Classifer ensemble for an effective cytological image analysis

Paweł Filipczuk b, Bartosz Krawczyk a,*, Michał Woźniak a

a Department of Systems and Computer Networks, Wrocław University of Technology, Wrocław, Poland
b Institute of Control and Computation Engineering, University of Zielona Góra, Zielona Góra, Poland

ARTICLE INFO

Article history:
Available online 13 May 2013
Communicated by Manuel Graña

Keywords:
Pattern recognition
Machine learning
Multiple classifier system
Classifier ensemble
Trained fuser
Image analysis

ABSTRACT

Breast cancer is the most common type of cancer among women. As early detection is crucial for the patient’s health, much attention has been paid to the development of tools for effective recognition of this disease. This article presents an application of image analysis and classification methods for fine needle biopsy. In our approach, each patient is described by nine microscopic images taken from the biopsy sample. The images are related to regions of the biopsy that seem interesting to the physician who selects them arbitrarily. We propose four different hybrid segmentation algorithms dedicated to processing these images and examine their effectiveness for the nuclei feature extraction task. Classification is carried out with the usage of a classifier ensemble based on the Random Subspaces approach. To boost its effectiveness, we use a linear combination of the support functions returned by the individual classifiers in the ensemble. In the proposed medical support system, the final decision about the patient is delivered after a fusion of nine separate outputs of the classifier – each for a different image. Experimental results carried out on a diverse dataset collected by the authors prove that the proposed solution outperforms state-of-the-art classifiers and shows itself to be a valuable tool for supporting day-to-day cytologist’s routine.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

According to the International Agency for Research on Cancer, breast cancer is the most common cancer among women. In 2008, there were 1,384,155 diagnosed cases of breast cancer and 458,503 deaths caused by the disease worldwide (Ferlay et al., 2010; Bray et al., 2012). In 2010, there were 15,784 diagnosed cases of breast cancer among Polish women, with 5226 resulting in death (National Cancer Registry, 2012). There has also been an increase in the incidence of breast cancer by 3–4% a year since the 1980s. The effectiveness of treatment largely depends on the timely detection of the disease. An important and often used diagnostic method is the so-called triple-test, which is based on three medical examinations, and is used because of its effectiveness in diagnosing breast cancer (Britton et al., 2009). The triple-test includes self examination (palpation), mammography or ultrasonography imaging, and fine needle biopsy (FNB) (Underwood, 1987). FNB is an examination that consists of obtaining material directly from the tumor. The collected material is then examined under a microscope to determine the prevalence of cancer cells.

This approach requires extensive knowledge and experience of the cytologist responsible for the diagnosis. Automatic morphometric diagnosis can help make the results objective and assist inexperienced specialists. Along with the development of advanced vision systems and computer science, quantitative cytopathology has become a useful method for the detection of diseases, infections, and many other disorders (Gurcan et al., 2009; Śmietanka and Tadeusiewicz, 2010; Hassan et al., 2010).

Recently, there has been an increase in interest in computer-aided cytology. Several researchers have studied the segmentation of cytological images of breast tumors, proposed new features or classification algorithms. However, only a few of these researchers have tested the efficiency of their methodology in a comprehensive computerized cancer classification system. Jeleń et al. (2010) presented an approach based on the level sets segmentation method. Classification efficiency was tested on 110 (44 malignant, 66 benign) images with results reaching 82.6%. Niwas et al. (2010) presented a method based on the analysis of nuclei texture using a wavelet transform. Classification efficiency with the k-nearest neighbor algorithm on 45 (20 malignant, 25 benign) images reached 93.33%. Another approach was presented by Malek et al. (2009). They used active contours to segment nuclei and classified 200 (80 malignant, 120 benign) images using fuzzy C-means algorithm achieving 95% efficiency. Breast cancer diagnosis was also discussed by Xiong et al. (2005). Partial least squares regression was used to classify 699 (241 malignant, 458 benign) images...
yielding 96.57% efficiency. However, the authors did not describe the segmentation method used to extract nuclei.

