Cirrhosis classification based on MRI with duplicative-feature support vector machine (DFSVM)

Liu Hui\textsuperscript{a,},*, Guo Dong Mei\textsuperscript{b}, Liu Xiang\textsuperscript{c}

\textsuperscript{a} Department of Biomedical Engineer, Dalian University of Technology, Dalian 116024, China
\textsuperscript{b} Department of Radiology, Second Affiliated Hospital, Dalian Medical University, Dalian 116027, China
\textsuperscript{c} School of Materials Science and Engineering, Dalian Jiaotong University, Dalian 116028, China

A R T I C L E   I N F O
Article history:
Received 3 December 2012
Received in revised form 4 February 2013
Accepted 7 March 2013
Available online xxx

Keywords:
Cirrhosis detection
Computer-aided diagnostic (CAD)
Duplicative-feature support vector machine
Feature selection
Classification

A B S T R A C T
Magnetic resonance imaging (MRI) is a sensitive diagnostic method in improving the diagnostic capacity for hepatic cirrhosis and determining the accurate characterization of hepatic cirrhosis. But hepatic MRI has some shortcomings in detection and classification hepatic cirrhosis in clinical, especially using non-enhanced MRI for diagnosing early hepatic cirrhosis. And computer-aided diagnostic (CAD) system, including quantitative description of lesion and automatically classification, can provide radiologists or physicians an alternative second opinion to efficiently apply the abundant information of the hepatic MRI. However, it is expected to characterize comprehensively the lesion and guarantee high classification rate of CAD system. In this paper, a new CAD system for hepatic cirrhosis detection and classification from normal hepatic tissue non-enhanced MRI is presented. According to prior approach, six texture features with different properties based on gray level difference method are extracted from regions of interest (ROI). Then duplicative-feature support vector machine (DFSVM) is proposed for feature selection and classification: Firstly, the search process of DFSVM imitates diagnosis of doctors: doctor will take a more feature for consideration until the final diagnoses regardless of whether the feature is used in advance. So our algorithm is consistent with the process of clinical diagnosis. Secondly, the impact of the most valuable features will be well strengthened and then the high prediction performance can be got. Experimental results also illustrate the satisfying classification rate. Performance of extracted features and normalization are studied. And it is also compared with typical classifier ANN.

© 2013 Published by Elsevier Ltd.

1. Introduction

Hepatic cirrhosis increases the risk of hepatocellular carcinoma and is the leading causes of death by disease. Hepatic cirrhosis is characterized by the presence of widespread fibrosis and regenerative nodules in the liver [1]. And the fibrosis and nodules formation causes distortion of the normal hepatic architecture, resulting in characteristic texture patterns. Examining hepatic cirrhosis has been performed with various medical imaging modalities in clinical, such as ultrasonography (US), computer tomography (CT), or magnetic resonance imaging (MRI) [2]. MRI is more sensitive than CT and US in improving the diagnostic capacity and determining the accurate characterization of hepatic cirrhosis [3]. However, information obtained from MRI is of quite a variety, and it is difficult for radiologists to interpret MR images of the hepatic tissue, especially by non-enhanced MRI. Computer-aided diagnostic (CAD) system, as an alternative second opinion, has become one of the major research subjects in medical imaging, including MRI, CT, and ultrasound imaging [3]. So recent years, the evolution in medical image processing and artificial intelligence technologies has provided researchers the opportunity to apply CAD systems to classify hepatic tissue based on MRI [2,4,5].

Nowadays a few researches concerning hepatic cirrhosis with MRI and CAD are reported as MRI has just been popular in clinical recently. Texture analysis is the primary method at present and usually based on region-of-interest (ROI). Cogbert Lee has applied unsupervised classification to distinguish cirrhosis from non-cirrhotic hepatic tissues [1,6] based on texture features. And the area under the curve (AUC) obtained by unsupervised classification is 0.704. However, as labels (diseased or non-diseased) can provide beneficial prior knowledge, supervised classification is another helpful method. Zhang [7] integrates the shape and texture features of the hepatic tissue and applied a three-layer feed-forward artificial neural network (ANN) for classification. In testing of the whole hepatic tissue regions, 82% cirrhosis and 100% normal cases were correctly differentiated from 18 test cases.

