DOI:10.11817/j.issn.1672-7347.2019.03.003
http://xbyxb.csu.edu.cn/xbwk/fileup/PDF/201903244.pdf

不同类型小波滤波对影像组学特征相关性和诊断效能的影响

程梓轩[1,2]，黄燕琪[2]，黄晓媚[2]，吴小媚[2]，梁长虹[1,2]，刘再毅[1,2]

(1. 华南理工大学医学院，广州 510006；2. 广东省人民医院放射科，广州 510080)

[摘要] **目的：**探讨不同小波滤波对影像组学特征相关性和诊断效能的影响。**方法：**回顾性收集143例结直肠癌患者(淋巴结转移阳性64例，阴性79例)的术前CT图像，经放射科医师勾画肿瘤区域后，使用Matlab编写的软件提取不同类型小波的影像组学特征。通过计算相关系数分析不同小波间同名特征的相关性。采用最小绝对收缩和选择算子(the least absolute shrinkage and selection operator，LASSO)构建不同的小波特征集预测淋巴结转移的影像组学标签并采用Delong's检验比较其效能。**结果：**随着小波阶数差异的增大，小波间高相关同名特征数量减少。部分特征在不同小波间易出现高相关性。单个小波的特征集中rbio2.2，sym7和db7的特征子集构建的影像组学标签诊断效能最高。Daubechies系列小波特征集构建的标签预测淋巴结转移效能最高，Biorthogonal系列小波标签则最低，在去除同名高相关特征后全体特征集的诊断效能显著提高(P=0.004)。**结论：**建议选择阶数差异大的小波以降低影像组学特征的数据冗余度。为提高标签的诊断效能，有必要去除高相关特征。

[关键词] 影像组学；小波特征；特征冗余；诊断效能

Effects of different wavelet filters on correlation and diagnostic performance of radiomics features

CHENG Zixuan[1,2], HUANG Yanqi[2], HUANG Xiaomei[2], WU Xiaomei[2], LIANG Changhong[1,2], LIU Zaiyi[1,2]

(1. School of Medicine, South China University of Technology, Guangzhou 510006;
2. Department of Radiology, Guangdong Provincial People's Hospital, Guangzhou 510080, China)

ABSTRACT

Objective: To investigate the effects of different wavelet filters on correlation and diagnostic performance of radiomics features.

Methods: A total of 143 colorectal cancer (CRC) patients (64 positive in lymph node metastasis and 79 negative) with contrast-enhanced CT examination were recruited. After labeling the tumor area by experienced radiologists, radiomics wavelets features based on 48 different wavelets were extracted using in-house software coded by Matlab. The correlation coefficients of the features with

收稿日期(Date of reception)：2018-08-10

第一作者(First author)：程梓轩，Email: czxtlbb@126.com, ORCID: 0000-0002-7221-4345

通信作者(Corresponding author)：刘再毅，Email: zyliu@163.com, ORCID: 0000-0003-3296-9759

基金项目(Foundation item)：国家重点研发计划(2017YFC1309100)；国家自然科学基金(81771912)；广东省省级科技计划项目(2017B020227012)。This work was supported by the National Key Research and Development Program (2017YFC1309100), the National Natural Science Foundation (81771912), and the Science and Technology Planning Project of Guangdong Province (2017B020227012), China.

same names between different wavelets were calculated and got the distribution of high-correlation features between each wavelet. The least absolute shrinkage and selection operator (LASSO) was used to build signatures between lymph node metastasis and wavelet features data set based on different wavelets. The numbers of features in signatures and diagnostic performance were compared using Delong's test.

Results: With the difference of wavelet order increased, the number of high-correlation features between two wavelets decreased. Some features were prone to high correlation between different wavelets. When building radiomics signature based on single wavelet, signatures built from 'rbio2.2', 'sym7' and 'db7' did well in predicting lymph node metastasis. The signature based on Daubechies wavelet feature set had the highest performance in predicting lymph node metastasis, while the signature from Biorthogonal wavelet features was worst. Improvement was significant in diagnostic performance after excluding the high-correlation features in the whole features set (P=0.004).