This paper presents recent progress in the development of a comprehensive fully automatic breast cancer diagnostic system based on analysis of cytological images of FNB material. The task at hand is to classify a case as benign or malignant. Recently in our research, we have introduced the third class – fibroadenoma. Fibroadenoma is a benign tumor of the breast that often occurs in women. Despite the fact it is not cancerous, it might have some morphometric features similar to malignant neoplasm. This may confuse the system and cause an incorrect diagnosis. The diagnosis is done by using the morphometric and topological features of nuclei isolated from microscopic images of the tumor. The segmentation is based on a set of four hybrid segmentation methods combining adaptive thresholding and clustering (see Section 4).

At present the classifier ensembles, known also as multiple classifier systems (MCS) or combined classifiers (Kuncheva, 2004), are the focus of intense research (Jain et al., 2000) because usually we may have more than one classifier dedicated to a given problem at our disposal. Since each classifier has its own domain of competence (Wolpert, 2001), designing a method that can exploit strengths of individual predictors while preventing us from choosing the worst model seems to be an attractive proposition (Kurzynski and Woźniak, 2012).

The process of building such a compound classifier is not trivial, and there are several problems that must be considered carefully as classifier selection and choosing an appropriate method of classifier combination. The first task is known as the classifier selection (Ruta and Gabrys, 2005) or ensemble pruning (Martinez-Munoz et al., 2009), and we would like to assure that the chosen classifiers are characterized by the high accuracy and diversity because we expect them to be complimentary. A key issue is how to measure classifier diversity (Brown and Kuncheva, 2010) and how to select classifier using the proposed measure. The second problem is how to combine individual classifiers’ outputs. The first group of methods has their origin in voting algorithms (Biggio et al., 2007). For many years, a standard majority voting was the most widespread approach. In recent years, more advanced approaches have been proposed that take into account not all decisions coming from particular committee members should influence the voting procedure in an identical way (van Erp et al., 2002). Here, we should mention those works that propose training the weights, which seems to be an attractive alternative method, often outperforming static weight assignment (Woods et al., 1997; Woźniak and Jackowski, 2009). More advanced propositions use support functions of individual classifiers. Their main form is the posterior probability typically associated with probabilistic pattern recognition models, although outputs of neural networks or other functions whose values are used to establish the decision of the classifier could be considered as well. Aggregation methods that do not require learning use simple operators, like minimum, maximum, product, or mean. However, they are typically subject to very restrictive conditions (Duin, 2002), which limit their practical use. Therefore, the design of new fusion classification models, especially trained fusers, are currently the focus of intense research (Woźniak and Krawczyk, 2012).

In this work we will focus on the weighted combination of support functions, and where weights depend on a given classifier and class number. As shown in the previous works of authors (Woźniak and Zmyslony, 2010), this type of fusion achieves fairly good quality and it can be easy trained solving simple optimization task. Classifiers ensembles are nowadays widely used in the medical decision support (Jackowski et al., 2012). This paper shows the continuations of authors work on early breast cancer detection (Krawczyk et al., 2012a,b). For the presented task of fine needle biopsy image classification, we employ a MCS build on the basis of the Random Subspace approach (Ho, 1998) to assure the initial diversity among base classifiers. As it is common knowledge that not all predictors build on the basis of this method should take an identical part in the final decision we propose to combine their outputs with a novel trained fuser based on the discriminants. This way we achieve better combination results than when using traditional fusers. Exhaustive computational tests proves that our ensemble outperforms canonical classifier committees.

The proposed ensemble is used to classify independently all nine images representing the examined patient. Then, a final decision about the state of the patient is made on the basis of these individual ones. Two heuristic approaches are proposed for this problem. The results presented in this paper demonstrate that a computerized medical diagnosis system based on our method would be effective and can provide valuable and accurate diagnostic information.

