

·技术研究·

模糊遗传神经网络算法提取乳腺微钙化点的效果

鲁 雯^{1,2}, 张光玉²

(1. 同济大学电子与信息工程学院, 上海 200092; 2. 泰山医学院放射学院, 山东 泰安 271016)

摘要:【目的】微钙化点是早期乳腺癌的重要征象之一,本研究联合运用遗传算法、模糊数学和人工神经网络,建议一种乳腺微钙化点提取的新方法,为乳腺病变的自动识别提供前期处理,为早期乳腺癌的临床诊断提供帮助。【方法】首先利用随机方法产生大量的样本,然后,利用模糊遗传算法对产生的随机样本进行分类,将分类后的样本输入人工神经网络进行训练,将 310 幅乳腺图像的感兴趣区域输入训练后的人工神经网络分类器进行分类。【结果】与微钙化点提取方面的同类文献相比较,结果表明该算法在相同误检率下得到较高的阳性检出率。【结论】研究表明综合运用遗传算法、模糊数学和人工神经网络进行乳腺微钙化点提取比单纯运用人工神经网络提取效果好。

关键词:遗传算法; 人工神经网络; 微钙化点; 模糊数学

中图分类号: TP391

文献标识码: A

文章编号: 1672-3554(2008)02-0216-05

Effect of Fuzzy Genetic Artificial Neural Networks Algorithm Extracting Microcalcification

LU Wen^{1,2}, ZHANG Guang-yu²

(1. College of Electronic and Information Engineering, Tongji University, Shanghai, 200092;
2. College of Radiography, Taishan Medical University, TaiAn 271016, China)

Abstract: 【Objective】 Microcalcification is one of the most important characteristics of early breast tumors. In this paper, we proposed a new method of microcalcification detection by integrating genetic algorithm, fuzzy mathematics and artificial neural networks. The method could provide preprocessing for automatic recognition of breast cancers, and assist doctors to diagnose early breast cancer. 【Methods】 A lot of random training samples were firstly produced; then, these samples were classified into the background and microcalcifications using the fuzzy genetic method. Finally, the 310 regions of interest were classified into the background and microcalcifications using the trained neural networks. 【Results】 Compared with similar literature about microcalcification detection, we obtained better positive detection ratio with the same false detection ratio. 【Conclusions】 Experimental results demonstrate that our method obtain better extraction effect by integrating genetic algorithm, fuzzy mathematics and neural network, compared with the method simply using artificial neural network.

Key words: genetic algorithm; artificial neural networks; microcalcification; fuzzy mathematics

[J SUN Yat-sen Univ(Med Sci), 2008, 29(2):216-220,234]

乳腺癌是妇女常见的恶性肿瘤之一,发病率逐年升高^[1],已成为威胁人类健康的一大杀手。早发现、早治疗是乳腺癌防治的重要措施。微钙化点是早期乳腺癌的重要征象之一,因此,微钙化点的自

动检测已成为乳腺癌计算机辅助诊断中的国内外研究的热点和难点问题之一。国外在这方面的研究起步较早,已提出许多算法,主要包括阈值分类法、统计学分类法、人工神经网络分类法、支持向

收稿日期: 2008-01-07

基金项目: 国家自然科学基金(60572113)

作者简介: 鲁雯(1964-),女,博士,教授,研究方向:主要从事医学图像信息分析与处理, E-mail: aluwen@163.com

量机分类法以及混合分类法等^[2]。国内起步较晚,但也提出许多算法,主要集中在利用小波多分辨率分析、神经网络和支持向量机来进行微钙化点的提取。目前已有的算法,虽然取得一定效果,但由于误检率高,阳性检出率低,不能满足临床应用的需求。本文综合运用遗传算法、模糊数学和人工神经网络进行微钙化点的检测,可以在低误检率下得到较高的阳性检出率,比其它单独用人工神经网络检测方法效果好。