Conclusion: In order to reduce the data redundancy of features, it is recommended to select wavelets with large differences in wavelet orders when calculating radiomics wavelet features. It is necessary to remove high correlation features for improving the diagnostic performance of radiomics signature.

KEY WORDS radiomics; wavelet feature; feature redundancy; diagnostic performance

近年来兴起的影像组学，通过从医学影像中提取高通量影像组学特征，构建预测模型，可用以解析临床信息，评估疗效和预测预后，辅助临床决策，有重要的临床转化价值[1]。影像组学的本质是通过图像信息处理技术，将医学图像转化为可挖掘数据，反映肿瘤异质性[2]。早期研究提取的影像组学特征少，仅有数百个[3]，其对临床信息的解析可能不全面，因此有必要提取更多的特征。研究者发展了多种图像处理方法，使影像组学特征数量大幅增加，目前可达成上千个特征[4]。在增加的特征中，小波特征占据了很大的比例；通过使用不同类型小波对图像进行滤波，可产生不同的影像组学特征。然而，随着小波数量增加，影像组学特征总数急剧增加，给数据分析或模型构建带来巨大挑战，比如海量特征之间可能存在高度线性相关，由此产生的特征冗余会导致构建的线性模型过拟合[5]。不同研究者使用的小波类型不尽相同，有的仅使用一个小波[6]，也有的使用多个小波[7]。目前对于小波类型的选取没有统一的标准。选择合适类型和数量的小波进行图像滤波处理，既可避免特征数量过多，也可全面反映肿瘤的异质性，还能保证影像组学预测模型的诊断效能，因此有重要的临床意义。

本研究将以结直肠癌(colorectal cancer，CRC)CT数据集评价淋巴结转移为例，探讨5大类共48种不同的小波滤波后计算的影像组学特征之间的线性相关性，以及其标签对诊断淋巴结转移效能的影响，从而指导选择合适的小波类型进行影像组学特征提取。

1 对象与方法

1.1 对象

回顾性收集2011年1月至2014年9月在广东省人民医院经手术证实的CRC患者的临床资料和术前CT图像。纳入标准：接受CRC手术并进行淋巴结清扫的患者，有明确病理N分期[8]。排除标准为：1)曾接受术前治疗；2)病理N分期不明确；3)CT图像质量差导致肿瘤无法识别。共纳入143例患者，其中男86例，女57例；年龄29~89(62.2±11.23)岁；淋巴结转移阳性64例，阴性79例。

1.2 CT图像采集

采用GE LightSpeed Ultra 8排螺旋CT或LightSpeed VCT 64排螺旋CT对患者行腹部螺旋CT增强扫描，序列扫描参数为：管电压120 kV，管电流160 mA，视野350 mm×350 mm，层厚1 mm或1.25 mm。予静脉注射90~100 mL碘化造影剂，60 s后行静脉期延迟扫描。扫描获得的CT图像上传至影像归档和通信系统(picture archiving and communication systems，PACS)中，导出DICOM格式图像。

1.3 肿瘤感兴趣区勾画

选取每位患者术前静脉期CT图像。每位患者的图像由1名具有5年相关工作经验的放射科医师勾画整个肿瘤区域作为感兴趣区域，勾画肿瘤区域使用的软件为ITK-SNAP。

1.4 影像组学特征提取

使用课题组开发的基于Matlab 2016b(Mathswork公司，美国)编写的软件进行影像组学特征提取。首先，使用Matlab库中的5个系列共48个小波的小波基对图像进行滤波处理，包括Daubechies小波8个(db，1~8)，Symlets小波7个(sym，2~8)，Coiflets小波5个(coif，1~5)，Biorthogonal小波14个(bior，1.3~6.8)和Reverse Bior小波14个(rbio，1.3~6.8)(小波后的数字称为小波的阶数；带小数点的阶数并非指小数，是指重构和分解滤波器各自的阶数)。然后进行特征提取[9-10]。其中，使用小波基对图像进行滤波处理的方法依据以下公式[11]：

$$X'(i,j,k)=\sum_{p}^{N_x}\sum_{q}^{N_y}\sum_{r}^{N_z}H_x(p)H_y(q)H_z(r)\cdot X(i+p,j+q,k+r)。$$