The paper is divided into six sections. Section 1 presents an introduction into breast cancer diagnosis. Section 2 describes the acquisition process of the medical images used for testing. A framework for the proposed medical decision support system is introduced in Section 3. Segmentation and feature extraction are described in Section 4, while Section 5 presents the classification algorithm. Section 6 shows the experimental results obtained by the proposed method. The paper ends with our conclusions.

2. Medical images database

All methods presented in this work were tested on real medical data. For this purpose, 675 images were collected from 75 patients (25 benign, 25 malignant and 25 fibroadenoma). Each patient is represented by 9 images selected by a pathologist arbitrarily. The number of images was recommended by the specialists from the hospital and allows for their correct diagnosis.

The cytological material was obtained by FNB from patients of the Regional Hospital in Zielona Góra, Poland. Biopsies without aspiration were performed under the control of an ultrasonograph with a 0.5 mm diameter needle. Smears from the material were fixed in spray fixative (Cellfix by Shandon) and dyed with hematoxylin and eosin (H&E). The time between preparation of smears and their preservation in fixative never exceeded three seconds. All cancers were histologically confirmed and all patients with benign disease were either biopsied or followed for a year.

The images were recorded by a Sony CCD Iris color video camera mounted atop Axioshot microscope. The slides were projected into the camera with 160× objective and 2.5× ocular giving together an enlargement of 400×. Images are BMP files, 704 × 578 pixels, 8 bit/channel RGB.

3. Framework for patient classification

In the proposed diagnostic procedure, each patient is characterized by nine separate cytological images. The physician examines all of these images and on the basis of her/his experience gives a final diagnosis. One should note that these images are not correlated to each other, because they are chosen according to arbitrary decision of the cytologist i.e., the first image from the i-th patient may describe a completely different area of the biopsy sample than the first image from the j-th patient.

In the described medical decision support system, we decided to preserve this schema. We propose to train the classifier on the basis of all images available in the training set and then use it to produce nine independent decisions about the images belonging to a tested patient. On the basis of these individual outputs, an aggregation procedure is being conducted that produces a final
decision about the state of the examined person. The schema of such a framework is presented in Fig. 1.

Let us present the advantages of the proposed approach:

- The final decision is made on the basis of several images. This prevents a situation where the image shows an unrepresentative part of the biopsy sample. The cancerous cells may be distributed differently along the sample, and referring to only a single image may easily produce a false diagnosis – which is especially dangerous if cancerous cells get overlooked. We propose decision rules for combining the results from individual images (see Section 5.2 for details).
- One classification algorithm that is trained on the data coming from all the images improves the competence area of the recognition system. Cancerous cells have similar properties, so even if the tested image came from a completely different part of the biopsy sample, the generalization properties of the system will allow for a precise classification.
- By using an fusion procedure on the individual outputs for nine images, we get a more elastic system that may reduce the number of wrong diagnosis for the patients, even if some of the individual images were misclassified.
- The classifier module can be filled by any model. In our approach, we propose an efficient MCS for this task that will be described in detail later in this section.

4. Nuclei segmentation and feature extraction

To determine the type of tumor, nuclei need to be isolated from other objects (e.g., red blood cells) and the background. Then from the nuclei, certain features can be extracted (Nguyen et al., 2012) and the classification of either benign, fibroadenoma, or malignant can be determined. In the literature, many different approaches have been proposed to extract cells or nuclei from microscope images (Al-Kofahi et al., 2010; Clocksin, 2003; Cloppet and Boucher, 2008; Hrebien et al., 2010; Obuchowicz et al., 2008; Peng et al., 2010). This task is usually done automatically or semi-automatically (Plissiti et al., 2011), using one of the well known methods of image segmentation (Gonzalez and Woods, 2001; Naz et al., 2010; Suri et al., 2002).

