

基于SVM算法的乳腺X光片辅助诊断系统的设计与实现

牛琳, 张雨薇, 张露馨

(海南医学院医学信息学院, 海南 海口 571199)

摘要: 运用MATLAB软件对乳腺癌X光片做预处理、特征提取、基于SVM算法的分类器识别及辅助诊断。经反复实验完成纹理特征提取、核函数选取、参数优化、样本优化等步骤,为影像学医师的诊断预测提供一种便捷有效的辅助方式,减轻了医生的工作量,提高了医生的工作效率,降低漏诊误诊的机率。

关键词: 乳腺癌X光片; 辅助诊断; SVM算法; 特征提取

中图分类号: TP311 **文献标识码:** A

Design and Implementation of the Breast X-Ray Assisted Diagnosis System Based on SVM Algorithm

NIU Lin,ZHANG Yuwei,ZHANG Luxin

(Hainan Medical College,Haikou 571199,China)

Abstract: MATLAB software is used to preprocess breast cancer X-rays,feature extraction,classifier recognition based on SVM algorithm,assisted diagnosis and process information storage.After repeated experiments,texture feature extraction,kernel function selection,parameter optimization,sample optimization and other steps have been completed.It provides a convenient and effective assisting way for imaging doctors to diagnose and predict,reduce the workload and improves the efficiency of doctors,and reduces the chance of misdiagnosis.

Keywords: breast X-ray;assisted diagnosis;SVM algorithm;feature extraction

1 引言(Introduction)

当前,健康问题变得愈加备受关注,影响健康的因素有很多,例如各方面的压力、饮食习惯、特殊环境影响等。正因如此,越来越多的女性,患上乳腺癌^[1]。按照WHO 2014年报道,乳腺癌是中国女性最常见的癌症,在全球范围内,中国占据新诊断乳腺癌病例的12.2%,占据乳腺癌死亡的9.6%^[2]。欧美国家显然是这些年乳腺癌发病率筛查最为成功的国家,乳腺癌的病死率已经逐年在下降。而在我国,乳腺癌的排查技术和早期预防乳腺癌疾病的意识不足,中国新发病例在全世界乳腺癌新发病例的占比为12%^[3]。超声检查是乳腺疾病诊断中常规的检查方式,在乳腺疾病超声图像中,进行定量化分析是十分重要的,可以找出病变区域,辅助治疗,减小误差。对于乳腺疾病的超声图像,区别与诊断,主要依赖医生的

经验,由于没有量化指标,而且有时判断会出现误差^[4]。针对这些问题,国内外学者提出了很多计算机辅助诊断系统,以提高乳腺超声图像诊断的效率与正确率。计算机辅助诊断系统可为临床医生的诊断供给较为真实有效的辅助讯息,尤其是在提升乳腺癌诊断准确率方面施展着重要的功能,是初期检测乳腺癌的重要方式^[5]。本文主要研究基于SVM算法的乳腺X光片辅助诊断系统,运用该系统能够减少影像科医师的工作量,提高其工作效率,缩小漏诊误诊的几率,为影像学医师的诊断预测提供一种便捷有效的方式。同时结合影像学医师丰富的经验能力可以有效避免患者错过治疗的最佳时间。

2 相关技术研究(Research on related technology)

2.1 特征提取研究

图像的特征提取是模式识别算法中一个及其重要的部

分, 它的目的是用于量化目标的重要特性, 比如说图像的纹理、灰度、亮度、边缘、轮廓、形状等都是相对于图像的特征, 在模式识别的算法中, 被量化的图像特征将作为智能算法机器学习的输入, 构成特征空间, 学习算法将在该特征的基础上进行学习, 进而构建所需的分类器模型^[6]。

乳腺的特征提取一般有纹理特征、空间集方法提取乳腺肿块特征、光学特征、多种图像的底层全局特征、基于边缘领域的乳腺肿块的特征提取方法。

2.2 支持向量机(SVM)分类算法研究

计算机辅助诊断在临床上对早期乳腺癌的检测中起到很大的作用, 常用的算法有: 支持向量机、人工神经网络分类、模糊逻辑分割、小波变换处理、统计学特征提取等, 以及这几种方法的结合智能算法处理方法^[7]。

支持向量机(Support Vector Machine, SVM)是一种模式识别算法, 它根据统计学习理论, 主要是VC维理论和结构风险最小原理, 在复杂的模型和学习的能力中探求最佳中性, 以便达到最优推行^[8]。SVM的特点是解决小样本的问题。

影响支持向量机分类方法的因素主要有: 核函数的选取和参数的优化。核函数包括径向基函数、多项式函数、S形函数。主要的参数中惩罚因子的数值取值决定了支持向量机的准确率。惩罚因子的数值越大就代表了对训练误差值的惩罚越大^[9]。当它的值太小时, 会导致训练误差较大; 取值过大时, 虽然会降低误差值, 但会导致过度拟合的情况。

3 系统设计与实现(System design and implementation)

