

•研究原著 •

# 以组织病理细胞学和影像结构特征为分析依据的乳腺导管增生 计算机辅助诊断系统

陶永鹏,刘朝霞,项 聪(大连外国语大学软件学院,辽宁省大连市 116044) **DOI**:10.3969/j.issn.2095-4344.0328 **ORCID**: 0000-0002-7129-5980(陶永鹏)

## 文章快速阅读:



陶永鵬,男,1981年生, 辽宁省大连市人,硕士, 讲师,主要从事医学图像 处理方面的研究。

中图分类号:R318 文献标识码:B 稿件接受: 2018-06-10



# 文题释义:

**乳腺导管癌的结构特征**:①细胞直径明显大且呈不规则形状;②有乳头状或囊状等形状;③细胞的排列不规则且布满整个导管,内部核仁较大且明显。

普通导管增生结构特征:①细胞排列成流水型结构;②次级管腔大小和形态不规则,细胞间的腔隙不规则分 布;③增生细胞核分布呈现不均匀,但可见不同程度的细胞核重叠;④细胞质染色后颜色深度哟差异,核形 状呈圆、卵圆至梭形,染色色泽和大小均匀。

# 摘要

**背景**: 对疑似乳腺癌诊断而言,细胞组织病理诊断是重要的诊断步骤。若医生需要针对每一张苏木精-伊红 染色影像都需要凭借自己的经验以肉眼进行判断,不但具有较大的主观性,医生的疲劳度和专注力也会影响 整体判断结果。目前国内外一直缺乏一套定量、客观的乳腺癌细胞计算机辅助诊断系统。

**目的**:将发展一套计算机辅助诊断系统,基于多重分辨率的概念,以组织病理细胞学特征和影像结构特征为 分析依据,从而判别乳腺导管细胞增生的类型。

方法:针对患者粗针穿刺切片的乳管组织影像,引入多重分辨率的概念,首先将细胞影像进行 Sigmoid 强化,然后将细胞的形状、大小、排列及颜色均匀度作为计算机辅助诊断系统的判别特征依据,利用最大期望 算法结合 Lab 色彩空间对细胞影像进行颜色分离,初步分割得到感兴趣的细胞核;针对细胞核不完全分割 的问题,采用空洞修复算法进行解决,对于相邻细胞核重叠的问题,采用分水岭转换进行解决,最终成功分 离得到候选细胞核。一旦细胞完整呈现后,将进行候选细胞核特征获取,采用椭圆拟合、图论以及纹理特征 等方法取出需要的细胞核特征,通过特征选择和线性判别分析进行特征获取和筛选,最后由支持向量机分类 器进行乳腺导管细胞增生的类型判别,判断是否存在癌症活动。

结果与结论:内导管增生病变判读的实验结果证实,当使用特征选取配合线性判别分析筛选得到 14 维度的特征维度,并使用支持向量机-径向基函数作为分类器时,此文所设计的诊断辅助系统的准确率高达 88.4%,能够作为疑似乳腺癌诊断的辅助系统。

#### 关键词:

乳腺癌;导管增生;计算机辅助诊断

主题词:

乳腺肿瘤;诊断,计算机辅助;组织工程

基金资助:

大连外国语大学创新团队资助项目(2017CXTD01)

Tao Yong-peng, Master, Lecturer, School of Software, Dalian University of Foreign Languages, Dalian 116044, Liaoning Province, China

# A computer-aided diagnosis system of breast ductal lesion based on histopathological and imaging characteristics

Tao Yong-peng, Liu Zhao-xia, Xu Cong (School of Software, Dalian University of Foreign Languages, Dalian 116044, Liaoning Province, China)

# Abstract

**BACKGROUND:** Histopathologic diagnosis of cells is an important method for the diagnosis of suspected breast cancer. If doctors need to use their own experience to judge each hematoxylin-eosin stained image with naked eyes, it is not only subjective, but also doctors' fatigue and concentration affects the judgment results. At home and abroad, there is a lack of quantitative and objective computer-aided diagnosis system for breast cancer cells.

**OBJECTIVE:** To develop a computer-aided diagnosis system, based on the concept of multiple resolution, histopathological cytology and image structural characteristics, in order to identify the types of breast ductal hyperplasia.

**METHODS:** In order to introduce the concept of multi-resolution in the image of breast tube tissue of the patient's thick needle puncture section, the cell image was first subjected to Sigmoid enhancement. Then the shape, size, arrangement and color uniformity of the cells were used as the discriminative features of the computer-aided diagnosis system, and the maximum expectation algorithm combined with the Lab color space were used for color separation of cell images, the cell nucleus of interest were preliminary segmented. The holes & cracks repair algorithm was used for the problem of incomplete nucleus division, and the watershed conversion was used for the problem of overlapping adjacent nuclear cells. Finally, the candidate cell nucleus was successfully isolated. The candidate nuclear feature acquisition was performed. Ellipse fitting, graph theory, and texture features were used to extract the required nuclear features, feature selection and linear discriminant analysis for feature acquisition and screening, and finally classified by the support vector machine. The device identifies the type of breast ductal cell hyperplasia to determine whether cancer activity exists.