In this work, we decided to use a segmentation method based on the combination of adaptive thresholding in grayscale and pixel classification in color space using one of four clustering algorithms: K-means, fuzzy C-means, Gaussian mixture model, and competitive neural networks. The approach is presented in our previous papers (Filipczuk et al., 2011a,b; Kowal et al., 2011a,b).

First, the RGB image is converted to grayscale by eliminating the color information while retaining luminance

\[ L = 0.2126R + 0.7152G + 0.0722B. \]  

Then, adaptive thresholding is employed to separate objects from the background. In the considered approach, the threshold value is calculated for each pixel and equals the mean intensity of pixels from the neighborhood defined by a square of size 21 × 21 pixels. The result is relatively well isolated individual objects. Unfortunately, erythrocytes and nuclei have similar luminance, so their distinction is difficult. To tackle this drawback, we use clustering performed in RGB. The idea of image segmentation using clustering algorithms boils down to a search for clusters of pixels in color space. We assume the presence of three types of objects defined respectively by their color: nuclei, red blood cells, and background. Clustering allows to distinguish nuclei from erythrocytes and the background, but obtained objects tend to merge and their edges are inaccurate. In order to combine advantages of both methods, results are fused together by multiplication of resultant images of both algorithms. Finally, objects that are either too small or on the edges of the image are removed. Fig. 2 summarizes the segmentation process. All clustering methods and feature extraction are described in detail later in this section. Fig. 3 shows a sample segmentation result.

4.1. K-means

The clustering procedure of the K-means (KM) algorithm is based on minimizing the within-cluster sum of squares for K clusters:

\[ J = \sum_{k=1}^{K} \sum_{y=1}^{Y} \sum_{x=1}^{X} \mu_{xy,k} D_{xy,k}^2, \]

where X and Y defines the size of the analyzed image, \( \mu_{xy,k} \) is a function specifying which (x,y)th pixel belongs to the kth cluster, and \( D_{xy,k}^2 \) is an Euclidean distance measure:

\[ D_{xy,k}^2 = (c_{xy} - \mu_k)^T(c_{xy} - \mu_k), \]

where \( c_{xy} \in \mathbb{R}^3 \) is a vector of the coordinates of the (x,y)th pixel in RGB space and \( \mu_k \in \mathbb{R}^3 \) is a vector of the coordinates of the kth cluster center in RGB space. The K-means clustering procedure iteratively change pixel assignments based on the distance to the nearest mean (cluster center) and updates the cluster centers to match the proper means of the data points that they are responsible for. Detailed expressions for iterative updating the cluster centers.

![Fig. 1. The proposed framework for patient classification based on the information fusion coming from nine individual cytological images.](image-url)
and pixel assignments can be found in the following papers (Kanungo et al., 2002; Lloyd, 1982; MacKay, 2003).

4.2 Fuzzy C-means algorithm

The clustering procedure of the fuzzy C-means (FCM) algorithm is based on finding the local minimum of the non-linear cost function given by Eq. (4) using the Picard iteration through first order conditions for stationary points:

\[
J(U, V) = \sum_{x=1}^{X} \sum_{y=1}^{Y} \sum_{k=1}^{C} \mu_{xy}^{m} f(c_{xy}, v_{k}),
\]

where the matrix \( U \in \mathbb{R}^{X \times Y \times C} \) contains the membership degrees of pixels to the defined clusters, \( v_{k} \in \mathbb{R}^{d} \) is a vector of the coordinates of the \( k \)th cluster center, \( X \) and \( Y \) defines the size of the analyzed image, \( f_{c} \) is a function used to determine the distance between the data points and cluster centers, \( \mu_{xy}^{k} \) is the membership of the \((x,y)\)th point to the \( k \)th cluster.
pixel in the fuzzy cluster \( k \), \( m \in (1, \infty) \) is the fuzziness of the clustering procedure, and \( \mathbf{e}_{xy} \) is a vector of \((x, y)\)th pixel parameters describing its color. The function \( f_k \) defines the metric used to determine the distance between the data points and cluster centers:

\[
 f_k(\mathbf{e}_{xy}, \mathbf{v}_k) = \| \mathbf{e}_{xy} - \mathbf{v}_k \|^2 = (\mathbf{e}_{xy} - \mathbf{v}_k)^T \mathbf{A}(\mathbf{e}_{xy} - \mathbf{v}_k). \tag{5}
\]

