

文章编号 1004-924X(2013)08-2137-09

## 基于两阶段集成支持向量机的前列腺肿瘤识别

周 涛<sup>1</sup>, 陆惠玲<sup>1\*</sup>, 陈志强<sup>2</sup>, 马 苗<sup>3</sup>

(1. 宁夏医科大学 理学院, 宁夏 银川 750004; 2. 宁夏医科大学附属总医院 放射科, 宁夏 银川 750004;  
3. 陕西师范大学 计算机科学学院, 陕西 西安 710062)

**摘要:**从核磁共振成像(MRI)对前列腺肿瘤的诊断入手,提出了一种基于两阶段集成支持向量机(SVM)的前列腺肿瘤辅助诊断方法。首先,提取 MRI 图像中的前列腺感兴趣区域(ROI)的统计特征、纹理特征和不变矩特征;然后,在不同的特征空间里,使用不同的核函数来扰动 SVM 参数并在不同的特征空间生成个体 SVM,通过相对多数投票进行第一次集成;接着把第一次集成结果用相对多数投票进行第二次集成;最后,以前列腺患者的 MRI 图像为原始数据,采用两阶段融合集成 SVM 对前列腺肿瘤进行辅助诊断。实验显示,第一次集成分类准确率最高比单 SVM 提高了 26.67%,第二次集成分类准确率比第一次集成 SVM 提高了 3.33%,结果表明本文算法能够有效提高前列腺肿瘤的认识精度。

**关键词:**前列腺肿瘤;核磁共振成像;辅助诊断;集成支持向量机

中图分类号:R445.2;TP391.4 文献标识码:A doi:10.3788/OPE.20132108.2137

## Prostate tumor recognition based on two-stage integrating SVM

ZHOU Tao<sup>1</sup>, LU Hui-ling<sup>1\*</sup>, CHEN Zhi-qiang<sup>2</sup>, MA Miao<sup>3</sup>

(1. School of Science, Ningxia Medical University, Yinchuan 750004, China;

2. Department of Radiology, The Affiliated General Hospital of Ningxia Medical University, Yinchuan 750004, China;

3. College of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

\* Corresponding author, E-mail: lu-huilin@163.com

**Abstract:** On the basis of prostate tumor diagnosis by nuclear Magnetic Resonance Imaging(MRI), a two-stage ensemble Support Vector Machine(SVM) method were proposed to realize the prostate tumor aided diagnosis. Firstly, the statistical features, invariant moment features and the texture feature of the Area of Interest(ROI) for the prostate in a MRI image were extracted. Then, SVM parameters were disturbed by using different kernel functions in different feature spaces, and the first ensemble was carried out by relative majority voting. Furthermore, the results of first ensemble were integrated again by the relative majority voting. Finally, MRI images of prostate patients were regarded as original data, and two-stage ensemble SVM were utilized to aid tumor diagnosis. Experiment results show that the classification accuracy from the first ensemble has improved by 26.67% as com-

收稿日期:2012-12-27;修订日期:2013-02-16.

基金项目:国家自然科学基金资助项目(No. 81160183);教育部“春晖计划”资助项目(No. Z2011051);宁夏自然科学基金资助项目(No. NZ12179);宁夏高等学校科研重点项目(No. NXY2011042);陕西省青年科技新星资助项目(No. 2011kjxx17);宁夏医科大学特殊人才计划资助项目(No. XT2011004);宁夏医科大学青年基金资助项目(No. XQ2011011)

pared with that of single-stage SVM and that from the second ensemble has improved 3.33% than that of the first ensemble. These results illustrate that the proposed algorithm can improve the recognition accuracy of prostate tumor effectively.

**Key words:** prostate tumor; Nuclear Magnetic Resonance Imaging(MRI); aided diagnosis; integrating SVM

## 1 引言

前列腺肿瘤主要包括前列腺癌与前列腺增生,其中前列腺癌是一种常见的恶性肿瘤<sup>[1-2]</sup>,多发于老年人且复发率极高,并且有逐步年轻化的趋势。一般情况下,前列腺增生是不会转变为前列腺癌的,首先前列腺增生主要发生在前列腺中央区域的移行带,而前列腺癌则主要发生在前列腺的外周带,两者在解剖结构上有很大的差别;其次前列腺增生与前列腺癌病理进程不同,到目前为止,只有雄激素能促使病理性前列腺癌向临床前列腺癌转变的证据,并无促使良性前列腺增生向前列腺癌转化的证据。然而,前列腺增生和前列腺癌是可以同时存在的,临床统计表明有少部分前列腺增生(10%)会演变为前列腺癌,因此前列腺癌与前列腺增生的诊断在临床中具有重要的意义<sup>[3]</sup>。