* Corresponding author. Tel.: +86 0411 84706009x3015; fax: +86 0411 84706706.
E-mail addresses: liuhui@dlut.edu.cn, liuhui118@163.com (L. Hui),
drguangdong@163.com (G.D. Mei), luxiang@dlut.edu.cn (L. Xiang).

1746-8094/5 – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.bspc.2013.03.001

Please cite this article in press as: L. Hui, et al., Cirrhosis classification based on MRI with duplicative-feature support vector machine (DFSVM), Biomed. Signal Process. Control (2013), http://dx.doi.org/10.1016/j.bspc.2013.03.001
CAD for hepatic cirrhosis always consists of a two-step process: the first step involves feature extraction and effective feature selection, and the second step is to train classifier based on the selected features and identify cirrhosis. In feature extraction step, texture analysis techniques have been proved to be critical [1,6–9]. As each feature reflects different property of the medical image, we found that it is beneficial to efficiently combine them based on feature selection. Typically, feature selection refers to remove the irrelevant and duplicative features and increase the classification accuracy. Then duplication is usually been regarded as garbage feature and useless in classification. However, in our study we find that duplication can be used as an effective method to improve the classification performance. It is also studied that duplication cannot be simply reduced to a weighted feature.

And as for classification, ANN is one of a powerful computational tool for learning the relationship between the input, such as clinical parameters and radiological findings, and the output, such as interstitial diseases [4,7,10]. Although NN has many successful examples in practice, its generalization ability is an important bottle-neck. And NN is really sensitive to parameter changes, which make it cost long time for stable results. In recent years, since support vector machine (SVM) is capable of generating a hyper plane to separate two classes and providing good generalization, it has become the most popular machine learning classifier in biomedical engineering [11–13].

In this paper, six texture features of non-enhanced MRI hepatic tissues are applied in feature extraction step. However, each feature reflects different property of the image and has different effects on classification. Therefore, an efficient feature extraction and feature selection method is required to get the minimum possible error in the classification. Numbers of study focus the feature selection problem in pattern classification area [14–19]. A common method of feature selection is sequential feature selection, which is to build up a suboptimal feature subset incrementally starting from the empty set (forward selection: SFS) or to start with the full set (backward selection: SBS) of features and remove duplicative features successively. In this paper, a new sequential features selection method named duplicative-feature SVM (DFSVM) is proposed. Unlike classical sequential feature search approaches, DFSVM not only choose important features but also strengthen the impact of the most valuable ones. So DFSVM can improve the classification performance, which is proven by our experiments. And the generalization of SVM also guarantees high sensitivity and specificity of our algorithm. Then, a new CAD system for detection hepatic cirrhosis from normal hepatic tissue MRI is presented and DFSVM algorithm is proposed for classification. To compare with typical classifier ANN is also studied.

The rest of this paper is organized as follows. Section 2 describes related medical image and machine learning approaches, which includes support vector machines, feature extraction. And the DFSVM is proposed to classify hepatic tissues. Section 3 presents the experiment results.