其中，X为原始图像，X'为滤波后图像，H_x，H_y，H_z分别为x，y，z 3个方向上的分解滤波器。根据3个方向上滤波器可能为低通(L)或高通(H)滤波器，按x，y，z的顺序排列，共有LLL，LLH，LHL，LHH，HLL，HLH，HHL，HHH 8种不同的滤波方式。每个滤波方式各包含14个一阶统计量和63个纹理特征，因此每个小波特征子集包含616[8×(14+63)=616]个特征，每例患者的CT图像提取影像组学特征共29 568(616×48=29 568)个。影像组学特征按照“小波名+滤波方式+特征名”进行命名。

1.5 影像组学特征相关性分析

提取的影像组学特征间可能存在高度线性相关的冗余特征，因此，笔者对所有小波特征子集进行两两比较。具体方法是：1)根据不同特征数据的正态性采用Pearson或Spearman相关分析[12]，计算每两个特征子集中的同名特征(滤波方式+特征名)间的相关系数r；2)以r的绝对值大于0.9的标准定义高相关特征。在计算同名特征相关性时统计以下信息：1)每两个小波间同名特征对的数量，每两个小波之间最多可能有616个同名特征对；2)记录该特征被判断为高相关特征的总次数。一个特征子集中的某个特征需要与其他子集中的特征比较1 128(48×47÷2=1 128)次，8种滤波方式总计9 024次。

1.6 不同小波影像组学特征诊断效能分析

为检验每个小波的影像组学特征的诊断效能，本研究以基于影像组学术前预测CRC淋巴结转移为例，对每个小波特征子集单独建立影像组学标签，评价不同小波影像组学标签术前预测CRC淋巴结转移状态的诊断效能。

影像组学标签构建流程如下：用最小绝对收缩和选择算子(the least absolute shrinkage and selection operator，LASSO)-logistic回归方法，采取10折交叉验证方式选择适当的惩罚因子λ，筛选出关键的影像组学特征并构建影像组学标签[13]。使用受试者工作特征(receiver operating characteristic，ROC)曲线下面积(area under curve，AUC)评价每个小波特征子集的最佳标签的诊断效能。在控制影像组学标签特征数量不超过14的条件下，每个特征子集选择诊断效能最高的影像组学标签[14]，并按诊断效能由高到低进行排序。

将相同系列的小波的所有特征组合成5个小波特征集。同时，将每个系列单个小波诊断效能最高的一个进行额外组合，另外将所有特征组成一个全体特征集。这样得到的7个特征集均进行高相关同名特征去除，并于去除前后以相同的方法进行标签构建，记录去除前后的特征数量、诊断效能以及使用Delong's检验评价去除前后ROC曲线的变化差异[15]。以上流程如图1所示。采用R3.3.1软件行统计学处理；P<0.05为差异有统计学意义。

2 结 果

2.1 患者基线资料

79例淋巴结转移阴性CRC患者中男44名，女35名，年龄为29~89(61.57±11.74)岁。64例淋巴结转移阳性患者中男42名，女22名，年龄44~84(63.09±10.61)岁。淋巴结转移阴性组和阳性组患者的性别(χ^2检验，P=0.2279)与年龄(t检验，P=0.4168)差异无统计学意义。