核磁共振成像(MRI)具有很好的软组织分辨率和三维成像的特点,并可通过选择不同的序列和扫描参数成像,能较好地显示前列腺的分区解剖及相邻的组织结构,因此可用于前列腺癌的定位,以及癌组织范围、有无穿透包膜、有无远处转移的检测。目前,MRI检查已成为前列腺肿瘤诊断和分析的常用方法。在前列腺肿瘤识别方面,文献[4]根据 t 检验,互信息,最小冗余和最大相关性准则,提取 MRI 图像中的前列腺癌的统计特征、Haralick 特征和梯度特征等,采用 4 种分类器非线性支持向量机(SVM),线性判别分析,k 近邻分类和朴素 bayes 分类)对其进行识别,结果表明这种计算机辅助诊断系统对外周带前列腺癌的辅助诊断具有潜力;文献[5]通过分析 303 例患者的 4 944 例肠超声图像,基于 k 近邻和隐马尔科夫模型提出了一种前列腺癌计算机辅助诊断方法,结果表明泌尿科专家在计算机系统辅助下的诊断能力比没有辅助系统的情况明显提高;文献[6]主要研究了多特征前列腺诊断和前列腺在组

织学图像中的 Gleason 分级。这些文献对前列腺肿瘤的辅助诊断的方法都是个体分类器的应用,而集成分类技术未见文献报道,本文以前列腺肿瘤为研究对象,给出了一种有效的两阶段集成辅助诊断方法。

本文首先介绍了基础理论,主要包括 MRI 图像中前列腺肿瘤的 3 类特征和集成 SVM;接着阐述了本文提出的基于两阶段集成 SVM 的前列腺肿瘤识别算法;最后,进行了算法仿真,给出并比较了单支持向量机、一次集成 SVM、二次集成 SVM 的实验结果。实验结果表明,该方法对于前列腺肿瘤的诊断有应用潜力。

## 2 基础理论

### 2.1 MRI 图像中前列腺肿瘤的特征

#### 2.1.1 统计特征

统计特征是对一定范围内灰度的大小,出现的频率和规律进行统计后得到的数值,本文选取以下 5 个统计特征:

(1)均值:

$$\mu = \frac{1}{N} \sum_{n=1}^N x_i.$$

其中,  $N$  为目标像素总个数,  $x_i$  为像素灰度值;

(2)矩形度:

$$R = \text{Area} / (H \times W).$$

其中, Area 为目标区域面积,  $H, W$  为长和宽;

(3)周长:

$$C = \sum_{i=1}^M \sum_{j=1}^N P(i, j).$$

其中,  $P(i, j)$  为目标边缘像素,  $M, N$  为长和宽;

(4)粗糙度:

$$d(i) = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2},$$

$$R = \frac{1}{N} \sum_{n=1}^N |d(i) - d(i+1)|.$$

$d(i)$  为  $(x_i, y_i)$  到中心  $(x_0, y_0)$  的距离;

(5)伸长度:  $E = \min(H, W) / \max(H, W)$ 。

2.1.2 纹理特征

由于纹理是灰度分布在空间位置上反复出现而形成的,因而在图像空间中相隔某距离的两像素之间会存在一定的灰度关系,即图像中灰度的空间相关特性。灰度共生矩阵<sup>[7]</sup>就是一种通过研究灰度空间相关特性来描述纹理的常用方法。灰度共生矩阵的产生如下:取(N×N)矩阵的任一点(x<sub>i</sub>, y<sub>i</sub>),其像素值为 P<sub>1</sub>,记与它距离为 d,方向为 θ 的点(x<sub>0</sub>, y<sub>0</sub>),像素值为 P<sub>2</sub>,用(P<sub>1</sub>, P<sub>2</sub>, d, θ)表示这样的像素对出现的频率。归一化为出现的概率 P(P<sub>1</sub>, P<sub>2</sub>, d, θ),这样的方阵称为灰度共生矩阵。在本文中 θ 的值取经典的 0、45、90、135°, d 的值取 1。基于灰度共生矩阵的特征有很多,本文选取以下几种<sup>[8]</sup>:

(1)角二阶矩(能量):

$$t_1 = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} c_{ij}^2$$

其中, c<sub>ij</sub> 表示(P<sub>1</sub>, P<sub>2</sub>, d, θ)的像素对出现的频率, K 是图像灰度级。 t<sub>1</sub> 反映图像的均匀性。该测度对区域内部有无灰度变化较敏感,但对灰度变化数值大小不敏感,即具有高的局部灰度对比度的区域角二阶矩值不一定高。