3.1 系统总体设计

基于SVM算法的乳腺癌X光片辅助诊断系统主要分为四个模块, 分别是样本图像预处理、特征提取、分类器处理和诊断结果。具体的系统实现步骤为: 首先是训练建模阶段, 导入待训练的乳腺癌X光片作为SVM分类器的训练样本, 对图像进行预处理, 依据乳腺癌X光片特点提取纹理特征, 用于SVM分类器训练, 经过反复调整参数, 优化样本等步骤, 建立符合乳腺癌X光片分类的SVM模型。然后进入测试模型阶段, 导入测试样本, 经过图像预处理、特征提取后在训练好的SVM模型中进行分类判断。系统流程图如图1所示。

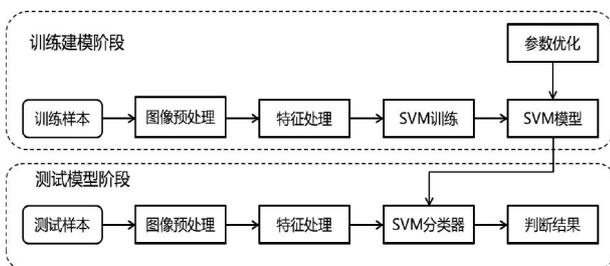


图1 系统流程图

Fig.1 System flow chart

3.2 乳腺癌X光片的预处理和特征提取模块

超声检查是乳腺疾病诊断中常规的检查方式, 这样的成像机制会在图像上产生一些斑点状的噪声, 也会造成图像的亮度分布不均匀。图像预处理的功能是除去图像中无关的信息, 寻找有价值的数值信息, 增强信息的可检测性, 进而提高特征提取、分类识别的准确率。本文的图像预处理采用中值滤波算法进行图像去噪。中值滤波法的优点: 抑制效果很好, 画面的清晰度基本保持; 缺点: 对高斯噪声的抑制效果不是很好。中值滤波法图像去噪效果如图2所示。

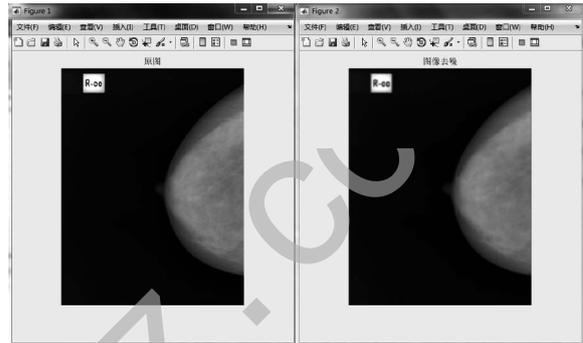


图2 图像去噪前后对比图

Fig.2 Image denoising contrast diagram

得到消除噪点的图像后, 可以更清楚的观察乳腺肿块的大小和形状, 但有部分肿块的边缘比较模糊, 无法直接进行特征提取。此时需要通过图像增强方法增强乳腺X光图像的明暗度对比, 使肿块的边缘更清晰, 方便医生观察与计算机的特征提取。本文选择了直方图均衡化进行图像增强, 效果如图3所示。

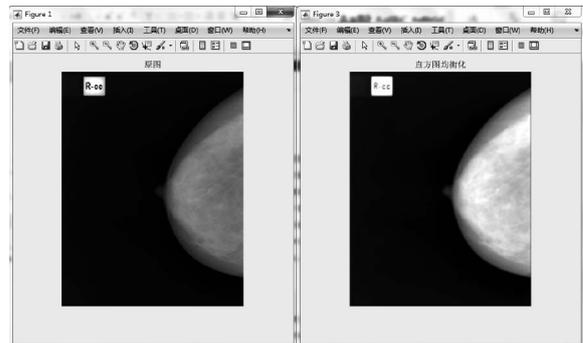


图3 图像增强前后对比图

Fig.3 Image enhancement contrast diagram

乳腺癌X光片中, 识别肿块可以通过观察, 一般局部相对较亮, 形状接近于椭圆状, 具有无明显缝隙内核的星芒状物或实质团块^[10]。

乳腺癌的特征提取一般有纹理特征、空间集方法提取乳腺肿块特征、光学特征、多种图像的底层全局特征、基于边缘领域的乳腺肿块的特征提取方法。在本文中, 使用乳腺的纹理特征作为图像特征表示乳腺病灶区域, 采用灰度共生

矩阵方法提取纹理特征。灰度共生矩阵描述的是具有某种空间位置关系的两个像素的联合分布。为了定义描述纹理的方法，前人使用过三种主要方法，本文所采纳的是基于共生矩阵的纹理特征提取。

纹理提取的步骤如下：

(1)灰度级量化：依据人的视觉特性，通过部分较粗的纹理特性判定，能够得到大多数图像的相近指数。首先，对原始图像质量进行压缩，然后生成图像共生矩阵。例如，级别分成16、32和64。通过训练发现16级的图像特征提取的效果较为明显。(2)特征值测定：本文选用能量、熵、惯性矩、相关度，四个纹理参考指标。求能量、熵、惯性矩、相关度的均值和标准差作为最终八维纹理特征。(3)特征向量的归一化：因为上述特征值的取值范围不相同，所以要对这些特殊的值进行归一化处理。