1 材料与方 法

1.1 对 象

山东省医学影像研究所提供的 40 个病例,患者全部为女性,年龄 35~64 岁,病理确诊为早期乳腺癌,从 40 病例中选择了 100 幅钼靶 X 线影像,在有经验的乳腺放射诊断专家指导下,从这些图像中选择了包含微钙化点的 310 个不同类型的感兴趣区域。

1.2 训练方 法

1.2.1 训练思路 本文方法的第一步是对人工神经网络进行训练,训练框图如图 1 所示,图 2 是框图中人工神经网络和神经网络分类器的结构图。训练算法可分为以下几步:

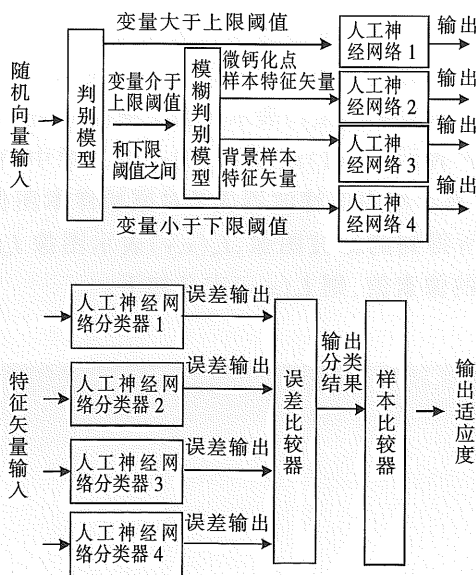


图 1 模糊遗传神经网络训练框图

Fig.1 Fuzzy genetic artificial neural network training diagram

1.2.2 仿真程序产生特征值 利用计算机仿真程

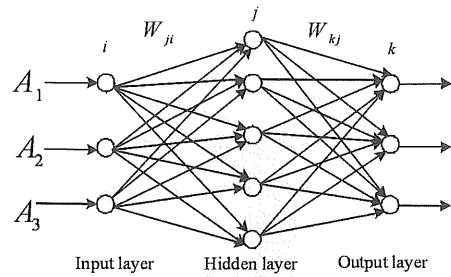


图 2 人工神经网络分类器

Fig.2 Artificial neural network classifier

序产生一组随机向量,每个随机向量都包含 3 个随机变量,分别对应于样本的方差、灰度和残差 3 个特征值,如图 1 所示,将这组随机向量输入判别模型,经判别模型判断后,大于上限阈值 u 的变量组成微钙化点特征矢量输入人工神经网络 1 进行训练,小于下限阈值 v 的变量组成背景特征矢量输入人工神经网络 4 进行训练,如果输入随机变量值位于上限阈值和 u 下限阈值 v 之间,则输入模糊判别模型,经模型中的隶属度函数(模糊函数)判别后,得到微钙化点和背景特征矢量,分别输入人工神经网络 2 和 3 进行训练。隶属度函数如式(a)和(b):

$$\mu_m = \begin{cases} 1 - \frac{2(x-u)^2}{(v-u)^2} & \frac{v+u}{2} < x \leq u \\ \frac{2(x-v)^2}{(v-u)^2} & v < x \leq \frac{v+u}{2} \end{cases} \quad (a)$$

式中, μ_m 表示微钙化点的隶属度, x 表示输入的随机变量, u 和 v 分别表示上限和下限阈值。设 μ_b 表示背景隶属度,则 μ_b 定义为:

$$\mu_b = 1 - \mu_m \quad (b)$$

1.2.3 BP 学习算法训练神经网络 利用 BP 学习算法^[3]对人工神经网络 1-4 训练,将训练后网络的最优权值系数分别输入到人工神经网络分类器 1-4。在 31 个感兴趣区域中随机选择 N_3 个感兴趣区域(ROI)作为样本图像,提取特征后组成样本的特征矢量输入到人工神经网络分类器 1-4,各分类器输出与预设期望值相减得到的误差输入到误差比较器进行比较,如果人工神经网络分类器 1 或 2 的输出误差最小,则分类器输出点为微钙化点,否则,输出点为背景。将分类结果输入到样本比较器与对应样本的手工分类结果比较,比较结果作为遗传算法的适应度输出。设 F_n 表示遗传算法适应