The matrix \( \mathbf{A} \) from the expression (5) is used to tune the shape and orientation of the clusters in space. In the simplest approach, the matrix \( \mathbf{A} \) is unitary, thus the distance measure \( f_k(\mathbf{e}_{xy}, \mathbf{v}_k) \) is an Euclidean distance. In this case, the study metric is defined as Euclidean norm. In this case, the study metric is defined as Euclidean norm.

Detailed expressions for iterative updating of the means and covariances can be found in the following paper (Bezdek, 1981). In order to obtain the final result, the membership to the cluster representing the nuclei needs to be sharpened. This task is done by thresholding, with a threshold arbitrary chosen to 0.8. Fig. 4 shows a sample membership to the clusters.

### 4.3. Competitive learning neural networks

In the considered case pixel affiliation to background, nuclei or red blood cells is determined using a competitive learning neural network (CLNN) with three neurons, each representing a cluster in color space corresponding to a given type of object. The network was trained using Kohonen learning rule (Dong and Xie, 2005). Each neuron has three inputs interpreted as RGB values. Learning samples (RGB values) were manually selected parts of the original images. Kohonen learning rule assumes that only the weights of the winning neuron will be adjusted. The winner is the neuron that best represents the input. The update procedure is given by the following formula:

\[
 w^{\text{win}}_i(k+1) = w^{\text{win}}_i(k) + \alpha(p_i(k) - w^{\text{win}}_i(k)), \tag{6}
\]

where \( w^{\text{win}}_i(k) \) is \( i \)th weight of the winner neuron during \( k \)th learning pass, \( p_i(k) \) is the \( i \)th input value, and \( \alpha \) is the learning rate. Fig. 5 shows a sample membership to the clusters.

### 4.4. Gaussian mixture model

In this work, the Gaussian mixture model (GMM) is built with three normal distribution components: \( N1(\eta_1, \Sigma_1) \), \( N2(\eta_2, \Sigma_2) \), and \( N3(\eta_3, \Sigma_3) \), using pixels from the RGB image. The components represent background, nuclei, and red blood cells. GMM parameters \( \theta = [\eta_1, \eta_2, \eta_3, \Sigma_1, \Sigma_2, \Sigma_3] \), where \( \eta_i \) represents means and \( \Sigma_i \) covariances, were estimated using the expectation maximization algorithm. This algorithm is well known in statistics, and detailed expressions for iterative updating of the means and covariances can be found in Dempster et al. (1977). The same data that was used to create the GMM is next partitioned based on the largest posterior probability computed for each pixel and GMM component.

### 4.5. Feature extraction

After isolation of nuclei from the images, 16 features are extracted from each slide and used in the classification procedure. First, for each nucleus we calculate all 8 features, which are described below. Then, for each of the 8 features, the mean and variance are determined, giving a total of 16 features. Below is a detailed description of all features:

- **Area** – The actual number of pixels of the nucleus.
- **Perimeter** – The distance between each adjoining pair of pixels around the border of the nucleus.
- **Eccentricity** – The scalar that specifies the ratio of the distance between the foci of the ellipse that has the same second-moments as the segmented nucleus and its major axis length.

---

Fig. 4. Original image (upper-left) and the membership to the three clusters obtained using fuzzy C-means algorithm (the brighter the stronger membership).
Major Axis Length – The length of the major axis of the ellipse that has the same normalized second central moments as the nucleus.

Minor Axis Length – The length of the minor axis of the ellipse that has the same normalized second central moments as the nucleus.

Luminance Mean – The mean luminance of pixels of the nucleus.

Luminance Variance – The variance of luminance of pixels of the nucleus.