2.2 小波之间特征相关性分析

两两比较所有小波中同名特征的相关性，小波间高相关特征的数量以颜色形式显示在图2中，从图中可以看出：1)sym2与db2，sym3与db3这两对小波对应方块颜色为最深的红色(蓝色箭头)；2)在同一系列的小波中，如db1与其他db系列小波对比(绿色箭头)，随着阶数差异的增大，方块颜色变浅，即高相关同名特征数量减少；3)即使不同系列的小波，阶数相近的小波其高相关同名特征的数量也会更多，如db7与sym系列的小波比较时(黄色箭头)，与sym7对应的方块颜色最深，颜色随阶数差异的增大而渐浅；4)比较特别的是bior系列和rbio系列，无论是系列内的比较，还是与其他系列的比较，均无明显规律。

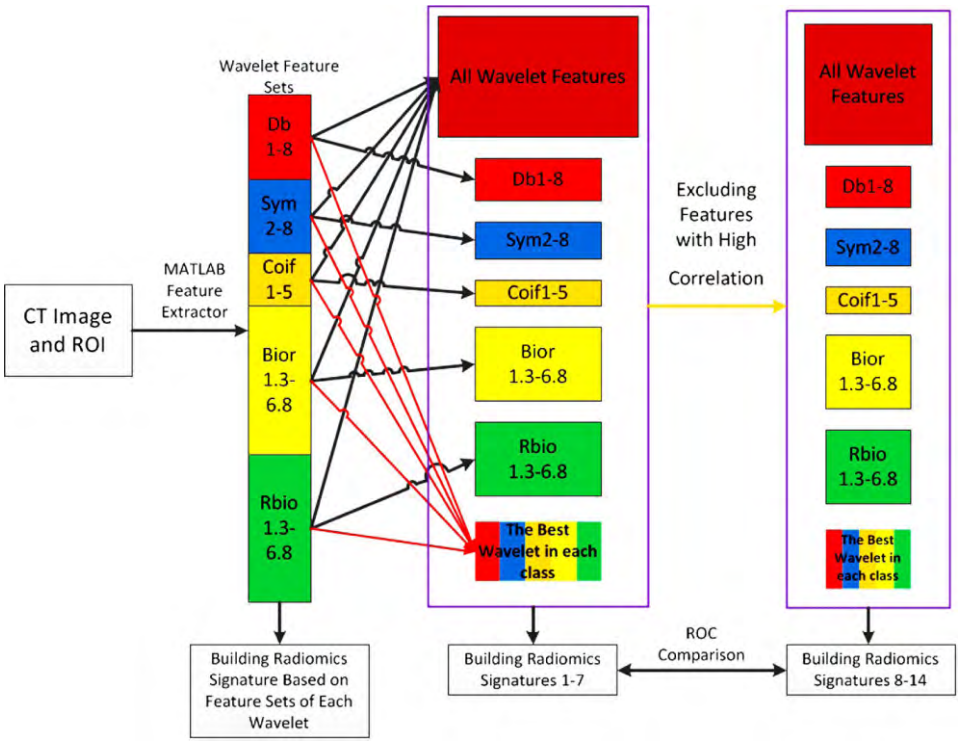

图1　不同小波的影像组学特征诊断效能分析流程图

Figure 1　Flow chart of analysis about diagnosis performance of different wavelet features