**RESULTS AND CONCLUSION:** The experimental results of the interpretation of ductal hyperplasia confirmed that when using feature selection and linear discriminant analysis to obtain the 14-dimensional feature dimension and using support vector machine-radial basis function as a classifier, the accuracy of the diagnostic aid system designed in this paper was verified. With an accuracy of 88.4%, it can be used as an auxiliary system for the diagnosis of suspected breast cancer.

Subject headings: Breast Neoplasms; Diagnosis, Computer-Assisted; Tissue Engineering Funding: the Innovation Group Project of Dalian University of Foreign Languages, No. 2017CXTD01

# 0 引言 Introduction

乳腺癌是由乳腺管细胞或腺泡细胞不正常分裂、繁殖 后所形成的恶性肿瘤,是全世界女性最常见的癌症之一。 世界各地1年约有50万人死于乳腺癌,但是如果能在早期 就及时诊断发现并治疗,其治愈的成功率高达80%<sup>[1-2]</sup>。

乳腺癌形成原因最初是乳腺导管细胞增生分化,其中, 初步乳腺导管细胞增生分化又分为3个类型:普通导管细胞 增生、非典型性导管细胞增生和导管原位癌<sup>[3-5]</sup>。临床上普 通导管细胞增生是一种良性导管增生病变,轻度普通导管 细胞增生并无临床意义,但是重度普通导管细胞增生将使 乳导管内被增生的细胞所填满,造成与非典型性导管细胞 增生的判别困难。此外,临床上对于非典型性导管细胞增 生和导管原位癌的辨别也十分困难,某些较严重的非典型 性导管细胞增生病例会具备低级导管原位癌细胞和结构特 点,但病变范围和严重性还不能诊断为导管原位癌。一般 情况下,小范围轻度非典型性导管细胞增生危险不大<sup>[6]</sup>。 因此,对于乳腺导管细胞增生类型判别的准确与否,对于 患者是否需要接受进一步妥善治疗具有重要意义。

从乳腺癌的细胞研究分析可知,普通导管细胞增生、非 典型性导管细胞增生和导管原位癌在每一阶段的细胞形态、 排列特性都不相同,若医生需要针对每一张苏木精-伊红染 色影像都需要凭借自己的经验以肉眼进行判断,不但具有较 大的主观性,医生的疲劳度和专注力也会影响整体判断结 果<sup>[7-9]</sup>。因此,文章将发展一套计算机辅助诊断系统,基于 多重分辨率的概念,以组织病理细胞学特征和影像结构特征 为分析依据,从而判别乳腺导管细胞增生的类型。

# 1 乳腺导管增生系统构架 Systematic framework of breast ductal hyperplasia

1.1 系统基本构架 由于苏木精-伊红染色的效果在每家 医院都不尽相同,细胞影像会因取像环境因素而改变,为 解决此问题, 文章引入多重分辨率的概念, 首先将细胞影 像进行Sigmoid强化, 解决染色对比不均的问题; 将普通导 管细胞增生、非典型性导管细胞增生和导管原位癌的细胞 形状、大小、排列及颜色均匀度作为计算机辅助诊断系统 的判别特征依据,利用最大期望算法结合Lab色彩空间对 细胞影像进行颜色分离,初步分割得到感兴趣的细胞核。 然后,针对细胞核不完全分割的问题,采用空洞修复算法 进行解决;对于相邻细胞核重叠的问题,采用分水岭转换 进行解决,最终成功分离得到候选细胞核。一旦细胞完整 呈现后,将进行候选细胞核特征获取(包括颜色特征、形状 与结构特征、纹理特征等3类),采用椭圆拟合、图论以及 纹理特征等方法取出需要的细胞核特征,然后通过特征选 择和线性判别分析进行特征获取和筛选,最后由支持向量 机分类器进行乳腺导管细胞增生的类型判别。系统的构架 流程图如图1所示。

1.2 乳腺导管增生计算机辅助诊断系统的系统设计

1.2.1 苏木精-伊红乳腺导管细胞影像的处理

(1)苏木精-伊红乳腺导管细胞影像的对比强化: 经苏 木精-伊红染色的乳房细胞切片组织会有下列特征: 细胞核 部分会呈现深紫蓝色, 细胞质、间质会呈现淡红色至深红 色。但由于部分导管细胞影像会因取像环境、设备造成染 色质量不同,造成细胞核颜色并不明显, 而不能直接进行 实验分析,如果不加强颜色对比,在后续色彩分割处理上 会存在困难<sup>[10]</sup>。文章采用Sigmoid Filter强化影像的颜色对 比,实际处理结果对比如图2所示,增加对比后的影像不仅 可增加后续色彩空间的分离性,在执行后续分割步骤时, 也可顺利地将影像分离为细胞核、细胞间质(坏死细胞或红 血球等)以及背景或者杂质等,有助于对细胞核进行特征获 取<sup>[11]</sup>。