Distance to the Centroid of all Nuclei – The distance between the geometric center of the nucleus and the centroid of all nuclei.

5. Classification

In this section, the proposed classification scheme, employed as a part of the presented medical decision support system, is presented in detail.

5.1. Proposed multiple classifier system

Let us now present in detail the proposed classifier ensemble dedicated to this task.

5.1.1. Pool of base classifiers

The pool is a set of n base classifiers:

$$\Pi^\Psi = \{ \Psi_1, \Psi_2, \ldots, \Psi_n \},$$

built on the basis of the Random Subspace approach (Ho, 1998). In this work, we assume that the pool is homogeneous i.e., consists of classifiers built on the basis of the same model.

The performance of the Random Subspace method is strongly affected by the size of subspaces. Ones consisting of too many features will lead to creating a pool of similar models, while the ones consisting of few features will lead to creating too weak classifiers. There are no a priori rules on how to select the size of the subspaces – this must be fitted by the user to the problem at hand. There is an additional property of this method that can be seen as a cause for concern – due to its randomness the classifiers will vary significantly in their individual accuracies. Therefore, it is important to boost the influence of good models on the final decision. For this, we employ a novel trained fuser based on discriminants.

5.1.2. Trained fuser

Let’s assume that we have n classifiers $$\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(n)}$$ at our hand. For a given object $$x \in X$$, each individual classifier decides whether it belongs to class $$i \in M = \{1, \ldots, M\}$$ based on the values of discriminants. Let $$F^{(l)}(i, x)$$ denote a function that is assigned to class i for a given value of x, and that is used by the lth classifier $$\Psi^{(l)}$$. The combined classifier $$\Psi$$ uses the following decision rule (Jacobs, 1995)

$$\Psi(x) = i \quad \text{if} \quad \tilde{F}(i, x) = \max_{k \in M} \tilde{F}(k, x),$$

where

$$\tilde{F}(i, x) = \sum_{l=1}^{n} w_l^{(l)} F^{(l)}(i, x) \quad \text{and} \quad \sum_{l=1}^{n} w_l^{(l)} = 1.$$

We propose to use the weighted combination of support functions, where weights depend on class and classifier, because the previous works of authors proved that is a valuable and effective proposition (Wozniak and Zmyslony, 2010). Such a combination rule can be implemented as one layer perceptron, which may be trained with any standard procedure used in neural network learning; the input weights established during the learning process are then the weights assigned to each of the base classifiers. The implementation of the proposed fuser is presented in Fig. 6.

One should bear in mind that in order for the presented trained fuser to work, the ensemble must consist of classifiers that are able to output support functions as Naive Bayes, Neural Networks, or Support Vector Machines.

Fig. 5. Original images (left) and determined clusters using competitive neural network (right).
In our research, such a fuser is used independently for all of the nine images. In order to combine nine outputs returned by the linear combination method, we use one of the two proposed aggregation methods described in the next section.

5.2. Final decision rule

An important part of the proposed decision support system is how to combine the individual decisions made separately for the images into a final prediction for the patient. In order to provide a high quality output, we tried to mimic the behavior of experienced physicians. After observations and discussions, we have formulated two heuristic approaches for this task:

- **Majority** – where the patient is classified to the class pointed out by the majority of the decisions made on the individual images. One should bear in mind that there is some probability (although low) of the occurrence of a tie, as we have nine individual decisions and three possible classes. In such a case, the final decision points to a malignant case, as this gives the lowest risk for the patient health (overlooking is more dangerous). This approach is based on the opinions of physicians who state that the case can be classified as malignant only if the cancerous cells are present in most parts of the biopsy sample.

- **Singular** – where the patient is classified to the malignant class if at least one decision made on the individual images points to it. In case of no outputs pointing out to the malignant case, the dichotomization between benign and fibroadenoma classes is carried out via majority procedure. This approach is based on the opinions of physicians who state that even a single observation of cancerous cells should be treated as a malignant case for early prevention.