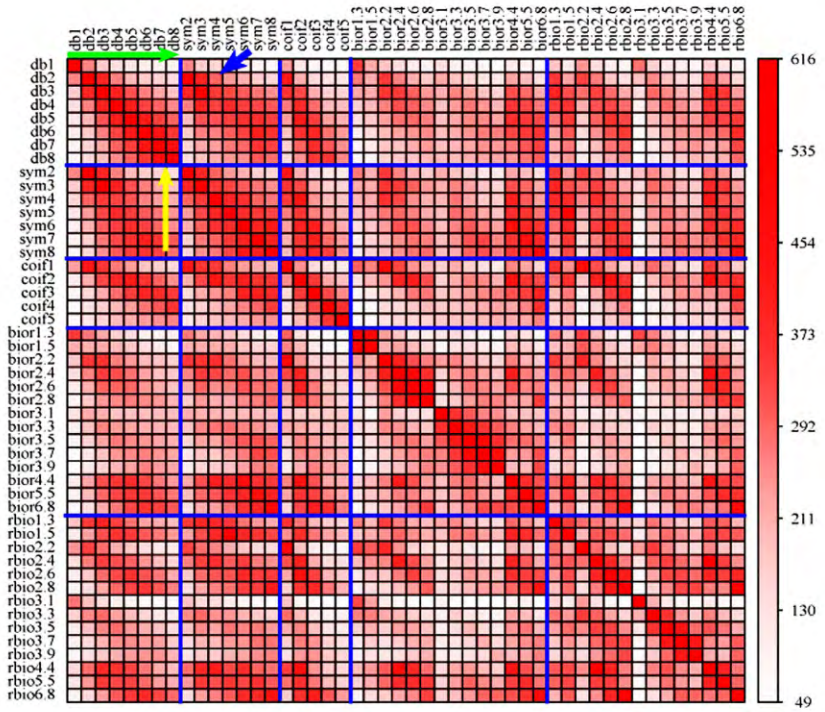

图2　小波间高相关同名特征对的数量统计彩图

Figure 2　Color map of feature pairs with high correlations between wavelets

The axis represents 48 wavelets in this study. Each block represents the number of high-correlation same feature pairs between the two wavelets by color, as the tab shows. The blue line separates the five classes of wavelets. The figure is symmetric about the "upper left-lower right" diagonal

2.3 不同特征相关性分析

通过统计每个特征被判断为高相关特征的总次数，发现部分同名特征在不同小波间易出现高相关性，而部分同名特征在不同小波间难以表现出相关性。其中，GLRLM_GLN，GLRLM_RLN，GLDM_GLN，GLDM_DM和NGTDM_Coarness这5个特征，任何小波在各个滤波方向的相对应特征，均与另一个小波的相对应特征有高相关性，统计次数总数达到最大值9 024次，频率达到100%。

2.4 各小波特征集诊断效能分析

通过每个小波特征子集建立诊断患者淋巴结转移的影像组学标签。由于限制了标签特征数量，所有标签均包含14个特征。包含不同小波子集的特征集对应的最佳影像组学标签AUC经排序后如图3所示。

经排序后发现，rbio2.2，sym7和db7特征子集构建的标签诊断效能最高，而最差的标签是sym4子集构建的标签。其他研究常用的sym8小波在48个小波中排在第12位，而db1和coif1则分别排在第20和23位。同一系列的小波中既有诊断效能较高的，也有诊断效能较低的。

将每个系列单个小波诊断效能最高的一个(rbio2.2，sym7，db7，coif2，bior5.5)进行额外组合成一个特征集(Best 5)，加上每个系列小波的特征子集和全体特征集。7个特征集进行高相关特征去除前后的特征数量和诊断效能的变化差异见表1。

从表1可以看出，当没有进行高相关特征去除前，Daubechies系列小波特征集构建的标签诊断效能最高，Biorthogonal系列小波特征集构建的标签诊断效能最低。各系列诊断效能最佳的5个小波组合后构建的标签诊断效能并不突出。在对各特征集进行同名特征高相关去除后，所有小波特征集构建的标签诊断效能有显著提升；而Biorthogonal所有小波构建的标签诊断效能反而明显下降，且显著低于全体特征集构建的标签。原本表现最好的Daubechies小波特征集构建的标签诊断效能也有下降，但不明显。其他系列小波特征集构建的标签则没有显著的变化。

