image to a class 'C' is computed as follows [2, 5]:

$$d(\mathbf{x}, C) = \frac{\sum_{t \in T_C} \left[\sum_{j=1}^{|\mathbf{x}|} F_j (x_j - t_j)^2 \right]^p}{|T_C|} \quad (6)$$

where T_C is the training set of class 'C', t is a feature vector from T_C , $|\mathbf{x}|$ is the dimension of the feature vector $\mathbf{x}$,

x_j is the value of feature j in the vector $\mathbf{x}$, F_j is the Fisher score of feature j , $|T_C|$ is the number of training examples of class 'C', and p is the exponent which is set to -5, which has been empirically determined [2, 3].

4. Experimental Studies

We conducted experiments with IICBU-2008 benchmark dataset [7]. We used Intel Pentium i3 core processor with 2 GB RAM and 500 GB hard disk with 2.3 GHz clock speed. The features are extracted using wndcharm tool. The command used to extract features in wndcharm is [2] **Wndchrn train [options] images feature_file**. The resulting image feature values are store in **feature_file**, **images** is a path to the top folder that contain all the images of the data set and [options] field is for specifying the optional switches . Images belonging to different classes can be kept in separate directories inside the top folder. Wndchrn supports *tiff* and *ppm* image file formats.

We implemented and analyzed three variations of Bayesian classifiers. These variations are based on the approximations of covariance matrix of the training set. We named these classifiers as Bayesian-1, Bayesian-2 and Bayesian-3 respectively. The classification accuracies for these classifiers were experimentally determined for various sets of feature vectors. We implemented the SVM classifier using the tool SVM Torch [8].

The IICBU-2008 benchmark dataset [7] includes biological images of different subjects such as organelles, cells, tissues, and full organisms using different magnifications and different types of microscopy [5]. The details of the data sets are given in Table I. An indicative representation of images from IICBU-2008 dataset is given in Figure 2. From Table II we observe that, feature set 2 (transform features) give fairly good classification accuracy compared with the other feature sets for Bayesian-2 classifier. It is also observed that SVM classifier has better classification accuracy than Bayesian classifiers.

Table IV and Table V give details about the performance of Bayesian-2 and SVM classifiers for reduced set of features obtained using PCA and Fisher score based methodologies respectively for Cho dataset. It can be seen that the classification accuracies for Bayesian-2 and SVM based classifiers are almost remaining stable for considerable reduction of the number of features in the representation of

images. This is in fact a pleasant surprise from the point of view of the design of CBIR system for biological images. The images can be represented in terms of compact, most informative and discriminant features in the CBIR database there by reducing the computational complexity. This also helps to improve the retrieval speed of the CBIR system.

TABLE I
DETAILS OF THE DATASET CONSIDERED FOR EXPERIMENTS
FROM IICBU -2008 BENCHMARK DATA SET [9].

Sl. no	Dataset	No. classes	No. of Images
1	Binucleate	2	40
2	C.elegans	4	252
3	Cho	5	340
4	Hela	10	860
5	Terminal Bulb	7	970
6	Rnai	10	200

Figure 2. Representative images from IICBU 2008 dataset
(Source: <http://ome.grc.nia.nih.gov/iicbu2008>).

TABLE II

COMPARISON OF CLASSIFICATION ACCURACIES FOR DIFFERENT SETS OF FEATURE VECTORS FOR BAYESIAN -2 CLASSIFIER.

Feature set	Classification Accuracies (%) for datasets				
	Binucleate	Celegans	Cho	Hela	T.Bulb
Feature set 1	91.6	44.4	97.9	73.4	44.3
Feature set 2	100	62.5	97.9	53.9	43.9
Feature set 3	100	56.9	98.9	56.7	38.8
Feature set 4	100	59.7	97.9	59.3	39.1

TABLE III

COMPARISON OF CLASSIFICATION ACCURACIES FOR DIFFERENT SETS OF FEATURE VECTORS FOR SVM CLASSIFIER.

Feature set	Classification Accuracies (%) for datasets				
	Binucleate	Celegans	Cho	Hela	T.Bulb
Feature set 1	100	63.9	98.9	78.6	51.5
Feature set 2	100	72.2	98.9	78.7	51.9
Feature set 3	100	55.6	98.9	83.7	51.9
Feature set 4	100	75.0	97.9	77.5	49.1

TABLE IV

VARIATION OF CLASSIFICATION ACCURACIES FOR BAYESIAN -2 AND SVM CLASSIFIER WITH REDUCTION IN NUMBER OF FEATURES USING PCA FOR CHO DATASET.