(2)惯量(反差、对比度):

$$t_2 = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} (i-j)^2 c_{ij}$$

惯量又称反差,作为图像对比度的度量。该量对非均匀的局部对比度敏感。非均匀区的惯量值大,均匀区的惯量值小。

(3)熵:

$$t_3 = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} c_{ij} \log c_{ij}$$

反映区域内的随机程度。当所有 c<sub>ij</sub> 都相等时,熵有最大值。该测度可用于区分有结构纹理的组织与缺少纹理的组织,前者熵值要小些。

(4)相关度:

$$t_4 = \frac{\sum_{i=0}^{K-1} \sum_{j=0}^{K-1} c_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$$

其中, μ<sub>x</sub>、μ<sub>y</sub>、σ<sub>x</sub>、σ<sub>y</sub> 是共生矩阵第 i 行 c<sub>x</sub>(i),第 j 列 c<sub>y</sub>(j)的均值与方差, K 是图像灰度级。

2.1.3 不变矩特征

M. K. Hu 最早提出了几何矩<sup>[9]</sup>的概念,他利用连续状态下的二阶和三阶中心矩构造了 7 阶

不变矩,此处仿照 M. K. Hu 构造的 7 阶不变矩利用离散状态下的二阶和三阶中心矩来构造新的不变矩。由于这 7 个不变矩的变化范围很大,不便于比较,故利用取对数的方法进行了数据压缩。因此,实际采用的不变矩为 C<sub>k</sub>':

$$C_k = \lfloor \lg_{10} |C_k'| \rfloor \quad k=1, 2, \dots, 7$$

经修正后的不变矩特征同样具有平移、旋转和比例不变性。一般将修改后的 C<sub>1</sub>、C<sub>2</sub>、C<sub>3</sub>、C<sub>4</sub>、C<sub>5</sub>、C<sub>6</sub>、C<sub>7</sub> 作为目标的不变性特征。

2.2 集成 SVM 理论

支持向量机<sup>[10-12]</sup>是统计学习理论中一种十分重要的方法,对统计学习理论的发展起到了巨大的推动作用。其中心思想是利用核技巧将样本变换到高维特征空间,根据结构风险最小化原则和对偶原理,通过求解一个凸二次规划问题,找出支持向量及其系数构造最优分类面,使得该分类面不仅能够将两类样本分开,而且分类间隔最大。然而,和神经网络等其它机器学习方法类似,由于训练样本相对有限,所以 SVM 在训练集上的分类性能很好,而在测试集上相对较差。集成(ensemble)是解决上述问题的一个有效方法,它首先建立一组比较精确而又相对独立的分类器对测试样本进行分类,然后采用一定的策略或准则将各个学习机的输出结果进行融合得到最终的判别结果。近几年来,集成学习在机器学习、神经网络、统计学等领域都成为一个相当活跃的研究热点,并被认为是当前机器学习领域的 4 大研究方向之首<sup>[13]</sup>。文献[14]将集成方法应用于 SVM,并取得了较好的效果;文献[15]将集成 SVM 成功地应用于癌症识别,实验效果良好。

集成方法的关键是在给定训练集上构造多个性能较好而又相对独立的分类器,文献[16]给出了集成的个体分类器的生成方法有以下几种:

(1)扰动训练集中的样本。Bagging 和 Boosting 是此种方法的典型。这两种算法均对训练数据集进行重采样得到许多新的训练集,在这些新获得的训练集上构造相应的分类器,一般来说采用投票法就能获得最终分类结果;

(2)扰动样本的输入特征。通过随机产生、特征选择或特征提取算法可获得不同的特征向量子空间,在这些不同特征子空间上训练各成员分类器就能得到相应的成员分类器;

(3)扰动分类器的模型参数。一般分类器都

有自己特定的模型参数,扰动这些参数就可以得到同构条件下的不同成员分类器,例如  $k$  近邻中不同的  $k$  值和距离度量,神经网络不同的节点数目,SVM 的不同核函数和不同的核参数等;

(4) 扰动分类器的输出结果。即对分类器的输出结果进行处理;

(5) 组合不同的分类器。上述 4 种方法都是来自同一类型的分类器,称之为同构集成,不同类型的分类器组合,如神经网络,支持向量机和决策树等称之为异构集成<sup>[10]</sup>。

本文采用扰动分类器的模型参数方法和扰动样本的输入特征来构造个体分类器,最后通过两阶段相对多数投票方法<sup>[15]</sup>(其思想就是个体分类器  $C_i$  对样本  $x$  进行识别投票,得票数多的那个分类  $C_i$  作为对应样本  $x$  的最终分类)得到集成支持向量机。

