

基于形态学特征提取的人工神经网络在口腔鳞状细胞癌诊断中的应用

马开宇, 马开阳, 黎明, 代晓明, 李逸松
(昆明医学院附属口腔医院, 云南昆明 650031)

[摘要] **目的** 基于形态学图像处理方法, 应用径向基神经网络 (radial basis function, RBF) 寻找一种可行、便捷的方法辅助口腔鳞状细胞癌的诊断. **方法** 选择口腔鳞状细胞癌和口腔非癌的组织病理切片图像进行形态学方法处理, 提取表述特征的向量, 作为训练集训练 RBF 网络; 另选择 67 帧病理图像, 包含癌和非癌的病例, 作为测试集观察 RBF 的性能. **结果** 在 RBF 网络将测试标本分类结果的分析中可以看到不同输出值分类阈值的选择对应不同的诊断敏感度和特异度. **结论** 训练后的 RBF 虽然鉴别阳性、阴性的能力不能和金标准 (即病理诊断) 相比, 但是通过选择不同敏感度和特异度, 依然能够有效辅助病理医师, 提高诊断效率, 发挥机器的优势.

[关键词] 口腔鳞状细胞癌; 神经网络 (计算机); 图像处理

[中图分类号] R739.8 **[文献标识码]** A **[文章编号]** 1003 - 4706 (2011) 27 - 0029 - 04

Application of Artificial Neural Network based on Morphometric Feature Extraction in Diagnosis of Oral Squamous Cell Carcinoma

MA Kai - yu, MA Kai - yang, LI Ming, DAI Xiao - ming, LI Yi - song
(The Affiliated Stomatology Hospital of Kunming Medical University, Kunming Yunnan 650031, China)

[Abstract] **Objective** To find a feasible and convenient way to help pathologist to analyze cell features of oral squamous cell carcinoma, an artificial neural network (radial basis function) has been applied based on the morphometric image processing. **Methods** Some images of histopathological sections of patients suffering from oral squamous cell carcinoma and non-carcinoma epithelium were selected to train a RBF network. The network was based on a morphometric method to extract a feature vector. Another 67 images of sections including oral squamous cell carcinoma and non-carcinoma were tested by the trained network to evaluate the performance of RBF network. **Result** Through the analysis of the output of trained RBF network classification, different sensitivity and specificity of diagnosis was achieved by choosing different threshold value correspondingly. **Conclusion** RBF network is a feasible auxiliary tool in the diagnosis of oral squamous cell carcinoma, even though it can not be a precise diagnosis standard.

[Key words] Oral squamous cell carcinoma; Neural network; Image processing

口腔癌在其病理诊断中由于形态学的表述具有模糊的性质, 所以诊断的结果不可避免地带有主观的性质, 另外, 病理诊断过程是一个分析处理大量图像信息的过程, 因此适宜使用计算机技术作为辅助手段. 在上个世纪出现的神经网络方法如今随着计算机技术的发展逐步应用到各个领域, 其中重要

的一个应用就是用于医学上恶性肿瘤的诊断过程中. 神经网络技术 (artificial neural network, ANN) 具有处理超大规模数据的能力, 计算速度快, 能够并行处理数据及数据之间的非线性关系, 用常规方法不能求解时能够有效解决问题, 而且容错性能较好. 细胞学早期诊断技术就是其应用的一个例证,

[基金项目] 云南省科技厅 - 昆明医学院联合专项基金资助项目 (2010CD216)

[作者简介] 马开宇 (1973~), 男, 云南通海县人, 医学学士, 主治医师, 主要从事口腔颌面外科临床工作.

[通讯作者] 黎明. E-mail:remfxp@gmail.com

目前,该技术难于推广的原因之一是脱落细胞采集技术不完善,以及脱落细胞的诊断耗时,复杂^[1]。但这种技术如果得以改善,则带来的变化是巨大的。本实验拟采用形态学特征提取的方法处理病理图像,应用径向基神经网络(RBF)将特征分类,旨在探讨一种简便可行的方法服务于口腔癌的早期筛查,并作为一个有效的辅助手段服务于病理医师的日常诊断中,帮助排除相关疾病以及明确诊断。