(2)影像的分割:在得到经Sigmoid对比强化的影像后, 就可进行影像分割。此系统采用RGB[由红色通道(R)、绿 色通道(G)、蓝色通道(B)组成]颜色配合Lab色彩空间对影 像进行分割。采用Lab色彩空间的原因,是由于苏木精-伊 红彩色影像的颜色变化不大,其色泽主要偏向于白色、粉 红色和深紫色,因此HSV色彩空间中的H(色度)和S(饱和度) 对于苏木精-伊红影像的分割并无太大帮助<sup>[12]</sup>。图3为Lab 色彩空间成分影像。

然后进行影像分割,其过程为[13-16]:

首先对每一个象素点的RGB及Lab色彩空间中共六成 分参数作为训练的特征向量,以ColorFeatureSet表示,其 组成公式为:

 $ColorFeatureSet = \{L, a, b, r, g, b\}$ (1)

将向量导入最大期望算法中,寻找最大似然估计参数, 最大期望算法经2个步骤交替运算后,会收敛为设定的类 别,其中E步骤会将RGB和Lab 6个参数值导入期望最大化 (expectation-maximum, EM)算法进行群处理,即将RGB 和Lab作为一个数列X = [μ1, μ2, …, μk],然后根据E步 骤[公式(2)]计算4个期望值E,其中是μ均值,而σ为标准差, 分别用这4个期望值作为阀值,再通过M步骤[公式(3)]得到 max ln P(X|θ),然后计算数据X落在个别群心的概率总和, 得到特定群心k的比例分布,进一步得到X的概率分布,直 到RGB和Lab的值收敛为止,然后得到如**图4**所示的五群物 体。

E Step

$$E[z_{ij}] = \frac{e^{\frac{-1}{2\sigma^2}(x_i - u_j)^2}}{\sum_{n=1}^k e^{\frac{-1}{2\sigma^2}(x_i - u_j)^2}}$$
(2)

M Step

$$P(x_{n}|\theta) = \sum_{k=1}^{K} P(x_{n},k|\theta) = \sum_{k=1}^{K} P(k|\theta) = \sum_{k=1}^{K} c_{k} P(x_{n}|k,\theta)$$
(3)

由于苏木精-伊红影像常会出现细胞与背景或导管颜 色相近的情况,因此将影像分为5群后,应再进行颜色标识 和筛选,其中深与浅细胞核象素分别标示为红色与蓝色, 深与浅的间质细胞象素分别标示为绿色与黄色,其余属于 杂质、脂肪背景象素等则标示为白色,如图4B所示,分群 效果显著<sup>[17]</sup>。此系统将选取其中的蓝色与红色2类象素,其 余象素会被当作背景而去除。 (3)苏木精-伊红细胞核的形态学处理:在分群后,可 能在一些细胞核内颜色较淡的象素会被作为背景去除,间 接造成细胞核内破洞的产生,使得后续分析细胞形状时造 成误判,因此必须对候选细胞内的象素进行形态学补洞<sup>[18]</sup>。 但如果对较小的细胞核直接利用形态学侵蚀膨胀补洞时会 破坏细胞结构,反而会造成错误分割。为有效对有破洞的 合格细胞核补洞,首先测量细胞核的数量、位置与大小, 作为是否对候选细胞核补洞的依据,再配合物体轮廓为边 界条件,对内部破洞实施膨胀行为,直至接触物体轮廓为 止<sup>[19-21]</sup>。

内部破洞大致可分为两类,一类为相邻细胞间非核所 形成的过大破洞或是破碎杂质所形成的区域,另一类则是 实际细胞内部核所形成的破洞区域<sup>[22-23]</sup>。对第一类情况进 行补洞,将造成后续在分离重叠细胞核时产生过度分割的 错误结果,因此对于这类破洞不能进行补洞处理。因此, 在补洞运算前,首先在分割后影像上标记所有空洞为h<sub>i</sub>(i=1, 2,…, H),每个空洞内的象素则以h<sub>i</sub>(j)表示,I是分割结 果影像,将边界内任一象素设为0,I的边界象素设成1-I,  $H = [R_{I_c}^{\ \ D}(F)]^c$ 为所有洞填完的数字影像,R是膨胀目标区 域,见**表1**。