### 3 基于两阶段集成 SVM 的前列腺肿瘤识别算法

#### 3.1 算法思想

本文以 MRI 中的前列腺肿瘤为研究对象,采集 13 例前列腺患者(其中前列腺增生 5 例、前列腺癌 8 例)的 MRI 图像,提取前列腺 ROI 区域的统计特征(均值,矩形度,周长,粗糙度,伸长度)、纹理特征(基于灰度共生矩阵的熵,相关,能量,惯性矩的 4 个方向 16 个特征)和几何特征(Hu 七阶不变矩),共计 29 个特征,然后在 3 类特征分别构成的特征空间里,采用不同的核函数扰动模型和扰动样本的输入特征来构造个体 SVM,在此基础上通过相对多数投票融合结果进行第一次集成;再次把在三类特征构成的特征空间里的第一次集成结果通过相对多数投票进行第二次集成;利用二阶段融合策略来构造集成 SVM 对前列腺肿瘤进行分类识别。算法思想见图 1。

具体步骤如下:

步骤 1: 从 13 个病例中提取 100 幅前列腺 MRI 图像中的 ROI 区域;

步骤 2: 对前列腺 MRI 图像中的 ROI 区域进行预处理;

步骤 3: 提取统计特征(均值、周长、粗糙度、矩形度、伸长度)、Hu 不变矩、纹理特征(角二阶

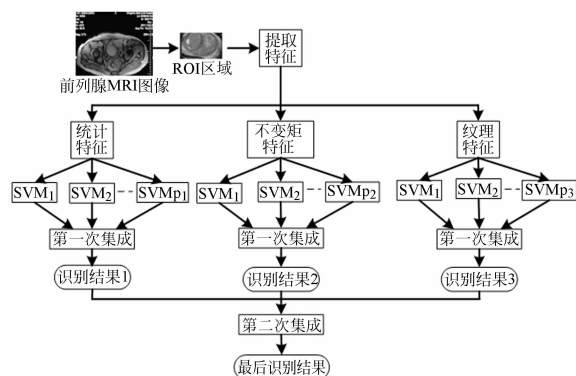


图 1 两阶段集成 SVM 算法流程图

Fig. 1 Flow chart of two-stage ensemble SVM algorithm

矩、惯性、熵、相关度);

步骤 4: 扰动样本的输入特征,即在 3 个不同的特征空间(统计特征空间、Hu 不变矩特征空间、纹理特征空间)中构造训练样本子集;

步骤 5: 扰动 SVM 模型参数,即在同一个特征子集中,通过不同的核函数实现扰动分类器的模型参数构造不同的个体 SVM 分类器;

步骤 6: 对 3 个样本子集分别进行各自训练;

步骤 7: 对个体 SVM 输出结果进行相对多数投票。

#### 3.2 算法描述

算法名称: 基于两阶段集成 SVM 的前列腺肿瘤识别算法

输入:

1) 训练样本集  $X = \{x_1, \dots, x_m, x_{m+1}, \dots, x_M\}$ ,  $m$  为前列腺癌训练样本数,  $M-m$  为前列腺增生训练样本数;

2) 样本类别数  $n=2$ ;

输出:

训练好的两阶段集成 SVM--Two\_Stage ES-VM;

步骤:

Begin

//分别提取  $M$  个样本图像的不变阶矩特征,纹理特征和统计特征

For  $i=1$  to  $M$

$T_1(i) = \text{Moment}(X(i));$

$T_2(i) = \text{Co\_Occurrence}(X(i));$

$T_3(i) = \text{Geometry}(X(i));$

End for

//在七阶矩特征空间  $T_1$  里利用核函数差异构

造个体分类器并第一次集成

```
Mom_Y1 = Moment_SVM (poly, T1); //用
poly 核函数
```

```
Mom_Y2 = Moment_SVM (sigmoid, T1); //用
sigmoid 核函数
```

```
Mom_Y3 = Moment_SVM (erfb, T1); //用
erfb 核函数
```

```
Mom_Y = Mom_Y1 + Mom_Y2 + Mom_Y3 //在
T1 空间里第一次相对多数投票
```

//在纹理特征空间  $T_2$  里利用核函数差异构造个体分类器并第一次集成

```
Co_Occ_Y1 = Co_Occurrence_SVM(poly, T2);
//用 poly 核函数
```

```
Co_Occ_Y2 = Co_Occurrence_SVM (rbf,
T2); //用 rbf 核函数
```

```
Co_Occ_Y3 = Co_Occurrence_SVM(spline,
T2); //用 spline 核函数
```

```
Co_Occ_Y = Co_Occ_Y1 + Co_Occ_Y2 + Co_
Occ_Y3; //在  $T_2$  空间里第一次相对多数投票
```