6. Experimental results

6.1. Aims of the experiment

The aim of the experiment was to examine the usefulness of the proposed hybrid segmentation methods and ensemble with trained combination method for the cytological image analysis. The main goals of the investigations are listed below:

- To examine the usefulness of the segmentation methods for the process of feature extraction from biopsy images by checking their discriminant abilities.
- To check the quality of the proposed MCS based on Random Subspaces and compare it with several models popular in medical informatics.
- To validate the quality of the proposed linear combiner trained with a neural computation approach and check if the more complex design will repay with a boost of recognition accuracy for the ensemble.
- To compare the two approaches for diagnosing the status of the patient on the basis of nine individual images and see if there are any clinically significant differences in their quality.

6.2. Set-up

As a base classifier for the proposed MCS, a Support Vector Machine with the Radial Basis Function Kernel, trained with the SMO procedure, was applied.

The ensemble was trained with 10 base classifiers, each built on the basis of a feature subspace consisting of 60% of original feature space. This parameter value was established experimentally through a grid search.
Table 1

<table>
<thead>
<tr>
<th>Segmentation</th>
<th>SVM</th>
<th>Bagg</th>
<th>Boost</th>
<th>MV</th>
<th>MAX</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Majority 1</td>
<td>Majority 2</td>
<td>Majority 3</td>
<td>Majority 4</td>
<td>Majority 5</td>
<td>Majority 6</td>
</tr>
<tr>
<td>K-means</td>
<td>76.00</td>
<td>78.66</td>
<td>80.00</td>
<td>81.33</td>
<td>84.00</td>
<td>85.33</td>
</tr>
<tr>
<td>Fuzzy C-means</td>
<td>80.00</td>
<td>81.33</td>
<td>82.67</td>
<td>82.67</td>
<td>85.33</td>
<td>85.33</td>
</tr>
<tr>
<td>Competitive learning</td>
<td>82.67</td>
<td>84.00</td>
<td>86.67</td>
<td>86.67</td>
<td>90.66</td>
<td>90.66</td>
</tr>
<tr>
<td>Gaussian mixture</td>
<td>81.33</td>
<td>82.67</td>
<td>84.00</td>
<td>84.00</td>
<td>88.00</td>
<td>88.00</td>
</tr>
</tbody>
</table>

To put the results into context, we compared our proposed ensemble with: a single SVM, a bagged SVM, and a boosted SVM. Bagging consisted of 15 bags, while boosting was run for 20 iterations. These were the parameter values returning best performance for these models, established via the grid search. Additionally, we have presented the results for ensembles built from an identical pool of classifiers as our proposed one, but using different fusers – majority voting and max operator.

All experiments were carried out in the R environment (Team, 2008), with classification algorithms taken from the dedicated packages, thus ensuring that the results achieved the best possible efficiency and that the performance was not decreased by a bad implementation. All tests were done by a 10CV F-test (Alpaydin, 2010).

6.3. Results of the experiment

The results of the experiment are presented in Table 1. Bagg stands for bagged SVM, Boost for boosted SVM, MV for the proposed pool of classifiers with a majority voting fusion, MAX for the same pool with a max operator fusion, and Proposed for the MCS presented in Section 5.1.

6.4. Results and discussion

From the experimental comparison, we can clearly see that the proposed MCS based on the Random Subspaces and linear combiner trained with a neural computation approach always outperformed other reference methods. Single SVM always delivered the worst performance, not being able to handle the compound data. Bagging and Boosting lead to an increase in accuracy in comparison to a single-model approach and performed similar (or sometimes better) to a Random Subspace with majority voting. Interesting conclusions arise when comparing the fusion algorithms applied to the pool of classifier build on the basis of the Random Subspace. Majority voting always came last, which can be explained by the randomness in the classifier creation process, as it may have led to creation of some weaker classifiers that destabilized the voting procedure. Using simple max operator based on discriminants returned a significant improvement in the quality. The proposed trained fuser always turned out as the best solution, proving that additional computational cost spent on its training procedure returned a satisfactory boost of the performance of the proposed MCS. Trained fuser seems to work very well with the Random Subspace approach, allowing the MCS to explore valuable classifiers while reducing the influence of those who were built on an unfavorable random selection of features.