2. Material and methods

2.1. Image acquisition

The MR images are collected from Second Affiliated Hospital of Dalian Medical University. 18 cases are obtained, including 10 patients with cirrhosis and 8 normal, which are all diagnosed in clinics or laboratories. All cases are scanned by a 1.5-T T MR scanner (Signa, GE, USA). The parameters of T1-weighted FSPGR sequence: TR = 1750.0 ms, TE = 4.2 ms; 256 × 128 matrix; 8 mm slice thickness and 2 mm inter-slice gap. Considering the diffuse distribution of hepatic tissue, large blood vessels within the hepatic tissue are excluded. And functionalized contrast agent is not used as we want to explore the performance of plain scan T1 image. Then regions of interest (ROIs) are manually segmented by an experienced radiologist, and 181 ROIs are obtained. Among those, 110 (60 cirrhosis and 50 normal) are picked out as training samples and 61 (39 cirrhosis and 32 normal) from different individuals as independent testing samples. A close-up of ROI including cirrhosis and normal samples is shown in Fig. 1. On close visual inspection, both cirrhosis and normal ROI might be bright or dark and their gray values are much close. However, only gray values are not enough to differ the two classes, texture features are needed of cirrhosis ROIs.

2.2. Feature extraction for ROIs

According to clinical information, normal hepatic tissue appears as delicate texture and uniform medium gray level in T1 image, while cirrhosis appears as coarse particles or diffuse small nodular with saltatory gray. Based on image processing theory, texture is defined as the spatial interrelationships and arrangement of the basic pixels of an image. And the gray level difference method (GLDM) [8] is a very powerful method for statistical texture description in medical imaging, especially for ultrasonic, MR and CT image analysis. Then GLDM is adopted for describing the collected image.

For each ROI, the texture features are extracted based on the estimation of the joint probability density function p(i, j) for two pixels of gray i and j with distance d in direction specified by the angle θ. In this paper, p(i, j) is considered in d = 1 and θ = 0. Six texture features are adopted in our CAD system based on GLCM, including F1: angular second moment (Asm), F2: contrast (Con), F3: correlation (Cor), F4: inverse difference moment (Idm), F5: entropy (Ent), F6: variance (Var), which are shown in following equations:

\[ F1 = \text{Asm} = \sum_{i,j} [p(i,j)]^2 \]
\[ F2 = \text{Con} = \sum_{i,j} [p(i,j) \times (i-j)^2] \]
\[ F3 = \text{Cor} = \sum_{i,j} (i^2 + j^2) \times p(i,j) - \mu_x \mu_y / \sigma_x \sigma_y \]
\[ F4 = \text{Idm} = \sum_{i,j} [p(i,j)]^2 [1 + (i-j)^2] \]
\[ F5 = \text{Ent} = -\sum_{i,j} p(i,j) \log[p(i,j)] \]
\[ F6 = \text{Var} = \sum_{i,j} (i - \mu)^2 p(i,j) \]

Where \( \mu_x, \mu_y \) and \( \sigma_x, \sigma_y \) are respectively the mean and variance of \( p(i, j), p(i), p(j) \)

\[ \mu_x = \sum_i x \sum_j p(i,j) \text{ and } \mu_y = \sum_j y \sum_i p(i,j) \]
\[ \sigma_x^2 = \sum_i (i - \mu_x)^2 \sum_j p(i,j) \text{ and } \sigma_y^2 = \sum_j (j - \mu_y)^2 \sum_i p(i,j) \]

Among the above features, Asm is the measurement for the evenness of image grayscale distribution and the thickness of images textures. Its value will be large when the image grayscale distributes evenly and the image shows more coarse granules, and vice versa. Con reflects the gray-level contrast between the adjacent pixel, and small Con suggests obvious image texture. Cor measures the grainy. Small and regular grainy obtains higher Con. Idm provides measures of regularity, and Chaos will get small Idm. Entropy shows the complexity of the image, and the complex image corresponds to larger entropy. Var measures the contrast of the whole image.

Each feature reflects different property of the image and might have different effects on classification. However, classification by significant features can obtain higher accuracy than that obtained with all features. Therefore, an efficient feature evaluation and feature selection method is needed.
Feature selection is the process of selecting features from the original ones which should be involved by optimizing an objective function as a ranking criterion [20]. For classification problems, the ideal objective function is the expected value of the error, which is the error rate computed on finite number of examples. SVM proposed by Vapnik and co-workers is a kind of supervised classifier [21] and has become the most popular machine learning classifier. In this paper, SVM is applied as the classifier and its error rate for independent test datasets as objective function.