度,定义为:

$$F_n = \sum_{i=1}^S \left(1 - \frac{I_i^{ref} \oplus I_i^{dst}}{N} \right) \quad (c)$$

式中, S 表示样本图像总数, N 表示每个样本图像总的像素数, I_i^{ref} 表示用手工方法对第*i*个样本图像进行微钙化点提取后所得到的二值图像, I_i^{dst} 表示人工神经网络分类器对第*i*个样本图像分类后所得到的二值图像,其中,二值图像的像素取值为0和1,0表示背景像素点的值,1表示微钙化点的像素值。

1.2.4 遗传算法最优解为神经网络赋数值 按照文献[4]的遗传算法获得最佳上限和下限阈值,算法描述如下:把上限阈值u和下限阈值v看作是个体中的染色体进行基因编码,利用计算机仿真程序产生*N*个初始个体,对这些个体中的染色体进行基因编码,将初始个体随机配对并按照一定的交叉率将染色体进行交叉产生后代个体,在初始个体中按一定的几率对个体染色体中的基因进行变异产生新的后代个体,按公式(c)计算每个后代个体的适应度,选择适应度高的个体进行配对和变异产生下一代个体,按照这种方式,不断循环直到达到预设的进化代数,选择进化出的最优个体,将最优个体中的染色体进行基因解码,解码后的数值对应最佳上限和下限阈值。对应于最佳阈值所训练出的网络权值,是最优权值,作为后续的人工神经网络分类器的权值。

1.3 提取特征矢量

利用模糊遗传神经网络对图像进行分类,必须首先提取特征矢量。基于对微钙化点特征的研究,选择微钙化点的方差、残差和灰度3个特征组成输入特征矢量。下面介绍这3个特征的获得。

1.3.1 方差的提取 为了获得微钙化点的方差特

征,首先对样本图像进行边缘检测。本文中利用Kirsch算子^[5]进行边缘提取,由于所得到的边缘图像还有许多不连续的地方,利用边缘连接算法^[6]进行连接。为了减少运算量和防止乳腺中正常组织的干扰,基于图像的边缘图像进行灰度腐蚀运算^[7]。将被腐蚀的图像与原始样本图像相减,得到差值图像 I_d ,求出 I_d 的方差,用 I_d 中的最大方差进行归一化处理,得到微钙化点的输入特征。方差的获取方法如下:设 μ 表示图像 I_d 中的像素平均值, μ 定义如下:

$$\mu = \frac{1}{N} \sum_{y=1}^H \sum_{x=1}^W I_d(x,y) \quad (d)$$

式中, W,H 分别表示差值图像 I_d 的宽和高, $N=W \times H$ 表示图像 I_d 中的总像素数, $I_d(x,y)$ 表示图像 I_d 中 (x,y) 处的像素值。设 $\sigma(x,y)$ 表示图像 I_d 中像素 (x,y) 的方差, $\sigma(x,y)$ 定义如下:

$$\sigma(x,y) = [I_d(x,y) - \mu]^2 \quad (e)$$

基于式(e)对差值图像 I_d 中的每一像素点进行计算,得到差值图像中各像素点的方差值。图3的方差图像显示,微钙化点与背景像素差值更加明显。

1.3.2 残差的提取 由于边缘提取得到的边缘不仅包含微钙化点边缘,还包含对比度较大组织和噪声点边缘,仅仅利用上述算法得到的微钙化点的方差特征进行分类,还存在较大的误检率,为了降低误检率,利用中值滤波和快速傅立叶变换提取微钙化点的残差和灰度特征。设 I 表示待滤波的原始样本图像, $I(x,y)$ 表示图像 I 中 (x,y) 处的像素值, MF 表示 $(2k+1) \times (2k+1)$ 的一个二维中值滤波器, I_m 表示用中值滤波器 MF 对原始样本图像 I 滤波后所得到的滤波图像, $I_m(x,y)$ 表示图像 I_m 中 (x,y) 处的像素值,则 $I_m(x,y)$ 定义如下:

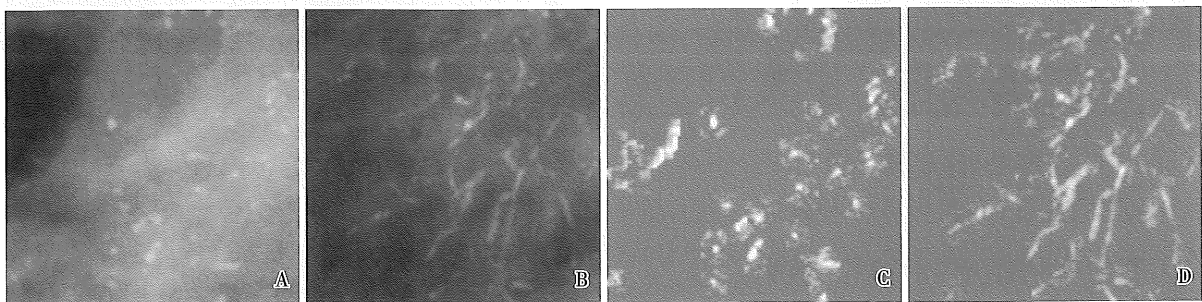


图3 图像求方差结果

Fig.3 Results of variance images

A,B: original sample images; C,D: variance images

$$I_m(x, y) = \text{Median}\{I(x+i, y+j)\} \quad (f)$$

式中,符号 *Median* 表示求像素的中值,和是整数,且 $i \in [-k, k]$, $j \in [-k, k]$,用 $(2k+1) \times (2k+1)$ 的二维中值滤波器对图像 *I* 滤波。具体方法描述如下:首先对图像 *I* 进行列边界像素的对称性扩展,扩展像素数为 *k*,然后对扩展后的图像进行边界像素的对称性扩展,扩展像素数也为 *k*,对图像 *I* 中的每一个像素点 (x, y) ,取一个以该像素点为中心的 $(2k+1) \times (2k+1)$ 区域,基于式(f)得到该区域的像素中值,用该中值代替 (x, y) 处的像素值,用这样的方式处理每一个像素点,就可以得到中值滤波图像。将原始样本图像 *I* 与滤波图像 I_m 相减得残差图像,利用最大残差值对残差图像中的像素值归一化,得到微钙化点的第二个输入特征。滤波后的图像与原始图像相减得到的残差图像中,在微钙化点处的像素值将明显高于周围的背景区域,可以较容易地分割微钙化点,因此,残差值可以作为微钙化点的输入特征。图4 A、B 是中值滤波的结果,图像中,微钙化点基本都被滤除。C、D 是残差图像,可以看出微钙化点的像素值明显高于背景像素值,非常有利于对微钙化点的分类。

1.3.3 灰度的提取 微钙化点在频域介于低频背

景和极高频噪声之间,利用这一特点可以对微钙化点的特征进行增强,然后,提取灰度特征。提取方法描述如下:利用二维傅立叶变换将原始样本图像 *I* 变换到频域,将频域的低频和极高频部分分别乘以小于1的衰减系数后进行二维傅立叶反变换,获得微钙化点相对增强的图像。由于低频背景和极高频噪声得到衰减,在反变换后的图像中微钙化点的灰度值得到相对增强,高于周围的背景灰度值,这样可以较容易地分离微钙化点和背景,因此,经过傅氏正反变换微钙化点得到增强的图像中,灰度值可以作为输入特征进行提取。图5的增强图像显示,微钙化点灰度值高于背景。

输入感兴趣区域,利用本文的提取算法得到每一像素点的输入特征,归一化后组成特征矢量,输入到图1所示的已经训练好的人工神经网络分类器1-4中,网络输出与预设的期望值相减得到输出误差,将误差输入到误差比较器进行比较,若人工神经网络分类器1或2的输出误差最小,则该像素点为微钙化点,否则,该像素点为背景,利用这种方法,逐像素点进行检测,就可以将感兴趣区域中的微钙化点提取出来。