在计算出第i个细胞核任一个破洞j的面积Area(i, j)后 [公式(4)],如任一破洞面积接近总体核破洞的平均面积, 则对该物体或候选细胞核进行补洞,其过程为如公式(5)所 示,其中d1、d2为细胞核破洞平均面积,A1为细胞核平均 面积,A2为细胞间隙区域的平均面积,而公式(6)是判断补 洞象素是否在细胞核内且需要进行补洞的处理,实际补洞 后的结果如图5所示呈现,同时,补洞过程中相邻细胞核间 隙的空洞区域并不会改变,如图6所示。

$$Area(i, j) = \sum_{x} \sum_{y} h_i(x, y)$$
(4)

$$d_{1} = \|Area(i, j) - A_{1}\|$$

$$d_{2} = \|Area(i, j) - A_{2}\|$$
(5)

$$h_i(j) = \begin{cases} nucleus \ area, \ if \ d_1 \le d_2 \\ The \ cell \ gap \ area, \ otherwise \end{cases}$$
(6)

(4)重叠细胞核的分离:相互接触的细胞会影响分割与 判读结果,因此该系统采用分水岭算法进行相邻细胞核的 分离<sup>[24-26]</sup>。实际分割的结果如图7所示。重叠细胞核会被进 一步分离,再运用噪声去除法滤除处理过小的噪声后,即 可得到细胞核被成功分离的数码影像。

(5)细胞轮廓的侦测:一般而言,正常细胞核的形状通 常介于椭圆到圆之间,但进行上述分割过程后,细胞形状 可能会受到破坏,产生不平滑轮廓<sup>[27]</sup>。因此,该系统将进 行细胞轮廓侦测,首先使用椭圆拟合侦测技术针对选细胞 核进行侦测,将形状偏向导管或者切割时由于角度问题而 造成严重变形的细胞核筛除,然后将通过主动轮廓法得到 细胞的轮廓。主动轮廓法可侦测到细胞核轮廓,而其拓朴 学的原理是将外轮廓视为平面而内轮廓视为高地,由外轮 廓向内轮廓逼近,得到彼此交界线当作真正细胞轮廓<sup>[28-30]</sup>。 该系统利用公式(7)得到侦测轮廓的能量函数E(C),其中 $\alpha$ 、  $\beta$ 、 $\lambda$ 为正常数,  $\alpha$ 和 $\beta$ 可控制曲线的平滑度, $\lambda$ 可吸引曲 线向边界靠近。加入水平集可发展成公式(8),其中c为常 数,k为曲率,而N是曲线的标准法向量,且 $N = -\left(\nabla u / \nabla u \right)$ 。 根据主动轮廓法得到的细胞核的轮廓,测量其面积,针对 面积过大或过小的细胞核进行排除,只保留面积在60象素 与140象素之间的候选细胞<sup>[31]</sup>。轮廓的结果影像见**图8**所 示。

$$E(C) = a \int_0^1 \left| C'(q) \right|^2 dq + \lambda \int_0^1 g \left| \nabla l C(q) \right|^2 dq \tag{7}$$

$$\frac{\partial u}{\partial t} = (c+k)g(l)|\nabla u| + \nabla u \cdot \nabla g(l)$$
(8)

(6)利用细胞形状清除人工杂物:在取得细胞影像轮廓 后,如能进行量测轮廓内的象素数目,可进一步得到细胞 核面积,再以长短轴比、圆形度等观察每个轮廓的形状是 否逼近于椭圆或圆,并以细胞形状是否接近椭圆或圆形作 为排除人工杂物的标准<sup>[32]</sup>。同时,还可以间接得到细胞核 的面积、圆形度和长短轴比等分类特征参数。清除人工杂 物的结果比较如**图9**所示。

# 1.2.2 乳腺导管细胞核特征的获取

(1)基于图论特征的特征获取:一般而言,癌化细胞核 会因细胞增生或病变产生群聚现象,而纹路则为最常用描 述此种特性方式之一,此外图论也可协助评估增生或者癌 化细胞核的空间群聚和分布性<sup>[33]</sup>。如能充分运用这些特性, 能更进一步解释导管原位癌与普通导管细胞增生细胞核的 群聚和分布差异性。根据以往研究发现,导管原位癌的细 胞核群聚性会较为松散,而普通导管细胞增生的细胞核群 聚性会较为密集<sup>[34-35]</sup>。在该系统采用图论为基础的分布特 征技术,测量相邻上皮细胞群聚的程度。

为获取结构特征,该系统使用了分离测量器、生成三 角网格器和最小生成树算法作为细胞群聚的特征参 数<sup>[36-37]</sup>。其中,通过上一节中测量细胞核的圆形特征,先 会得到细胞分布的多边形外接椭圆,从外接椭圆可得到分 离测量器的子特征参数:多边形面积特征(包括面积平均 值、面积标准差、面积最小变异数及面积标准差与平均值 比值)、多边形周长特征(包括周长平均值、周长标准差、 周长最小变异数及周长标准差与平均值比值)、多边形外接 最小椭圆周长特征(椭圆平均值、椭圆标准差、椭圆最小变 异数及椭圆标准差与平均值比值)。