2.3. Support vector machine

SVM, proposed by Vapnik and co-workers [21], is a promising learning machine based on statistical learning theory (SLT). Unlike traditional methods that minimize the empirical training error, SVM aims at minimizing an upper bound of the generalization error through controlling the margin between the separating hyper-plane and the data. This can be regarded as an approximate implementation of the structure risk minimization (SRM) principle. The most remarkable characteristics of SVM are the avoidance of local minima, the sparseness of solution, and the use of the kernel-induced feature spaces. Given a training data, the SVM training algorithm can obtain the optimal separating hyper-plane in terms of generalization error. We will briefly review the basic theory of SVM.

Given a set of labelled training points, \((x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\), where \(l\) is the number of training points. Each point \(x_i \in \mathbb{R}^n\) belongs to either of two classes and is given a label \(y_i \in \{-1, 1\}\) for \(i = 1, \ldots, l\). SVM transfers the original input space into a higher dimension feature space \(Z\) by mapping \(z = \phi(x)\) and searching the optimal hyper-plane in this feature space, which can be formulated as solving the following primal problem:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\
\text{subject to} & \quad y_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, \ldots, l \\
& \quad \xi_i \geq 0, \quad i = 1, \ldots, l
\end{align*}
\]

where the weight vector \(w\) and the bias \(b\) define the separating hyperplane.

And its dual problem is defined as follows:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\
\text{subject to} & \quad y_i \alpha = 0 \\
& \quad 0 \leq \alpha \leq C, \quad i = 1, \ldots, l
\end{align*}
\]

where \(Q\) is a \(l \times l\) positive semidefinite matrix, and its element \(Q_{ij} = y_i y_j K(x_i, x_j)\), and \(K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle\) is the kernel function. The kernel function we adopt is RBF:

\[
K(x_i, x_j) = \exp \left(-r \|x_i - x_j\|^2\right)
\]

and \(e\) is a vector of all ones, \(C\) is the penalty parameter of the error term, \(C > 0\) is the upper bound of all variables \(\alpha\).
2.4. Duplicative-feature support vector machine for feature selection and classification

Generally, the feature selection problem is defined as selecting a subset of whole feature set and no duplication should exist in the selected subset. It requires two essential factors: (1) criterion function for evaluating the quality of selected feature subset; (2) effective strategy for finding the optimal subset of features. An ideal criterion function for classification problem is the expected value of the error. In this paper, we keep the classification on independent test set to estimate the classification error rate.

As for search strategy, the classical feature selection methods can be split into two basic strategies: optimal methods and suboptimal methods. Exhaustive search methods are typical optimal methods which are feasible for only very small problems. And sequential forward selection (SFS) is a kind of typical suboptimal methods, which sequentially adds chosen feature without replacement to an empty candidate set until the addition of further features does not decrease the criterion. Both classical selection methods will not choose duplicate features. However, in our study we find out that even selected features have different effect on classification, while duplication can be adopted to represent those differences.

In our DFSVM, to find the best feature set, each feature is picked out with replacement and added into chosen feature set according to its importance in the criterion function. The iteration can be described as following:

Chosen features list \( ch = [1] \), candidate set \( r = \{ F_1, F_2, F_3, F_4, F_5, F_6 \} \), and the criterion function \( e = [1] \) which is the classification error rate on independent test set are initialized firstly. Then, the training samples with the corresponding target values

\[
x_{tr} = (x_{tr}, y_{tr}), x_{tr} = \{x_{1tr}, x_{2tr}, x_{3tr}, x_{4tr}, x_{5tr}, x_{6tr}\};
\]

\[
y_{1tr} = 1 \text{ for cirrhosis; } y_{0tr} = 0 \text{ for normal}
\]

and the test samples with the target values

\[
x_{te} = (x_{te}, y_{te}), x_{te} = \{x_{1te}, x_{2te}, x_{3te}, x_{4te}, x_{5te}, x_{6te}\};
\]

\[
y_{1te} = 1 \text{ for cirrhosis; } y_{0te} = 0 \text{ for normal}
\]

are input to DFSVM.