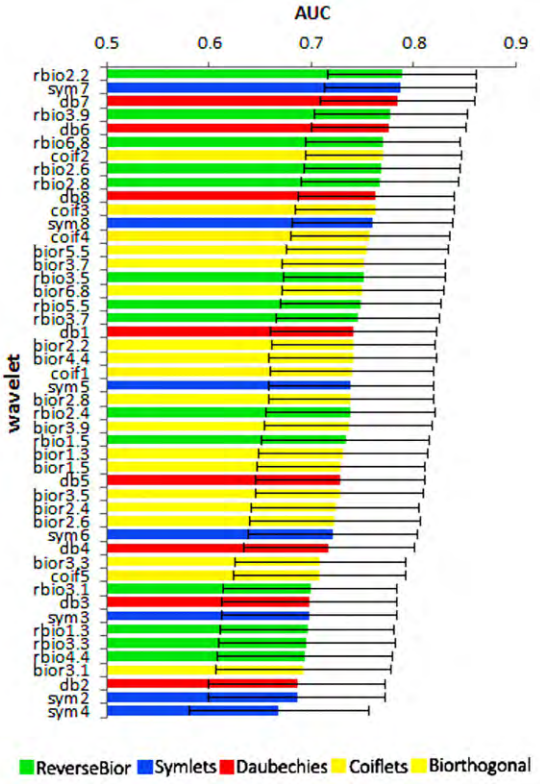

图3 不同小波构建的影像组学标签诊断效能

Figure 3 Diagnosis performances of radiomics signatures based on different wavelets

The diagnostic performances of the radiomics signatures based on each wavelet are sorted by AUC. Different colors represent different classes of wavelets. The black line segment represents the 95% confidence interval for the corresponding AUC

表1　不同类型小波特征标签特征数量和诊断效能的对比

Table 1　Comparison of feature numbers and diagnosis performance between signatures based on different wavelet features

Wavelet	All features included				Features with high correlation excluded				*P*
	Label	Feature number	AUC	*P*(vs Label 1)	Label	Feature number	AUC	*P*(vs Label 8)	
All	1	29 568	0.808		8	10 006	0.817		0.0040
Daubechies	2	4 928	0.812	0.8849	9	2 671	0.795	0.4607	0.1081
Symlets	3	4 312	0.785	0.2746	10	2 217	0.794	0.2771	0.3796
Coiflets	4	3 080	0.799	0.7773	11	2 008	0.799	0.5492	0.4795
Biorthogonal	5	8 624	0.779	0.3831	12	3 786	0.740	0.0158	0.0100
ReverseBior	6	8 624	0.796	0.7065	13	3 875	0.801	0.6090	0.6442
Best 5	7	3 080	0.796	0.4134	14	1 801	0.796	0.1316	1.0000

Best 5: The fusion of wavelets whose signatures perform best in predicting lymph node metastasis respectively from 5 wavelet classes; *P* value is calculated with Delong's test between AUCs

3 讨　论

因为sym2与db2，sym3与db3这2对小波滤波器长度和数值完全一样，因此算出来的特征也相同。在后续的分析中这两对小波可以互相代替，往后的研究中也不应重复选择。

阶数越接近的小波，出现高相关特征对的次数越多，冗余特征的数量越多。因此，选择小波的时候尽可能选择阶数差异大的小波，例如选择了db2小波，那db1和db3，db4应当不再考虑。

小波Biorthogonal系列和ReverseBior系列比较特别。由于这两个系列小波阶数由两部分构成，没有明确的高低差异，因此在高相关特征数量分布上没有明显规律，筛选该两类小波时需要参考图2中高相关特征数量在这类小波之间的分布情况。

一些特征由于自身特性，在不同的小波或滤波方式之间表现出高相关性，这些特征应当少作计算。但目前并未有研究指出明确的界限能以特征被认定为高相关的频率对特征进行筛选。本次研究找到5个特征，在不同的小波或滤波方式处理后的图像上，计算出来的结果之间均存在相关性。这些特征在今后的分析中可以考虑不算或只算一次。