生成三角网格器的子特征参数包括:三角形边长特征 (边长平均值、边长标准差、边长变异数以及边长标准差与 平均值比值)和三角形面积特征(面积平均值、面积标准差、 面积最小变异数以及面积标准差与平均值比值)。

最小生成树算法的子特征包括:边缘长特征(边长平均

值、边长标准差、边长变异数以及边长标准差与平均值比 值)。

此系统运用分离测量器、生成三角网格器和最小生成 树算法图学技术,选取了这3种类型的特征定量参数,共30 个子特征,其结果分布特征如**图10**所示。

(2)细胞核纹理特征的获取:纹理也是细胞核重要的特征之一,为描述细胞增生群聚现象,此系统利用灰阶共生矩阵来进行纹理分析。灰度共生矩阵能反映出关于细胞核内彼此之间分布的纹理方向、相邻间隔或变化幅度的信息<sup>[38-40]</sup>。此系统使用同距离4个不同方向(0°,45°,90°,135°)的灰阶共生矩阵的6个量化特征(能量、变异数、同构型、对比度、熵、相关性),从中测量或评估细胞核亮度分布均匀程度、纹理粗糙度、影像清晰度和纹理沟纹的深浅程度。灰阶共生矩阵算法如**表2**所示。

# 2 乳腺导管增生的类型判别结果分析 Results

该系统的开发基于Windows 7-64bits操作系统,并采 用Visual Studio 2008 Express C++,配合Open CV2.3与 MATLab作为程序开发工具,得到乳腺导管细胞计算机辅 助诊断系统。此文共选取了41张乳房导管细胞影像作为研 究样本,其中普通导管细胞增生细胞影像23张,导管原位 癌细胞影像18张。乳房导管细胞影像经患者同意后,取自 瓦房店第三医院。

前文中得到了细胞核的45个特征向量,但是如果45个 特征向量全部使用,会导致影像特征维度较高,对计算能 力和计算时间提出了较高的要求。因此,有必要对特征向 量进行筛选。此文通过特征选择(Feature Selection)和线性 判别分析进行特征向量的选择。最终,经特征选择筛选后 特征向量为18维,经线性判别分析筛选后的特征向量为17 维,而特征选择+线性判别分析筛选后的特征向量为14维。

因此,此文采用4种特征维度结合支持向量机分类器对 影像进行分类与比较,其中,4种特征维度包括:原始的全 部45个特征维度、特征选择所筛选的18个特征维度、线性 判别分析删选后的17个维度的特征,及特征选择配合线性 判别分析筛选后的14个特征维度。对细胞核增生进行分类 时,采用基于支持向量机分类模型,其主要参数设置kernel function核心为径向基函数核(RBF kernel),迭代次数设置 为25次,k-fold cross-validation设置为6(K=6)时得到最佳 结果。

表3为总共41张影像中,以不同判读方法分析分类结 果。其中,病例影像1-23为普通导管细胞增生病例;24-41 为导管原位癌病例。表中,打×表示以该方法判读时,会 产生伪阳性的误判;打√表示以该方法判读时,会产生伪 阴性的误判结果。比较判别结果可以发现,当采用的特征 维度为45个时,错误的影像为14张;当采用的特征维度为 特征选择选取的18个时,错误张数为12张;当采用特征维 度为线性判别分析选取的17个时,其判读错误张数为9;



图 1 系统的构架流程图

Figure 1 Flow chart of systematic framework



# 图 3 Lab 色彩空间成分影像

Figure 3 Lab color space results

图注:图A为L亮度信道影像;B为A(绿和红对立信道)影像;C为B(蓝和黄对立信道)影像。



图 5 影像补洞后的结果图
 Figure 5 Results after holes & cracks repair algorithm
 图注:图 A 为对比强化影像;B 为期望最大化算法分群后破洞的影像;
 C 为经补洞后的结果影像。



图 7 影像补洞后的结果图 Figure 7 Results after hole repair 图注:图 A 为经补洞后的结果;B 为执行距离变换后的结果;C 为经 分水岭算法分离重叠细胞核后的结果。



图注:图A为 染色对比不佳 的原始影像 (×40):B为 Sigmoid 对比 增强后影像。

图 2 苏木精-伊红染色影像 Figure 2 Hematoxylin-eosin staining results



# 图 4 期望最大化算法分群影像

 Figure 4
 Expectation maximization algorithm results

 图注:图A为Sigmoid影像;B示期望最大化算法分为5群结果图;C为影像分割结果图。



图 6 影像补洞后比较图

 Figure 6
 Comparison after holes & cracks repair algorithm

 图注:图A为对比强化影像;B为期望最大化算法分群后破洞的影像;