Step 1: Pick out each feature and the class label of the train dataset \( X_{tr} = \{x_{1tr}, y_{tr}\} \) to train SVM for classification model. Test each model on the independent test dataset \( X_{te} = \{x_{1te}, y_{te}\} \), and then get the classification error rate of the candidate set \( cr = \{r_1, r_2, r_3, r_4, r_5, r_6\} \). Rank the candidate features depending on \( cr \). Termination condition of iteration \( e_{min} = [1] \) is initialized.

Step 2: Initialize error list \( te = \{\} \). Pick out each candidate feature \( Fi \) from \( r \) according to the rank order, restrict training sample to feature set \( ch, Fi \). And train classification to get DFSVM \( ch, Fi \).

Step 3: Test the model on the independent test sample, and get the classification error rate \( e_i \) for each candidate feature. Update the error list \( te = \{te, e_i\} \).

Step 4: Find the feature with smallest error rate \( h = \arg \min (e_i) \).

Step 5: Update the chosen feature list \( ch = \{ch; h\} \) and the criterion function \( e = [e, e_i] \).

Step 6: Find the smallest error rate \( e_{k} \) of \( e \). Update \( e_{min} = [e_{min}, e_k] \).

Step 7: If \( e_{min} \) converges then iteration terminate; else go to Step 2.

Step 8: Get the optimal feature subset \( ch_{1..k} \) corresponding to the smallest error rate \( e_k \).

3. Experiments

To evaluate the effectiveness of DFSVM for features selection, five sets of experiments are performed. The performance of each feature is studied separately as the foundation of classification. Then DFSVM based on 6 texture features is carried out, with classification error rate and ROC curve are worked out for the selected features. For comparison purposes, feature selection without replacement is studied to illustrate the performance of duplication caused by replacement of DFSVM. Duplication strengthens the specific feature, which is similar with feature-weight. So feature-weight is also compared. As the ranges of 6 texture features vary largely, normalization is adopted to verify influence at last. And the typical classifier ANN is also carried out for compare. Above experiments are implemented in MATLAB and tested on an AMD Phenom(tm) 2.8 GHz machine with 2 GB of RAM. The generally used parameters with RBF kernel and \( C=1000 \) are chosen for SVM.

Commonly used parameters are adopted to evaluate the performance: \( N_c \) (error) is the proportion of incorrectly predicted samples as criterion function for evaluating the quality of selected feature subset, \( N_p \) (sensitivity) is the percent of true positive representing the correctly predicted cirrhosis, and \( N_m \) (specificity) is the percent of true negative for the correctly predicted normal.

3.1. Experiment 1: performance of each feature

Each extracted feature reflects different property of the image, then their differences among range and classification performance based on each feature are studied firstly. Fig. 2 shows the different range of each feature: Each column contains all values of a feature. And it is obvious that F2 has the largest range. Compared to F2, the ranges of other features just like a point. Generally, normalization is often applied to scale features to a specific range and avoid outweighing the larger-range features. However, in DFSVM, features can be strengthened with duplication so normalization is not adopted firstly. And the normalization is also studied in subsequent experiment.

Classification performance reflects the difference among extracted features in Table 1 and Fig. 3. The classifiers based on each feature and the whole features set are trained on training dataset and tested on independent test dataset to get classification error rate as shown in Table 1. To illustrate the sensitivity and specificity of classifier in detail, the receiver operating characteristic (ROC) is applied as a quantitative measure for each model in Fig. 3. ROC curve is defined as a plot of test sensitivity as they...
specificity
Classification
ROC
feature
Ne
18.18%
coordinate
or
got
with
and
the
features
that
are
and
and
After
this
Table 1
Classification performance of each feature (%).