1 材料与方法

1.1 材料

选取 21 例口腔鳞状细胞癌患者的病理切片和 21 例非口腔鳞状细胞癌病理切片(HE 染色),口腔鳞状细胞癌病例中包含 5 例高分化鳞癌,4 例中度分化鳞癌,12 例低度分化鳞癌,非口腔鳞状细胞癌样本中包含 8 例正常上皮,6 例过度角化,7 例不典型增生,作为数据库;图像采集系统为 Olympus PM10 全自动摄影机和北京航空航天大学第三代真彩色病理图像分析处理系统;计算机硬件(华硕 X42Intel 酷睿 i5 480M);图像处理软件: Adobe Photoshop CS5 和 Matlab7.0。

1.2 方法

病理标本由图像采集系统连接于光学显微镜上摄片,共生产 42 帧图,分辨率为 2 560 px × 2 048 px, Adobe Photoshop CS5 作为图像裁剪工具,图像选取尺寸定为 320 px × 240 px,裁剪出小图 85 帧。在病理医师指导下选取 9 帧口腔鳞状细胞癌的患者 320 px × 240 px 图像和 9 帧非口腔鳞状细胞癌的病理图像作为训练集,将图像用 Matlab 进行形态学图像处理,完成表述特征的向量的提取,将其作为输入神经元对神经网络进行权值训练,产生 RBF 网络。在剩余的 67 帧病理图像中通过查随机数字表产生随机编号,在双盲的条件下,由病理医师专家组通过阅图给出图像视野中是否有癌细胞;同理,用同样的特征采集程序采集特征向量形成测试数据,通过 RBF 判读得出阳性或阴性结果。

1.3 形态学图像处理

神经网络的应用是基于医学图像的特征提取之上的,提取的方法各有不同^[2-3]。本实验采用如下方法:(1)二值图像的逻辑运算:形态学上重要的运算是:膨胀和腐蚀、开操作和闭操作。对图像做开运算消除了这些杂散点并不改变全局的几何特征^[4]。数学形态学对膨胀和腐蚀运算的定义如下:

$$A \oplus B = \{Z / (\hat{B})_Z \cap A \neq \phi\}$$

$$A \ominus B = \{Z / (\hat{B})_Z \cap A \neq \phi\}$$

开操作就是先对图像进行腐蚀操作,再对其进行膨胀,定义开操作为:

$$A \circ B = (A \ominus B) \oplus B$$

闭操作则是先对图像进行膨胀操作,再对结果进行腐蚀,定义闭操作为:

$$A \bullet B = (A \oplus B) \ominus B$$

对病理图像进行处理后的结果见图 1。

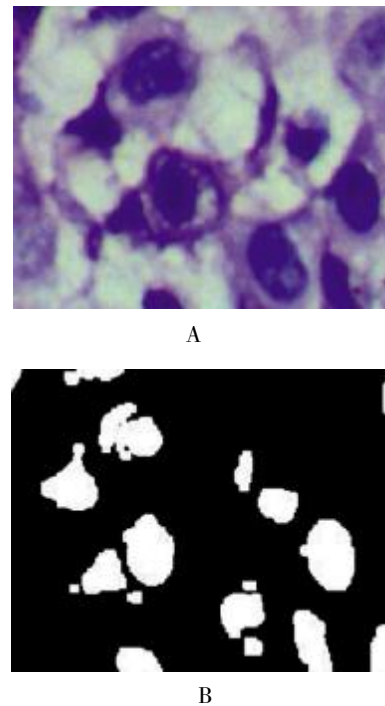


图 1 口腔鳞状细胞癌病理图像经图像处理前后比较
Fig. 1 Comparison of oral squamous cell carcinoma between before and after image processing

A:口腔鳞状细胞癌的镜下图像(放大倍数 400×); B:经过加运算、开运算得到的图像再取反、膨胀后的图像,代表机器提取的特征区域。

(2)形态特征计算:口腔鳞癌细胞形态特征明显,细胞及细胞核具备多形性,细胞具备分裂活性等特点。分化差的鳞癌细胞特征尤其明显:呈圆形或不规则形,成团或散在分布,核质比失常,畸形细胞核增大,居中,核仁增大,数目增多,染色质为粗颗粒状,不均匀。Nipon 等人曾经对显微镜下的骨髓细胞图像进行计算机辅助分析,将细胞核、细胞质、背景有效分割,并提取了细胞形态的 4 个特征值^[5],对不同细胞进行了有效分类。本实验采用 Matlab 中自带的 11 个基于二值图像的特征数值:面积(基于像素点的计算)、计算圆度(长轴