Feature set	Percentage reduction of features after PCA	Classification accuracy for Bayesian-2 Classifier	Classification accuracy for SVM Classifier
Feature set 1	87.24	97.96	97.96
Feature set 2	55.30	97.96	98.98
Feature set 3	61.97	98.98	98.98
Feature set 4	93.2	97.96	98.98

TABLE V

VARIATION OF CLASSIFICATION ACCURACIES FOR BAYESIAN -2 AND SVM CLASSIFIER WITH REDUCTION IN NUMBER OF FEATURES USING FISHER SCORE FOR CHO DATASET.

Feature set	Percentage reduction of features after PCA	Classification accuracy for Bayesian-2 Classifier	Classification accuracy for SVM Classifier
Feature set 1	76.57	97.96	97.96
Feature set 2	43.6	97.96	98.98
Feature set 3	40.0	98.98	98.98
Feature set 4	80.98	97.96	98.98

It is observed that we can obtain a high degree of reduction

in the number of features using PCA and Fisher score based methods. This will surely reduce the computational overhead and increase the speed and overall performance of the CBIR systems.

We analyzed the performances of Bayesian-2, SVM classifiers for data sets belonging to IICBU-2008 biological image data set. We used Gaussian kernel for the SVM implementation. We also compared the results with the available results of wndchrm classifier [2]. The results are shown in Figure 3. It is observed that SVM classifier outperforms the other classifiers.

Figure 3 Performance Analysis of Bayesian-2, SVM and wndchrm classifiers for datasets belonging to IICBU-2008 Biological image dataset.

5. Conclusions

We compared classification accuracies of Bayesian, SVM and Wndchrm Classifier for biological image data set. It has been observed that SVM classifier that uses Gaussian kernel yielded highest classification accuracy. We experimentally verified that extracting subsets of features from the canonical set of feature vectors and by the use of feature selection methods using PCA and Fisher score, one can speed up the retrieval process of CBIR system without much degradation in performance.

Acknowledgment

This research project has been granted financial assistance under Seed Money Project Scheme of Technical Quality Improvement Programme (TEQIP) phase II under State Project Facilitation Unit (SPFU) of the state of Kerala, India.

References

- (1) Jiliang Tang, Salem Alelyani, and Charu Aggarwal
“Feature Selection for Classification: A Review”,
CRC Press.
- (2) Lior Shamir, Nikita Orlov, D. Mark Ekley, Tomasz Macura, Josiah Ohnston, and Ilya G. Goldberg,
“Wndchrm- An Open Source Utility for Biological Image Analysis”, Source Code for Biology and Medicine 2008, pp3-13.
- (3) Orlov N., Shamir L., Macura T., Johnston J., Eckley D. M., and Goldberg I., “WND-CHRM Multi Purpose Classification using Compound Image Transforms”, Pattern Recognition Letters in Press.
- (4) Santosh K.C., Bart Lamiroy and Laurent Wendling.
“DTW for Matching Radon Features: A Pattern Recognition and Retrieval Method”, Advanced Concepts for Intelligent Vision Systems 6915 (2011) 249–260
- (5) Shamir L., Orlov N., Eckley D.M., Goldberg I.G. IICBU 2008, A “Benchmark Suite for Biological Imaging”, 3rd Workshop on Bio-Image Informatics. Biological Imaging, Computer Vision and Data Mining; Santa Barbara, CA 2008:1240.
- (6) R. O. Duda and P. E. Hart and David G. Stork, “Pattern Classification”, 2nd Ed., John Wiley, and sons Inc New York 2001.
- (7) IICBU-2008, <http://www.ncbi.nlm.nih.gov/iicbu2008>.
- (8) A small user guide for SVM Torch. August 9, 2000.
- (9) Simon O. Haykin, “Neural Networks and Learning Machines”, Copyright 2009 by Pearson Education Inc., New Jersey.
- (10) Quanquan, G. Zhenhui, L. Jiawei H., “Generalized Fisher Score for Feature Selection”. In: Proceedings of the International Conference on Uncertainty in Artificial Intelligence (2011).