<table>
<thead>
<tr>
<th>Feature</th>
<th>(N_s)</th>
<th>(N_p)</th>
<th>(N_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>45.07</td>
<td>51.28</td>
<td>59.38</td>
</tr>
<tr>
<td>F2</td>
<td>32.39</td>
<td>58.97</td>
<td>78.13</td>
</tr>
<tr>
<td>F3</td>
<td>50.70</td>
<td>48.72</td>
<td>50.69</td>
</tr>
<tr>
<td>F4</td>
<td>45.07</td>
<td>17.95</td>
<td>1</td>
</tr>
<tr>
<td>F5</td>
<td>36.62</td>
<td>48.72</td>
<td>81.25</td>
</tr>
<tr>
<td>F6</td>
<td>45.07</td>
<td>51.28</td>
<td>59.38</td>
</tr>
<tr>
<td>All</td>
<td>26.76</td>
<td>89.74</td>
<td>53.13</td>
</tr>
</tbody>
</table>

\(N_s\) is the proportion of incorrectly predicted samples, \(N_p\) is the sensitivity, and \(N_m\) is the specificity.

Fig. 3. The receiver operating characteristic (ROC) to illustrate the sensitivity and specificity of classifier based on each feature.

Fig. 4. the convergence curve of criterion function \(e\) and termination condition of iteration \(e_{\text{min}}\) of DFSVM.

coordinate versus its 1-specificity or false positive rate (FPR) as the \(x\) coordinate, which is an effective method of evaluating the quality or performance of diagnostic tests and is widely used to evaluate the performance of classifier in recent years. The more ROC curve hugs the left and top edges of the plot, the better the classification. And the area under the ROC curve can also measure the probability of correct classification.

From Table 1 and Fig. 3, it is clear that classifiers based on one feature are not ideal. Even the minimal error rate is above 30%. However, different characteristics have their own advantages. For example, although F4 has the lowest sensitivity, it has absolute specificity. Among the classifier trained by one feature, F2 hugs the left and top edges of the plot more than other features. F4 is another quality feature especially when FPR is just over 0.2. It is inspiring that just assembling those features can improve the classification performance to 26.76%. So the classification error rate, ROC curve and the improvement obtained by assembling all features illustrate that combining features might give a promising result, which indicates a good prospect to effectively combine the extracted features.

3.2. Experiment 2: performance of DFSVM

For feature selection procedure, training samples is randomly separated into two sets, 80% and 20% for training and validation respectively. DFSVM converges after 4 iterations, and the convergence curve is shown in Fig. 4. Table 2 shows the corresponding values. At the beginning of iterations, with the number of chosen features gradually increasing, the criterion function decreases to 18.18% and the minimal classification error rate drops to 18.18% simultaneously. When DFSVM converges, the chosen feature list includes F2, F4, F4, F4, which are the chosen features. After this point, the minimal classification error rate does not change any more. To illustrate the trend of criterion function and minimal classification error rate, the curve convergence is shown in Fig. 4 and the quantitative value is list in Table 2 with the number of iterations is manually set to 20.

3.2.1. DFSVM compared with feature selection of sampling without replacement

As is shown earlier, classification error rate can be decreased by adding the effective features even duplication, which is caused by sampling with replacement of search strategy. However, after it drop to a specific value, 18.18% in this problem, adding more features can not get better result, and even will lead to bad ones. When the classification error rate reaches the lowest point, the chosen feature list by DFSVM contains F2 and three F4 with duplication existing. The two duplicative features F4 is duplicative in the sense of set while they strengthen F4 and decrease the error rate from 27.27% to 18.18%. Therefore, the duplication brought by sampling with replacement becomes a good mechanism of DFSVM to weight the specific feature.