//在统计特征空间  $T_3$  里利用核函数差异构造个体分类器并第一次集成

```
Geo_Y1 = Geometry_SVM (poly, T3); //用
poly 核函数
```

```
Geo_Y2 = Geometry_SVM (erfb, T3); //用
erfb 核函数
```

```
Geo_Y3 = Geometry_SVM (spline, T3); //用
spline 核函数
```

```
Geo_Y = Geo_Y1 + Geo_Y2 + Geo_Y3; //在  $T_3$ 
空间里第一次相对多数投票
```

```
Y = Mom_Y + Co_Occ_Y + Geo_Y; //第二次
相对多数投票
```

End

## 4 算法仿真

### 4.1 实验环境与数据

本文实验所涉及到的软、硬件环境如下:

软件环境: Windows 7 操作系统, MATLAB 7.0, efilm 3.4;

硬件环境: 2 G 内存, 320 G 硬盘, T4000 酷睿处理器。

实验数据: 本文实验数据来源于宁夏医科大学放射科 13 例前列腺患者(其中增生 5 例、前列

腺癌 8 例)的 MRI 图像, 其中训练样本和测试样本规模和组成分别如下:

(1)训练样本生成: 取 40 幅前列腺癌、30 幅前列腺增生的特征;

(2)测试样本生成: 取 15 幅前列腺癌、15 幅前列腺增生的特征;

通过分析, 结合放射科医生医嘱, 提取出了具有区分价值的 100 幅前列腺 MRI 图像中的 ROI 区域, 出于篇幅的考虑, 训练样本中每种病灶给出 9 张图像, 测试样本就不单独给出。图 2 给出了 9 幅前列腺癌 MRI 图像 ROI 区域, 图 3 给出了 9 幅前列腺增生 MRI 图像 ROI 区域。具体如下:

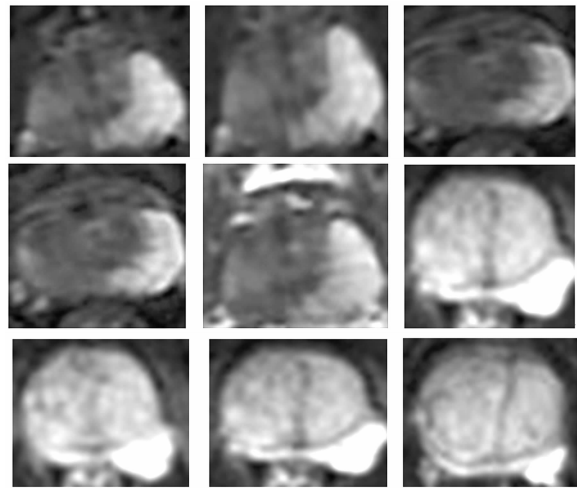


图 2 部分前列腺癌 MRI 图像 ROI 区域

Fig. 2 MRI images and ROI region of some prostate cancer

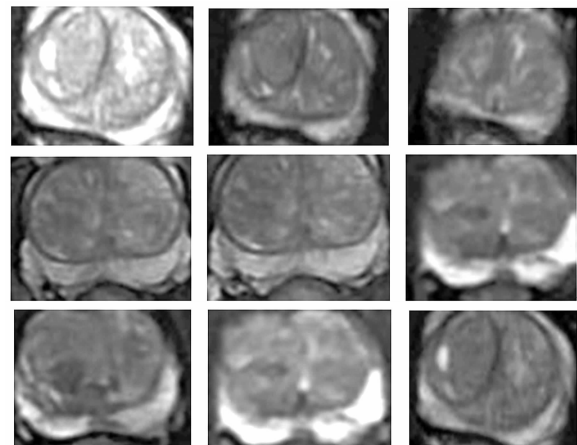


图 3 部分前列腺增生 MRI 图像 ROI 区域

Fig. 3 MRI images and ROI region of some prostate hyperplasia

4.2 前列腺肿瘤 MRI 图像特征提取举例

这里给出两例患者的前列腺 ROI 区域(图 4 是前列腺癌 ROI 区域,图 5 是前列腺增生 ROI 区域),并分别提取它们的 29 个特征值,表 1 给出了提取到的特征值。