As for the four proposed segmentation methods, the features with the most discriminative power were extracted through a competitive learning neural networks approach. This lead to a high recognition rate (96.00%) of the patients’ conditions. Simple K-means algorithm was not suitable for this task. Gaussian mixture returned satisfactory results, but features extracted through it had a slightly lower discriminative power than the ones extracted by the best method. An interesting observation can be made for the fuzzy C-means approach. For standard classification methods, its performance is not promising, but when coupled with the proposed ensemble and trained fuser, the accuracy increased rapidly, returning identical quality as Gaussian mixture algorithm. This shows that a carefully designed fuser may in some part compensate for weaker classifiers in the pool.

Finally, the aggregation method did not have a crucial impact on the quality of the proposed medical decision support system – quite contrary to what we were expecting. In most cases, there were no differences between the majority and singular approaches. In some cases there were small but statistically significant differences in favor of the singular method. This combined with the fact that the singular approach have never performed worse than the majority one allows us to draw a conclusion that although the differences are very small, the approach based on classifying a patient to the malignant class if at least one decision made on the individual images points to it, is a more preferable solution for the system discussed.

7. Conclusions

The paper proposed a novel framework for a medical decision support system dedicated to breast cancer recognition from biopsy images. We have discussed an approach that examines the patient on the basis of nine individual images, and then makes the final decision by aggregating the individual outputs of the classification module. For the evaluation purposes we use a wide collection of real-life images gathered over the recent years by authors.

We have proposed four different methods dedicated to the segmentation of nuclei cells that later served for the feature extraction process. One of the aims of the paper was to establish which one of them is most suitable for the practical implementation. Competitive learning neural networks, although computationally costly, returned a best results and features extracted on its basis contributed to the lowest classification error.

When having at our disposal nine different images assigned to a single patient, one should use all the available data to give the final diagnosis. Therefore, we tried to mimic the line of thought of an experienced physician and proposed two heuristic solutions for information aggregation from individual images. To our surprise, there were no significant differences between them. In some cases, there were small improvements when using an approach based on labeling patient as a malignant case when at least one of the individual images was labeled as such.
Finally, we have introduced a MCS based on the Random Subspaces and a novel trained fuser. The implementation of such a fuser was motivated by the fact that not all of the base classifiers created by the Random Subspace method are of the same quality – while it is a good method for assuring diversity it cannot always ensure the high individual accuracy. By using a trained fuser, we controlled the influence that each base classifier had on the final decision. This resulted in a high quality ensemble, which outperformed other committees, commonly used in the medical decision support task.

Our system returned a very high accuracy and overall satisfactory performance. The results presented in this paper demonstrate that a computerized medical decision support system based on our method can provide valuable and accurate diagnostic information.

In our future work, we would like to concentrate on the fusion of information coming from the different segmentation methods, as it is possible that such a combined feature set coming from the outputs of several segmentations may further improve the performance.

Acknowledgements

The authors thank Dr. Roman Monczak from the Regional Hospital in Zielona Góra, Poland for his great help and interesting discussions and Tracy Murcheski for her linguistic assistance in preparing this article.

Bartosz Krawczyk and Michał Wozniak are supported by the Polish National Science Centre under the grant NNS19650440, which is being realized in years 2011–2014.

Pawel Filipczuk is a scholar within Sub-measure 8.2.2 Regional Innovation Strategies, Measure 8.2 Transfer of knowledge, Priority VIII Regional human resources for the economy Human Capital Operational Programme co-financed by the European Social Fund and by state budget.

This research was partially supported by the National Science Centre in Poland.

References