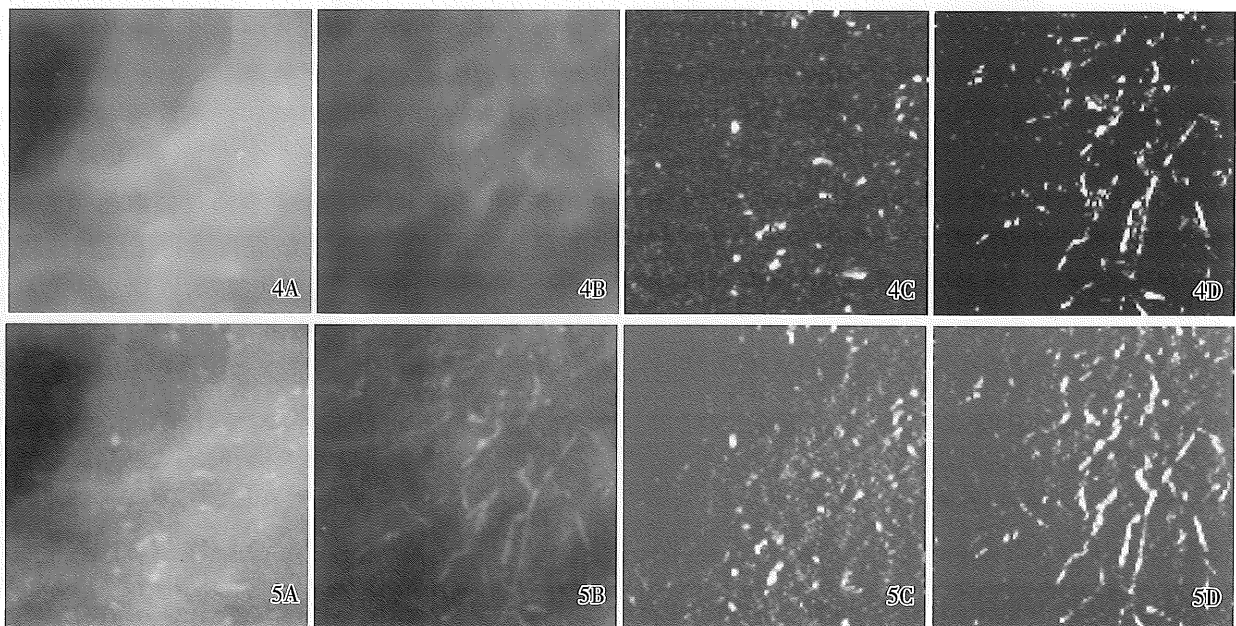


图4 中值滤波结果和残差图像

图5 微钙化点增强后的图像

Fig.4 Results of median filtering and residual error image

A, B: Results of median filtering; C, D: difference images

Fig.5 Enhanced images with microcalcifications

A, B: original sample images; C, D: enhanced images

2 结 果

用本文算法对 310 个感兴趣区域进行了检测,图 6 是部分实验结果,图中,A、B 是原始样本图像,C、D 是微钙化点提取结果。从图中可以看出,算法能够较完整地提取出微钙化点,阳性检出率较高,误检率低。本文运用遗传算法确定模糊隶属度中的最优上下限阈值,然后利用模糊数学的方法对样本进行分类,分类后的样本输入人工神经网络进行训练学习,这样训练出的人工神经网络

具有较强的识别能力,通过对个不同类型感兴趣区域中微钙化点的检测发现,混合人工神经网络分类器比单一的人工神经网络分类器效果好。

为了从统计学的角度判断本文算法的优劣,从 310 个不同类型的感兴趣区域中选取了有经验的放射诊断专家确认的 7 000 个微钙化点,通过改变边缘提取的阈值和训练样本集绘制了本文算法的 ROC 曲线,如图 7 所示。从图中可以看出,算法能在误检率为 0.5%时,得到 94%的阳性检出率,高于文献[10]单独用神经网络提取所报道 88.9%的阳性检出率。