长度、短轴长度、偏心率)、方向、填充区域面积、凸边面积、欧拉数、等效直径、固靠性、伸长度^[4,6,7]

1.4 RBF 神经网络的设计

在欠发达国家和地区, ANN 在计算机辅助诊断 (CAD) 领域的应用有着特别重要的意义, 因为对于从病人身上获取的第一手资料, 这是低成本的解决方案^[2]. 本实验选择 RBF 神经网络进行特征分类. RBF 是有监督的神经网络, 它具备一些明显的特点. 它具有三层结构, 包括输入层、隐含层 (包含径向基函数神经元)、输出层. 输出层给出我们需要分类的结果 (通常为 0-1 的二值结构). RBF 网络学习和分类的能力优于 Bp 神经网络方法. BP 神经网络是生物医学领域广泛使用的 ANN 方法, 但是该方法训练时间慢, 对于复杂数据处理能力差, RBF 训练速度快, 对未训练过的类别可以判读为: “未知”, 因为处理数据时会经常碰到从未处理过的数据, 所以这个特点有利于结果判读. 训练速度快很重要是因为当处理增加的分类时, 需要不断重复训练^[7,8]. 而且 RBF 对抗干扰的能力强于 BP^[9]. 但 RBF 无法对网络训练的设置进行预先调整^[11].

径向基神经网络本质是选择具备如下形式的函数:

$$f(x) = \sum_{i=1}^n w_i \varphi(\|x - x_i\|)$$

径向基神经网络的传递函数是高斯函数, 高斯

函数就是具备上述形式的径向基函数之一. 在图 2 所示 RBF 结构图中, 隐层神经元就是径向基神经元. 设计 RBF 的挑战在于正确设置隐含层的神经元数目, 调整散布常数 (spread) 的值, 使每个输入神经元的响应区域交叉减少^[11,12]. 输入变量很多时将会使网络设计难度增加, 若应用 newrb() 函数则能使设计更有效^[13].

2 结果

结果的判读和阈值的选择相关, 若将阈值选为 -1.3E-03, 见表 1. 经过 McNemar 检验, $P = 0.001$ (双侧), 在 $\alpha = 0.01$ 水平上, 二者有显著性差异. 虽然本实验方法和金标准有差距, 但仍不失为对金标准的有益补充.

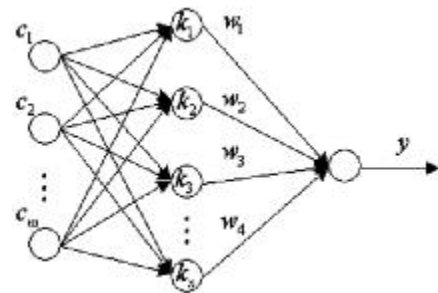


图 2 RBF 神经网络的拓扑结构
Fig. 2 The topological structure of RBF nerve net

表 1 阈值选为 -1.3E-03 时两种方法对疾病的诊断情况 (n)

Tab. 1 Comparison of disease diagnosis between two methods when the threshold value was -1.3E-03 (n)

项 目	病理医师专家组诊断为阳性	病理医师专家组诊断为阴性	合计
RBF 诊断为阳性	29	26	55
RBF 诊断为阴性性	7	5	12
合计	36	31	67

将 RBF 网络的输出结果进行分析, 通过了解其分布的特点来决定如何选取分类的阈值. 再分别计算不同阈值下 RBF 神经网络诊断实验的敏感度和特异度, RBF 输出值的判读可以从不同的阈值得出不同的结果, 见表 2.

阳性似然比为 $S_n/(1-S_p)$, 阴性似然比为 $(1-S_n)/S_p$, 从阳性似然比、阴性似然比来看, 实验中 RBF 区分阳性结果和阴性结果的能力无法和金标准相比, 但是分别选择高敏感性和高特异性的阈值却能在疾病的诊断过程中起到有益的辅助作用.