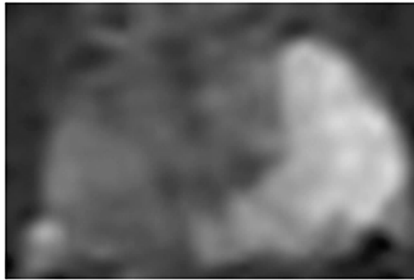


图 4 前列腺癌 ROI 区域  
Fig. 4 ROI region of prostate cancer

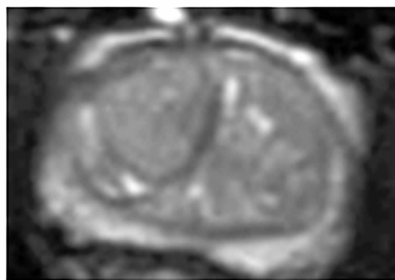


图 5 前列腺增生 ROI 区域  
Fig. 5 ROI region of prostate hyperplasia

表 1 前列腺 MRI 图像中 ROI 区域的特征值

Tab. 1 Eigenvalue of ROI region in prostate MRI image

特征	癌症特征值	增生特征值	特征	癌症特征值	增生特征值
能	0.027	0.023	均值	95.606	93.328
量	0.021	0.016	粗糙度	10.765	11.905
90°	0.027	0.021	对比度	34.621	41.151
135°	0.020	0.017	周长	7.327	12.648
熵	3.892	4.058	征 矩形度	1.468	1.3747
0°	4.144	4.464	伸长度	0.821	0.92
45°	3.938	4.239	H <sub>1</sub>	0.788	0.156
90°	4.187	4.384	不 H <sub>2</sub>	2.184	4.337
135°	1.363	1.581	变 H <sub>3</sub>	1.506	3.331
惯	2.409	4.029	矩 H <sub>4</sub>	3.411	5.847
性	1.600	2.517	特 H <sub>5</sub>	8.342	11.694
矩	2.727	3.264	征 H <sub>6</sub>	4.562	8.307
相	0.123	0.090	H <sub>7</sub>	6.763	10.958
45°	0.114	0.079			
关	0.120	0.086			
135°	0.111	0.082			

4.3 基于 SVM 的前列腺肿瘤的识别仿真实验

在单支持向量机中分别采用多项式核函数 poly、径向基核函数 rbf、样条核函数 spline、B 样条核函数 bspline 和 erfb 等 5 个核函数,在 70 个训练样本中进行训练,在 30 个测试样本上进行分类识别,分类结果如表 2 所示。结果表明,在前列腺癌的分类识别中,分类准确率最高的 3 个核函数是 poly、spline 和 bspline,都达到了 86.67%,在前列腺增生的分类识别中,分类准确率最高的是 rbf,达到 86.67%,而 erfb 核函数的分类准确率最低。可见,使用不同的核函数,分类准确率存在较大差异,为了提高精度,在后面集成的时候总是选择在单支持向量机中分类准确率最高的 3 个核函数作为个体 SVM 的核函数。

表 2 基于单 SVM 的前列腺肿瘤分类准确率

Tab. 2 Classification accuracy for prostate tumor based on single SVM (%)

准确率	核函数				
	poly	rbf	spline	bspline	erfb
前列腺癌	86.67	80.00	86.67	86.67	73.33
前列腺增生	80.00	86.67	73.33	73.33	66.67
整体准确率	83.34	83.34	80	80	70

4.4 基于两阶段集成 SVM 的前列腺肿瘤识别仿真实验

在不变矩特征构成的 7 维特征空间 T<sub>1</sub> 中,经过实验筛选,遴选出在单支持向量机中分类精度最高的前 3 个核函数 poly、sigmoid 和 erfb 作为个体分类器的核函数,分别构造 3 个分类器 Moment\_SVM (poly, T<sub>1</sub>)、Moment\_SVM (erfb, T<sub>1</sub>) 和 Moment\_SVM (sigmoid, T<sub>1</sub>) 对前列腺癌和增生的图像进行分类,其分类准确率如表 3 所示:

表 3 Moment\_SVM(T<sub>1</sub>) 分类准确率

Tab. 3 Classification accuracy for Moment\_SVM (T<sub>1</sub>) (%)

准确率	核函数		
	poly	sigmoid	erfb
前列腺癌	53.33	46.67	66.67
前列腺增生	100	53.33	86.67
整体准确率	76.67	50	76.67
平均准确率		67.78	

在灰度共生矩阵 16 维特征空间  $T_2$  中,经过实验筛选,遴选出在单支持向量机中分类精度最高的前 3 个核函数 poly、rbf 和 spline 作为个体分类器的核函数,分别构造 3 个分类器 Co\_Occurrence\_SVM (poly,  $T_2$ )、Co\_Occurrence\_SVM (rbf,  $T_2$ )和 Co\_Occurrence\_SVM (spline,  $T_2$ )对前列腺癌和增生的图像进行分类,其分类准确率如表 4 所示:

表 4 Co\_Occurrence\_SVM( $T_2$ ) 分类准确率

Tab. 4 Classification accuracy for Co\_Occurrence\_SVM( $T_2$ ) (%)