由于需要提取特征的图片数量较大，这里设置了批量提取特征的功能，可以一次性提取文件夹内的所有图片特征。使用循环函数，对上述图像处理到特征提取的步骤进行循环。运行结束后数据会生成一个Excel表格，存储在文件夹中，提供给分类器那边进行分类处理。此部分系统操作界面，如图4所示。

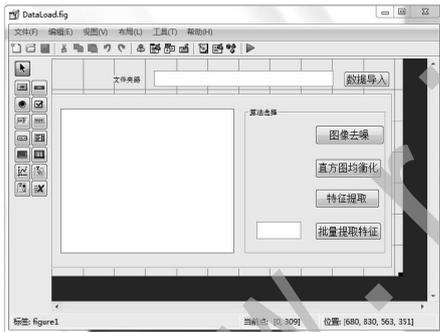


图4 图像预处理与特征提取界面

Fig.4 Image preprocessing and feature extraction interface

提取到的部分图片的特征值，如图5所示。

能量		熵		惯性矩		相关度	
均值	方差	均值	方差	均值	方差	均值	方差
0.428163	0.00116	1.651999	0.02927	0.085823	0.026088	0.028638	2.8E-05
0.343124	0.00166	1.892769	0.036078	0.08981	0.0269	0.028756	2.83E-05
0.299817	0.001238	2.046923	0.039108	0.088165	0.025854	0.028584	2.63E-05
0.252901	0.001584	2.158233	0.042841	0.0945	0.027868	0.027918	2.28E-05
0.604725	0.000643	1.186993	0.021814	0.049138	0.017712	0.042884	3.89E-05
0.551843	0.001108	1.323003	0.026409	0.07771	0.026615	0.042303	4.71E-05
0.443134	0.001078	1.652353	0.037702	0.085921	0.029098	0.03299	3.35E-05
0.374382	0.001488	1.845797	0.043044	0.097037	0.033211	0.031184	3.37E-05
0.317705	0.001944	1.606053	0.028212	0.07481	0.026215	0.06901	9.1E-05

图5 实验数据截图(部分图片的特征值)

Fig.5 Experimental data screenshot(feature values of some pictures)

3.3 乳腺癌X光片的分类器训练模块

提取乳腺X光片中的特征指标，以此输入来训练SVM。

本文使用林智仁教授提出的LIBSVM工具箱，进行样本图片训练和预测。选择径向基函数作核函数，是因为它有很多优点，例如：(1)具有非线性映射性质，可对数据非线性可分情况进行处理。(2)可退化为线性核函数，选取合适的核函数参数及适当的代价系数，径向基函数能发挥与线性核函数同样的作用。(3)参数指标符合需求，S形核函数与径向基函数的算法体现也较为接近，而且挑选S形核函数需要探讨有关的核函数参数^[9]。

分类器处理模块，包括导入训练样本、调试参数、训练样本图像、测试样本图像，判断预设标准和测试待测图像几个步骤。调试参数主要选择核函数和核函数参数的优化，使其训练学习达到最优值，测试样本图像准确率，若样本图像测试结果优良，则需进行测试待测医学图像，否则还需调整优化参数。

本训练建模选用最大分类正确率相对的惩罚因子c和核参数δ，用来支持向量机分类器，c取100，δ取1。

SVM训练界面如图6所示。

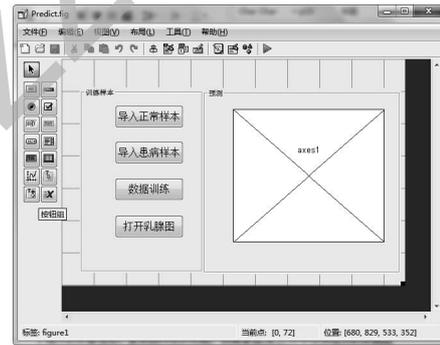


图6 SVM训练界面

Fig.6 SVM training interface

4 实验结果(Experimental results)

本设计选用MATLAB 2017a软件，系统运行截图如图4和图5所示。在训练建模过程中，共采集了180个志愿者的乳腺X光片，共32张正常乳腺X光片，以及148张乳腺肿瘤X光片。采用随机抽取的方式选取乳腺X光片，将所有样本180张图片平均分为六组，前三组作为测试集(即每30张乳腺X光片为一组)剩下三组当作训练组。使用交叉循环法反复训练六次，每次选出一个直接作为测试样本，将六次的交叉循环辨别率的平均值作为实验结果。分类建模完成后，检测150张待测图片准确率达到83.3%。

5 结论(Conclusion)

本文设计和实现了基于SVM算法的乳腺癌X光片辅助诊断系统，该系统的判断准确率还可以从三方面进一步提高。

(1)在模式识别的算法中，被量化的图像特征将作为智能

(下转第42页)