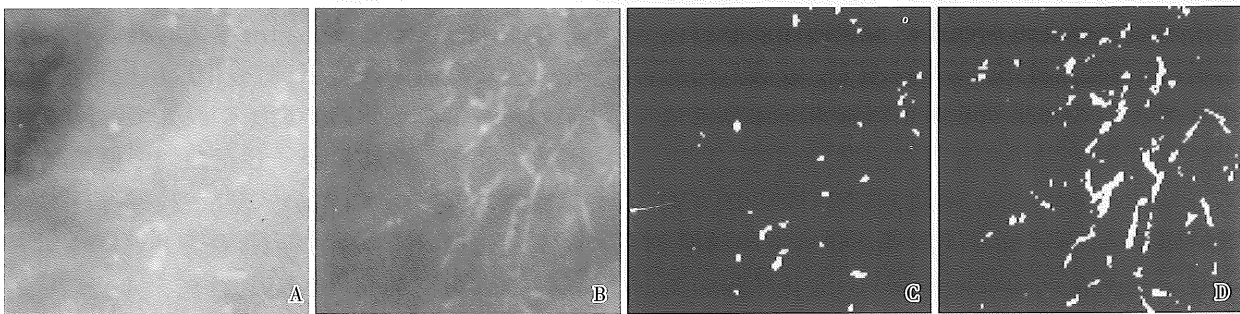


图 6 微钙化点提取结果

Fig.6 Results of microcalcification extraction

A、B: original sample images;C、D: images with extracted microcalcifications

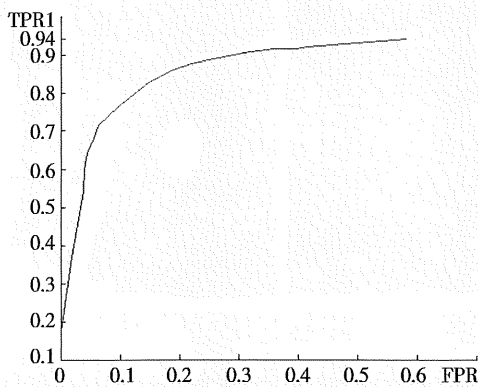


图 7 算法对诊断微钙化点的 ROC 曲线

Fig.7 ROC curve of algorithm diagnosing microcalcification

3 讨 论

由于微钙化点的形态和种类多种多样,密度相差也很大,使得微钙化点的检测成为乳腺计算机辅助诊断中的一个难点和热点问题。目前已有的算法中,阈值分类方法^[8]是简单地用一个或几个阈值将乳腺图像中的微钙化点检测出来。由于乳

腺图像复杂多样,很难找到一个合适的阈值满足绝大多数微钙化点检测的需要,因此,阈值法检测微钙化点有很大的局限性。这类算法通常在误检率为 0.5%时,阳性检出率为 80.5%,漏检率为 19.5%。统计学分类法^[9]主要通过提取图像的统计特征和利用统计学分类法进行微钙化点的检测,该方法通常计算量较大、方法复杂,该算法通常在误检率为 0.5%时,阳性检出率为 82.5%,漏检率为 17.5%。人工神经网络分类方法^[10]首先随机选择样本对网络进行训练,提取感兴趣区域中微钙化点的各种特征组成输入特征矢量,输入训练后的神经网络进行分类,该算法通常在误检率为 0.5%时,阳性检出率为 88.9%,漏检率为 11.1%。支持向量机分类法^[11]是基于结构风险最小的原则构造学习准则,然后对网络进行训练。该方法能在有限的训练样本下,得到全局最优解,具有更大的优越性,但该算法中,核函数的选择是难点。该算法通常在误检率为 0.5%时,阳性检出率为 94%,漏检率为 6%。本文算法综合运用了遗传算法、模糊方法和人工神经网络