 C为细胞经区域补洞后的结果影像。



图注: 图 A 为细 胞分水岭后的结 果影像; B 为主动 轮廓法获取轮廓 与筛选后的结果 影像。

图 8 轮廓的结果影像 Figure 8 Contour images



图注: 图 A 为原 始影像图; B 为 清除杂物后影 像图。

图 9 清除人工杂物结果比较图

 Figure 9
 Images before and after removing artificial sundries

 图注:图A为原始影像图;B为清除杂物后影像图。



图 10 不同图论方法的示意图 Figure 10 Schematic diagram of different kind of graph methods 图注:图 A 为苏木精-伊红原始影像; B 为分 离测量器结构分布图;C 为生成三角网格器 结构分布图;D 为上皮细胞最小生成树算法 结构分布图。

#### 表1 补洞算法

## Table 1 Holes & cracks repair algorithm

命令语句	命令注释
BEGIN HoleFilling (O[]) Procedure	/*O=Candidate nucleus*/
FOR $i \leftarrow 1$ to n do	/*I = Number of candidate hole*/
IF $(d_1 \leq d_2)$ THEN H= $\begin{bmatrix} R^{D}(F) \end{bmatrix}^{C}$	/*Formula 3-5, 3-6*/ /*H-Fill on the Candidate bole*/
ENDIF	
ENDFOR	
END Procedure	

表 3 不同维度下对病例以支持向量机判读细胞影像结果 Table 3 Case judgement results based on support vector machine under different dimensions

病例编号	45 维	18 维	17 维	14 维	病例编号	45 维	18 维	17 维	14 维
1					22				
2	×	×	×		23	×	×		
3					24				
4					25				
5	×	×	×	×	26	$\checkmark$			
6					27	$\checkmark$	$\checkmark$		
7					28				
8		×	×	×	29				
9					30	$\checkmark$	$\checkmark$	$\checkmark$	
10					31				
11					32	$\checkmark$	$\checkmark$	$\checkmark$	
12					33				
13	×				34	$\checkmark$	$\checkmark$		
14					35				
15					36				
16					37	$\checkmark$		$\checkmark$	$\checkmark$
17	×	×	×	×	38				
18					39	$\checkmark$	$\checkmark$		
19					40				
20	×	×	×		41	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
21									

当采用特征选择配合线性判别分析选取的特征维度为14个时,仅有5个影像产生判读错误,因此此种模式的判读错误 率最低。

经过特征选择+线性判别分析将特征维度降维至14维 后,分别以支持向量机分类器的2种kernal function进行比 较,其判别结果如**表**4,5所示。从表中发现,以支持向量 机-linear作为分类器进行分析时,其总体准确率达75.6%; 而支持向量机-径向基函数作为分类器进行分析时,其准确 率则达87.8%。综合可知,此系统确实具有较好的判读效 果,能够实际作为乳腺导管增生的计算机辅助诊断系统。

# 3 结论 Discussion

在临床上,导管原位癌病例会因细胞群聚与形状过于 复杂,即使有经验的医师有时也很难通过人眼进行判别。 此文通过对乳管组织影像进行对比强化、分割、补洞和分 离、侦测细胞轮廓、清楚杂物等处理,并提取乳腺导管细

#### 表 2 灰阶共生矩阵算法

Table 2 Gray-level co-occurrence matrix

~	~	11	1	
TH T	$\sim$	14	'nΠ	
чр.	~	мц	-	

BEGIN Gray Level Co-occurrence Matrix (GLCM[i]) Procedure
Pixel(i, j), d=1, θ[k], k=[0°, 45°, 90°, 135°]
$D1{=}Energy_{Function},\ D2{=}Variance_{Function},\ D3{=}Homogeneity_{Function},$
D4=Contrast <sub>Function</sub> , D5=Entropy <sub>Function</sub> , D6=Correlation <sub>Function</sub>
Do θ[k]→k=0, 45, 90, 135
Do $D_n(i,j)$ {i=1,,256; j=1,, 256} $\rightarrow n=1,, 6;$
OUTPUT 6 D
$Mean(n) = \sum \frac{D_n}{A}, \ n = 1, K, 6$
END Procedure

表 4	不同特征维度下支持向量机-Linear 判别结果
Table	4 Judgement results of support vector machine-Linear
under	different dimensions

特征维度	真正率	假正率	精度	F值	ROC 空间	总体准确率
所有 45 个特征维度	0.389	0.304	0.501	0.437	0.55	56.1%
经付征选择后的 18 个 特征维度	0.501	0.201	0.611	0.545	0.64	03.4%
经线性判别分析后的 17 个 特征维度	0.556	0.217	0.667	0.606	0.67	68.3%
经特征选择和线性判别 分析后的 14 个特征维度	0.667	0.174	0.750	0.706	0.76	75.6%
特征维度 经特征选择和线性判别 分析后的 14 个特征维度	0.667	0.174	0.750	0.706	0.76	75.6%