准确率	核函数		
	poly	rbf	spline
前列腺癌	73.33	86.67	73.33
前列腺增生	86.67	80.00	86.67
整体准确率	80	83.33	80
平均准确率	81.11		

在纹理特征 6 维特征空间  $T_3$  中,经过实验筛选,遴选出在单支持向量机中分类精度最高的前 3 个核函数 poly、erfb 和 spline 作为个体分类器的核函数,分别构造 3 个分类器 Geometry\_SVM (poly,  $T_3$ )、Geometry\_SVM (erfb,  $T_3$ )和 Geometry\_SVM (spline,  $T_3$ )对前列腺癌和增生的图像进行分类,其分类准确率如表 5 所示:

表 5 Geometry\_SVM( $T_3$ ) 分类准确率

Tab. 5 Classification accuracy for Geometry\_SVM ( $T_3$ ) (%)

准确率	核函数		
	poly	spline	erfb
前列腺癌	46.67	46.67	66.67
前列腺增生	73.33	73.33	53.33
整体准确率	60	60	60
平均准确率	60		

表 6 第一次集成 SVM 的肿瘤分类准确率

Tab. 6 Classification accuracy in 1st ensemble SVM (%)

分类器	Moment-SVM	Co-Occurrence-SVM	Geometry-SVM
准确率	83.33	90	86.67
平均准确率	86.67		

在 3 个不同的特征空间里(不变矩特征、灰度共生矩阵特征和统计特征),通过相对多数投票进行第一次融合,分别构造 Moment\_SVM、Co\_Occurrence\_SVM 和 Geometry\_SVM 3 个集成支持向量机,其对前列腺肿瘤识别结果如表 6 所示:

从 Moment\_SVM 可以看到,与单支持向量机的平均准确率相比,集成以后分类准确率提高了 15.55%,从 Co\_Occurrence\_SVM 可以看到,集成以后分类准确率提高了 8.89%,从 Geometry\_SVM 可以看到,集成以后分类准确率提高了 26.67%。

在 3 个不同的特征空间里,通过第一次集成得到了 Moment\_SVM、Co\_Occurrence\_SVM 和 Geometry\_SVM 3 个集成支持向量机,再通过第 2 次相对多数投票,对这 3 个集成支持向量机的结果进行再次融合,得到两阶段集成支持向量机 Two-Stage SVM,其准确率达到 90%,与第 1 次集成的平均准确率相比,准确率提高了 3.33%。

## 5 结 论

本文从 MRI 图像在前列腺疾病肿瘤的诊断入手,分析了前列腺肿瘤的辅助诊断现状,提出了一种基于两阶段融合集成 SVM 的方法来实现前列腺癌的辅助诊断。实验结果表明,第一次集成分类准确率最高比单支持向量机提高了 26.67%,二次集成 Two-SVM 的分类准确率与第一次集成 SVM 相比提高了 3.33%,可见该方法对前列腺肿瘤的诊断具有明显的应用价值。由于客观条件所限,本文样本数据的容量有限,提取的特征有待进一步优化,个体分类器是通过扰动模型参数,即采用不同的核函数来实现,在后续的工作中,应搜集更多的临床前列腺肿瘤病例,来丰富训练样本数量和质量;尽可能地提取更多的 ROI 区域的特征,并分析不同特征之间的相关性、不同特征组合对肿瘤识别精度的影响程度等;个体分类器的生成有 5 种不同的类型,不同的策略构造的分类器对最后集成的结果影响不相同,如何影响,影响到什么程度值得进一步的研究。从整体上讲,前列腺肿瘤的辅助诊断鲜有文献报道,采用集成的方法进行辅助诊断文献就更少,这是一个具有广阔应用前景的研究领域。