On the other hand, feature selection by sampling without replacement is brought into comparison and the convergence curve is shown in Fig. 5. When classification error rate hit the lowest point, 22.73%, three features F2, F4 and F6 are picked out and the value is 4.55% higher than that got by sampling with replacement. That is after F2 and F4 chosen, redundant F4 can bring better result than a fresh feature F1. So it is clear that sampling with replacement of DFSVM can get better classification performance by duplication.

Table 2
The corresponding data of convergence curve.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Choose feature</th>
<th>Criterion function: (e) (%)</th>
<th>Termination condition: (e_{\text{min}}) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F2</td>
<td>27.27</td>
<td>27.27</td>
</tr>
<tr>
<td>2</td>
<td>F4</td>
<td>27.27</td>
<td>27.27</td>
</tr>
<tr>
<td>3</td>
<td>F4</td>
<td>22.73</td>
<td>22.73</td>
</tr>
<tr>
<td>4</td>
<td>F4</td>
<td>18.18</td>
<td>18.18</td>
</tr>
<tr>
<td>5</td>
<td>F4</td>
<td>18.18</td>
<td>18.18</td>
</tr>
<tr>
<td>6</td>
<td>F2, F4</td>
<td>22.73</td>
<td>18.18</td>
</tr>
<tr>
<td>7</td>
<td>F1</td>
<td>22.73</td>
<td>18.18</td>
</tr>
<tr>
<td>8</td>
<td>F4</td>
<td>18.18</td>
<td>18.18</td>
</tr>
<tr>
<td>9</td>
<td>F2, F4</td>
<td>18.18</td>
<td>18.18</td>
</tr>
<tr>
<td>10</td>
<td>F4</td>
<td>18.18</td>
<td>18.18</td>
</tr>
</tbody>
</table>

Please cite this article in press as: L. Hui et al., Cirrhosis classification based on MRI with duplicative-feature support vector machine (DFSVM), Biomed. Signal Process. Control (2013), http://dx.doi.org/10.1016/j.bspc.2013.03.001
3.2.2. The effect of duplication

With the above selected features, the classification performance is assessed by independent testing samples in the following part. To study the effect of duplication, the chosen features F2 and F4 without duplication, named as CF, is picked out for comparison firstly. CF gets classification error rate 19.72\%, sensitivity 94.87\%, and specificity 62.50\% as shown in Table 3. Comparing to those of all features, just picking out F2 and F4 can improve both sensitivity and specificity. So the combination of F2 and F4 is a good foundation. However, specificity is not very good. Comparing to CF, DFSVM greatly improves the specificity to 88.50\% by adopting two repetitve redundancies to strengthen F4. Although it slightly reduces the sensitivity by 2.56\%, DFSVM is satisfying as low specificity is the major puzzle for image doctor.

It is worth noting that weighting is a more common method to outweigh a certain feature. So we also compare F2 and weighted F4 multiplied by three, named as WCF, and get classification error rate 16.9\%, sensitivity 94.87\%, and specificity 68.75\%.

Above results show that enhancing the influence of F4 by adding redundancies or weighting can improve specificity. Redundancies greatly improve specificity by 25\%, which is consistent with that F4 has absolute specificity. And it is clear that duplication is more effective for specificity than weighting as it only improve 6.25\%. However, F4 has the lowest sensitivity. So it might be F2 that guarantees the overall sensitivity. The same sensitivity of WCF and CF demonstrates weighing have no effect on F2. Although duplication of DFSVM greatly improves the specificity, it drops the proportion of F2 and then reduces the sensitivity by 2.56\%. So the following conclusion can be got: duplication greatly enhances the influence of specific features and might slightly affect the other features. And duplication cannot be simply reduced to an integral-weighted feature.