#### 表 5 不同特征维度下支持向量机-径向基函数判别结果

Table 5 Judgement results of support vector machine- radial basis function under different dimensions

特征维度	真正率	假正率	精度	F值	ROC 空间	总体准确率
所有 45 个特征维度	0.556	0.261	0.625	0.588	0.62	65.8%
经特征选择后的 18 个	0.611	0.217	0.687	0.647	0.71	70.7%
特征维度						
经线性判别分析后的 17 个	0.778	0.217	0.737	0.756	0.81	78.1%
特征维度						
经特征选择和线性判别分析	0.833	0.087	0.882	0.857	0.89	88.4%
后的 14 个特征维度						

胞核的特征,从而建立了乳腺细胞增生的计算机辅助诊断 系统。通过实验发现,具有相当高的实用价值。

**致谢**:感谢在文章写作过程中给予帮助与指点的诸多同仁,感谢 瓦房店第三医院在导管细胞影像处理上给予的帮助。

作者贡献: 陶永鹏负责系统设计、图像处理等主体工作; 顼聪负 责实验设计及实施; 刘朝霞对论文进行了指导和修改。

*经费支持*:该文章接受了"大连外国语大学创新团队资助项目 (2017CXTD01)"的基金资助。所有作者声明,经费支持没有影响文 章观点和对研究数据客观结果的统计分析及其报道。

*利益冲突*: 文中对乳管组织进行粗针穿刺采用巴德空心针穿刺活 检枪, 文章的全部作者声明, 没有接受该器械的任何资助, 在课题研 究和文章撰写过程, 不存在利益冲突。

*机构伦理问题*:该临床研究的实施符合《赫尔辛基宣言》和瓦房 店第三医院对研究的相关伦理要求。

知情同意问题:参与试验的患病个体及其家属为自愿参加,均对 试验过程完全知情同意,在充分了解本治疗方案的前提下签署了"知 情同意书"。

*写作指南*:该研究遵守《观察性临床研究报告指南》(STROBE 指南)。

**文章查重**:文章出版前已经过专业反剽窃文献检测系统进行3次 查重。

**文章外审**: 文章经小同行外审专家双盲外审,同行评议认为文章 符合期刊发稿宗旨。

**生物统计学声明:** 文章统计学方法已经大连外国语大学软件学院 统计学专家审核。

**文章版权**:文章出版前杂志已与全体作者授权人签署了版权相关 协议。

*开放获取声明*:这是一篇开放获取文章,根据《知识共享许可协议》"署名-非商业性使用-相同方式共享4.0"条款,在合理引用的情况下,允许他人以非商业性目的基于原文内容编辑、调整和扩展,同时允许任何用户阅读、下载、拷贝、传递、打印、检索、超级链接该文献,并为之建立索引,用作软件的输入数据或其它任何合法用途。

# 4 参考文献 References

- Mba HS, Ms SKG, Edwin Silverberg BS, et al. Survival experience in the breast cancer detection demonstration project. Ca A Cancer J Clin. 1987;37(5):258-290.
- [2] 吕桂泉. 癌症不可怕: 30 年肿瘤诊治手记[M]. 杭州:浙江大学出版 社, 2009.
- Poller DN, Barth A, Slamon DJ, et al. Prognostic classification of breast ductal carcinoma-in-situ. Lancet. 1995;345(8958): 1154-1157.
- [4] 龚西驗. 有关乳腺导管上皮增生的几个问题[J]. 临床与实验病理学 杂志, 1997,13(4):370-370.
- [5] 辛智芳. 乳腺增生症的分类和诊治[J]. 中华乳腺病杂志(电子版),
   2008, 2(6): 45-48.
- [6] 许艳春,姚晓虹,祝明洁. 乳腺非典型性导管增生的研究进展[J]. 上 海交通大学学报(医学版), 2013, 33(4):516-519.
- [7] Dundar MM, Badve S, Bilgin G, et al. Computerized Classification of Intraductal Breast Lesions Using Histopathological Images.
   IEEE Trans Biomed Eng. 2011;58(7):1977-1984.
- [8] Allison KH, Jensen KC. Intraductal Proliferations (DCIS, ADH, and UDH): A Comprehensive Guide to Core Needle Biopsies of the Breast. Springer International Publishing, 2016.
- [9] Allred DC, Mohsin SK, Fuqua SA. Histological and biological evolution of human premalignant breast disease. Endocr Relat Canc. 2001;8(1):47-61.
- [10] 杜辉,王树迎,徐淑敏. H.E.染色切片质量影响因素分析[J]. 中国兽医 杂志,2003,39(12):50-51.
- [11] Moon WK, Lo CM, Cho N, et al. Computer-aided diagnosis of breast masses using quantified BI-RADS findings. Comput Methods Programs Biomed. 2013;111(1):84-92.
- [12] Basavanhally A, Ganesan S, Feldman M, et al. Multi-Field-of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides. IEEE Trans Biomed Eng. 2013;60(8):2089-2099.
- [13] 江贵平,秦文健,周寿军,等. 医学图像分割及其发展现状[J]. 计算机 学报,2015,38(6):1222-1242.
- [14] 缪仁拉,郑馨,江伟. 基于Lab 空间的白细胞图像偏色校正算法[J]. 电子世界,2017,39(15):153-153.
- [15] 蔡隽,鲍旭东,吴磊,等. 基于活动轮廓模型的彩色白细胞图像自动分 割方法研究[J]. 生物医学工程研究, 2005,24(4):218-222.
- [16] 张艳玲,何鑫驰,李立. 基于最大类间方差与形态学的淋巴结图像分 割[J]. 计算机科学,2013,40(8):296-299.