表 2 不同分类阈值的敏感度和特异度

Tab. 2 The sensitivity and specificity of different threshold values

RBF 输出值分类阈值	$S_n(\%)$	$S_p(\%)$	$1-S_n$	$1-S_p$
-1.8E-02	91.67	6.45	0.082 4	0.935 5
-1.3E-03	80.56	16.13	0.194 4	0.838 7
-5.2868E-04	72.22	16.13	0.277 8	0.838 7
8.9E-03	36	70.97	0.64	0.290 3
7.308E-01	8.33	90.32	0.916 7	0.096 8

3 讨论

神经网络在分析大量的非线性数据时具有自适应的特点,具有很好的鲁棒性(指控制系统存在不确定因素的情况下,维持某些性能的特性),但是神经网络的计算是通过黑盒环闭控制数据分类的,在实际训练中带有偶然性,RBF中散布常数的确定具有一定的难度.神经网络需要大样本的训练才能逐步改善^[1].RBF的实验结果应用于临床诊断虽然有局限性,但其优越性可以体现为:若不断地进行机器学习,网络将逐渐优化,有利结果的判读;而且判断结果的过程比较客观;可以根据临床使用的不同目的来选择不同的分类阈值.高敏感性的指标可以辅助临床病理医师在口腔鳞状细胞癌的早期进行筛查,如果得到阴性结果将有助于疾病的排除,以本实验为例,如果敏感性为91.67%,则漏诊率仅为6.45%.对于病理医师来说,在显微镜下搜索癌细胞耗时耗力,RBF可以用于自动搜寻癌细胞程序的细胞辨识中,发挥机器学习和工作高效的优点,作为病理医师筛选细胞的有效工具,可以提高诊断效率,减少人力的耗费.如果采用高特异性的指标,由于假阳性率低,则有助于肯定诊断.综上所述,人工神经网络技术在实际应用中尚有很多不足需要改进和完善,但是人工智能的发展是确切无疑的,在医学图像诊断上的应用必将越来越精细,带来的改变必将是革命性的.

[参考文献]

- [1] BRICKLEY M R, COWPE J G, AND SHEPHERD J P. Performance of a computer simulated neural network trained to categorise normal, premalignant and malignant oral smears[J]. *J Oral Pathol Med*, 1996, 25(8):424 - 428.
- [2] MUKHERJEE A, PAUL R R, CHAUDHURI K, et al. Performance analysis of different wavelet feature vectors in quantification of oral precancerous condition [J]. *Oral Oncol*, 2006, 42(9):914 - 928.
- [3] PAUL R R, MUKHERJEE A, DUTTA P K, et al. A novel wavelet neural network based pathological stage detection technique for an oral precancerous condition [J]. *J Clin Pathol*, 2005, 58(9):932 - 928.
- [4] 张汗灵. MATLAB在图像处理中的应用[M]. 北京:清华大学出版社, 2008:334.
- [5] NIPON T U. White blood cell segmentation and classification in microscopic bone marrow images[J]. *Lecture Notes in Computer Science*, 2005, 3 614:787 - 796.
- [6] 张洪刚, 陈光, 郭军. 图像处理与识别[M]. 北京:北京邮电大学出版社, 2006:125 - 127.
- [7] WILKINS M F, BODDY L, MORRIS C W, et al. Identification of phytoplankton from flow cytometry data by using radial basis function neural networks [J]. *Appl Environ Microbiol*, 1999, 65(10):4 404 - 4 410.
- [8] WILKINS M F, LBODDY, CWMORRIS, et al. A comparison of some neural and non-neural methods for identification of phytoplankton from flow cytometry data [J]. *CABIOS*, 1996, 12:9 - 18.
- [9] VAN STAVEREN H J, VAN VEEN R L, SPEELMAN O C, et al. Selecting effective siRNA sequences by using radial basis function network and decision tree learning[J]. *BMC Bioinformatics*, 2006, 18;7(Suppl 5):S22.
- [10] POGGIO, TAND GIROSI F. Networks for approximation and learning [J]. *Proceedings of the IEEE*, 1990, 78(9):1 481 - 1 497.
- [11] 董长虹. Matlab神经网络与应用[M]. 北京:国防工业出版社, 2007:124.
- [12] AFKHAMI A, ABBASI-TARIGHAT M. Comparative determination of phosphate and silicate using molybdenum blue by radial basis function and feed-forward neural networks assisted by principal component analysis [J]. *Anal Sci*, 2008, 24(6):779 - 783.
- [13] 飞思科技产品研发中心. 神经网络理论与Matlab7实现[M]. 北京:电子工业出版社, 2005:79 - 80.
- [14] VAN STAVEREN H J, VAN VEEN R L, SPEELMAN O C, et al. Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study[J]. *Oral Oncol*, 2000, 36(3):286 - 293.

(2011 - 11 - 05 收稿)