Fig. 5 is ROC curve for classification based on CF, WCF and DFSVM in detail. The line for all features is the result for assembling all features and put here as a reference. CF, WCF and DFSVM are all better than just assembling all features, which proves that classification by significant features can obtain higher accuracy. ROC curve also shows that DFSVM hugs the left and top edges of the plot and WCF is slightly better than CF, which is consistent with above classification error rate.

3.3. Experiment 3: comparison with feature reduction methods

In order to assess the performance of the proposed method, an experimental comparison has been done with respect to other typical feature reduction approaches including PCA, LDA, all min–max normalized features, weighted feature based on information gain (Weight-InformGrain) as shown in Table 4 and Fig. 7. PCA and LDA incline to improve specificity, however reduce sensitivity too much. Classification with all normalized features keeps the sensitivity and specificity of classifier.

### Table 3

<table>
<thead>
<tr>
<th>Feature</th>
<th>$\mathcal{N}_f$</th>
<th>$\mathcal{N}_p$</th>
<th>$\mathcal{N}_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>26.76</td>
<td>89.74</td>
<td>53.13</td>
</tr>
<tr>
<td>CF</td>
<td>19.72</td>
<td>94.87</td>
<td>62.50</td>
</tr>
<tr>
<td>WCF</td>
<td>16.90</td>
<td>94.87</td>
<td>68.75</td>
</tr>
<tr>
<td>DFSVM</td>
<td>9.86</td>
<td>92.31</td>
<td>87.50</td>
</tr>
</tbody>
</table>

Please cite this article in press as: L. Hui, et al., Cirrhosis classification based on MRI with duplicative-feature support vector machine (DFSVM), Biomed. Signal Process. Control (2013), http://dx.doi.org/10.1016/j.bspc.2013.03.001
slightly improves the specificity comparing with those of original features. That implies different ranges of original features might have impact on classification. It should be kept in mind that normalization hasn’t been adopted before selection specific features in our proposed DFSVM. Then normalization will be studied for comparing in the following section. Feature weight strengthens the specific feature by measuring the discriminating ability of each feature in terms of category separation such as information gain [22]. So weighted feature based on information grain is listed. However, classification rate does not improve based on Weight-InformGrain. The reason might be that probability density is not well estimated with the small sample size when computing the information gain for Weight-InformGrain.

3.4. Experiment 4: effect of normalized

To avoid outweighing the lager-range features, linear min–max normalization transformation is also studied on the original features. That is, normalization is adopted before selection. Comparison is shown in Fig. 8 and Table 5. It is easy to find that the normalization does not largely improve the classification performance. As for feature selection based on normalized features, the minimal error rate is 11.27% which is higher than that of DFSVM. With above result, normalization is not adopted in our algorithm.

3.5. Experiment 5: compare with other typical classification methods

In order to assess the classification performance of the proposed method, the ANN is also applied. The result is that $N_c$ equals 26.76%, $N_p$ equals 84.62% and $N_m$ equals 59.38%, which indicates that the ANN obtains lower classification rate and low sensitivity and specificity than DFSVM.

4. Conclusions

In this paper, DFSVM algorithm is proposed for classification and adopted for detection cirrhosis from normal hepatic tissue MR imaging. Six GLDM based texture features are extracted from medical MRI. Experimental results show that DFSVM can select important features and strengthen the specific feature by duplication caused by sampling with replacement in iteration. Our proposed DFSVM is compared with typical feature reduction approaches such as PCA, LDA and Weight-InformGrain, and also compared with typical classifier ANN. The experiment result shows that DFSVM gets both high sensitivity and high specificity. Duplication is also studied that it cannot be simply reduced to an integral-weighted feature.

As the candidate feature set is small, less iteration can get the satisfying result. However, with more candidate features, it is necessary to accelerate the training process. In Step 2 of DFSVM, each candidate feature is picked out according to the rank order. So, our future work will involve abbreviating iteration based on above rank order.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61003175 and No. 81071127).

References