- [17] 杨俊,郑曲波,吴桂良,等. 基于 Metropolis-SA 算法的脑部磁共振血 管造影图像分割[J]. 生物医学工程与临床, 2013, 17(2): 113-118.
- [18] 冯秀华,左玉霞,唐爱. 彩色多普勒检查乳腺疾病筛选早期乳癌的分析[J]. 医学影像学杂志,2003,13(6):399-401.
- [19] 王蜀,李永宁,陈楷民,等. 基于数学形态学的医学图像分割[J]. 计算 机应用,2005,25(10):2381-2382.
- [20] 赵于前. 基于数学形态学的医学图像处理理论与方法研究[M]. 中南 大学, 2006.
- [21] 夏剑峰. 基于数学形态学的癌细胞的分割与识别[J].电子科技, 2016, 29(10): 11.
- [22] 马东, 曹培杰. 分割重叠细胞核的方法及比较研究[J]. 北京生物医 学工程, 1999, 18(3): 142-147.
- [23] Mouelhi A, Sayadi M, Fnaiech F, et al. Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. Biomed Sign Proc Contr. 2013;8(5):421-436.
- [24] 胡彩亮. 基于指纹图的白细胞图像分割算法研究[D]. 燕山大学, 2011.
- [25] Vincent L, Soille P. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. IEEE Trans Patt Anal Mach Intell. 1991;13(6):583-598.
- [26] 廖苗,赵于前,曾业战,等. 基于支持向量机和椭圆拟合的细胞图像自动分割[J].浙江大学学报(工学版), 2017, 51(4):722-728.
- [27] 王康,李晓春. 基于多空间图像融合的白细胞自动分割[J]. 计算机仿 真, 2015, 32(3): 258-262.
- [28] Caselles V, Kimmel R, Sapiro G. Geodesic active contours. Int J Comp Vis. 1997;22(1): 61-79.
- [29] Pham VT, Tran TT, Chiu YJ, et al. Region-aided Geodesic Active Contour model for image segmentation. IEEE International Conference on Computer Science and Information Technology. IEEE, 2010:318-321.
- [30] Aubert G, Barlaud M, Faugeras O, et al. Image segmentation using active contours: Calculus of variations or shape gradients? SIAM J Appl Math. 2003;63(6): 2128-2154.
- [31] Fatakdawala H, Xu J, Basavanhally A, et al. Expectation-maximization-driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology. IEEE Trans Biomed Eng. 2010;57(7):1676-1689.
- [32] 龚平,郭华雄,王文清,等. 基于人工神经网络的肺癌细胞形态学诊断 模型的建立及应用[J]. 中华生物医学工程杂志, 2012,18(1):73-76.
- [33] 董敏,马义德. 基于乳腺肿瘤细胞形态特征参数的乳腺癌诊断发展 研究[J]. 中华临床医师杂志:电子版, 2013, 7(11):5023-5026.
- [34] 龚西騟. 乳腺活检中热点问题的新认识(一)[J]. 临床与实验病理学 杂志, 2005, 21(1):1-5.
- [35] 姜英,陈杰. 乳腺导管上皮内瘤变(DIN)[J]. 诊断病理学杂志, 2005, 12(6):463-466.
- [36] 马闪闪,董明利,张帆,等. 基于核主成分分析的流式细胞数据分群方 法研究[J]. 生物医学工程学杂志, 2017,34(1):115-122.
- [37] 赵明珠. 细胞病理图像的特征分析与分类识别[D]. 浙江工业大学, 2012.
- [38] Soh LK, Tsatsoulis C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Trans Geosc Rem Sens. 1999;37(2): 780-795.
- [39] Yu J. Texture Image Segmentation Based on Gaussian Mixture Models and Gray Level Co-occurrence Matrix[C]. International Symposium on Information Science and Engineering. IEEE, 2010:149-152.
- [40] 白雪冰,王克奇,王辉. 基于灰度共生矩阵的木材纹理分类方法的研究[J]. 哈尔滨工业大学学报, 2005, 37(12):1667-1670.