## 参考文献:

- [1] 何元烈. 多模医学图像配准与融合技术及医学智能辅助诊断系统研究[D]. 广州:华南理工大学, 2006.  
HE Y L. *Multimodal medical image registration, fusion and intelligent diagnosis system research* [D]. GuangZhou: South China University of Technology, 2006. (in Chinese)
- [2] 施发表, 韦嘉瑚, 王占立, 等. 正常和良性增生前列腺的高场强磁共振成像特征[J]. 中华老年医学杂志, 1997, 16(2): 79-83.  
SHI F B, WEI J H, WANG Z L, *et al.* Normal prostate and benign prostatic hyperplasia; MR imaging at 1.5T[J]. *Chinese Journal of Geriatrics*, 1997, 16(2): 79-83. (in Chinese)
- [3] 华立新, 居小兵, 吴宏飞, 等. 良性前列腺增生和前列腺癌组织中雄激素受体的表达[J]. 中华老年医学杂志, 2003, 22(7): 405-408.  
HUA L X, JU X B, WU H F, *et al.*. The expression of androgen receptor gene in benign prostatic hyperplasia and prostatic cancer tissues [J]. *Chinese Journal of Geriatrics*, 2003, 22(7): 405-408. (in Chinese)
- [4] NIAF E, ROUVIÈRE O, MÈGE-LECHEVALLIER F, *et al.*. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI [J]. *Physics in Medicine and Biology*, 2012, 57(12): 3833-3851.
- [5] LIOMET R, TOSELLI A H, JUAN C, *et al.*. Computer-aided prostate cancer detection in ultrasonographic images [C]. *Pattern Recognition and Image Analysis*, 2003, 2652: 411-419.
- [6] TABESH A, TEVEROVSKIY M, PANG H Y, *et al.*. Multifeature prostate cancer diagnosis and gleason grading of Histological [J]. *IEEE Transactions on Medical Imaging*, 2007, 26(10): 1366-1378.
- [7] 钟桦, 杨晓鸣, 焦李成. 基于多分辨共生矩阵的纹理图像分类[J]. 计算机研究与发展, 2011, 48(11): 1991-1999.  
ZHONG H, YANG X M, JIAO L CH. Texture classification based on multiresolution co-occurrence matrix [J]. *Journal of Computer Research and Development*, 2011, 48(11): 1991-1999. (in Chinese)
- [8] 薄华, 马缚龙, 焦李成. 图像纹理的灰度共生矩阵计算问题的分析[J]. 电子学报, 2006, 34(1): 155-159.  
BO H, MA F L, JIAO L C. Research on computation of GLCM of image texture [J]. *Acta Electronica Sinica*, 2006, 34(1): 155-159. (in Chinese)
- [9] HU M K. Visual pattern recognition by moment invariants [J]. *IEEE Transactions on Information Theory*, 1962, 8(2): 179-187.
- [10] CARUANA R, NICULESCU-MIZIL A, CREW G, *et al.*. Ensemble selection from libraries of Models [C]. *Proceedings of the 21th International Conference on Machine Learning*, 2004, 137-144.
- [11] SAIN S R. *The Nature of Statistical Learning Theory* [M]. New York: Springer-Verlag, 1995.
- [12] 张仁杰, 庄松林, 臧道青. 基于聚类分析与支持向量机模型的缸盖座圈图像判别[J]. 光学精密工程, 2011, 19(10): 2478-2483.  
ZHANG R J, ZHUANG S L, ZANG D Q. Estimation of gap width of cylinder socket sleeve based on clustering analysis and SVM [J]. *Opt. Precision Eng.*, 2011, 19(10): 2478-2483. (in Chinese)
- [13] 张春霞, 张讲社. 选择性集成学习算法综述[J]. 计算机学报, 2011, 34(8): 1399-1410.  
ZHANG CH X, ZHANG J SH. A survey of selective ensemble learning algorithms [J]. *Chinese Journal of Computers*, 2011, 34(8): 1399-1410. (in Chinese)
- [14] KIM H C, PANG S N, JE H M, *et al.*. Constructing Support Vector Machine Ensemble [J]. *Pattern Recognition*, 2003, 36(12): 2757-2767.
- [15] VALENTINI G, MUSELI M, RUFFINO F. Cancer recognition with bagged ensembles of support vector machines [J]. *Neuro computing*, 2004, 56: 461-466.
- [16] DUIN R P W. The combining classifier: to train or not to train [C]. *16th International Conference on Pattern Recognition*, 2002, 2: 765-770.

## 作者简介:



**周 涛**(1977—),男,宁夏同心人,博士,副教授,硕士生导师,2010年于西北工业大学获得计算机科学与技术专业工学博士学位,主要从事多模态医学图像融合识别,智能计算,数据挖掘方面的研究。E-mail: zhout123@gmail.com



**陆惠玲**(1976—),女,河北定兴人,硕士,讲师,2010年于西安工业大学获得计算机应用专业硕士学位,主要从事医学图像处理、数据挖掘等方面的研究。E-mail: lu\_huiling@163.com



**陈志强**(1971—),男,宁夏银川人,硕士,副教授/副主任医师,1997年于宁夏医科大学获得学士学位,2007年于宁夏医科大学获得硕士学位,目前主要从事于腹部和神经系统 MR 功能影像研究。E-mail: czq642000@163.com



**马 苗**(1977—),女,山东聊城人,博士,教授,硕士生导师,2005年6月于西北工业大学获得信号与信息处理专业工学博士学位,2009年6月于西北工业大学博士后流动站完成博士后研究工作,目前主要从事图像处理、智能计算和灰色理论等方面的研究。E-mail: mmthp@snnu.edu.cn

(版权所有 未经许可 不得转载)