表3直观地反映各小波特征之间诊断效能的差异。Symlet的高阶小波sym7和sym8构建的标签诊断效能比较好。研究[16]指出：因为symlet是一系列正交且紧凑支持的小波，有助于局部保持图像的空间特征，因此诊断效能相对更高。值得注意的是常用的低阶小波db1(即haar小波)和coif1表现也一般。

当小波按系列分别组合成多小波特征集构建标签时，Daubechies系列小波比所有特征组合起来构建的标签诊断效能更好。这说明特征数量并不是越多越好，过多特征反而对标签的构建造成了干扰。

在进行高相关同名特征去除后，全体小波特征集显著地提高了诊断效能，说明在特征数量很大的时候进行高相关特征去除，降低特征集冗余是很有必要的。混合了不同系列小波的特征集在去除高相关特征后构建的标签完全没有变化。说明不同系列小波特征融合后受高相关特征的影响减小，选择小波时应注意系列的多元化。

无论是否去除高相关特征，Biorthogonal系列小波特征构建的标签诊断效能都是最低。考虑其单个小波诊断效能也不突出，这可能与该系列小波缺乏对称性有关，选择小波时建议少选择这个系列的小波[17]。

总的来说，在选择小波时，既要尽可能地选入诊断效能好的小波，如多选择Daubechies系列的小波，少选择Biorthogonal系列的小波；同时也要注意小波间高相关特征的数量，这方面可以从图2中得到一定的参考信息，尽可能选择小波间高相关特征数量较少的小波以降低特征之间的冗余性。

本研究探究了影像组学小波特征间的线性相关性情况，给影像组学研究中小波特征的选择提供了一定的指导。但本文仍有一定的局限性。本研究仅研究48种小波基所计算的特征，结论能否推广到其他小波乃至其他滤波方法尚不明确[18]。另外，本文以预测CRC淋巴结转移为例构建标签目的是寻找诊断效能好的小波，由于缺乏独立验证，这些标签不可作为临床应用，而且对于其他疾病或指标这些结论是否成立，还需进一步研究。

利益冲突声明：作者声称无任何利益冲突。

参考文献

[1]　Lambin P, Riosvelazquez E, Leijenaar R, et al. Radiomics: Extracting

more information from medical images using advanced feature analysis[J]. Eur J Cancer, 2007, 48(4): 441-446.

[2] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data[J]. Radiology, 2016, 278(2): 563-577.

[3] Ma Z, Fang M, Huang Y, et al. CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma[J]. Eur J Radiol, 2017, 91: 142-147.

[4] Zhang S, Song G, Zang Y, et al. Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery[J]. Eur Radiol, 2018, 28(9): 3692-3701.

[5] Cong Y, Liu J, Fan B, et al. Online similarity learning for big data with overfitting[J]. IEEE Trans Big Data, 2017:1-1.

[6] Vallières M, Freeman CR, Skamene SR, et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities[J]. Phys Med Biol, 2015, 60(14): 5471-5496.

[7] Liang C, Cheng Z, Huang Y, et al. An MRI-based radiomics classifier for preoperative prediction of Ki-67 status in breast cancer[J]. Acad Radiol, 2018, 25(9): 1111-1117.

[8] Chang GJ, Rodriguez-Bigas MA, Skibber JM, et al. Lymph node evaluation and survival after curative resection of colon cancer: Systematic review[J]. J Natl Cancer Inst, 2007, 99(6): 433-441.

[9] Huang X, Cheng Z, Huang Y, et al. CT-based radiomics signature to discriminate high-grade from low-grade colorectal adenocarcinoma[J]. Acad Radiol, 2018, 25(10): 1285-1297.

[10] Zwanenburg A, Leger S, Vallières M, et al. Image biomarker standardisation initiative-feature definitions[J]. arXiv: 1612.07003.

[11] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J]. Nat Commun, 2014, 5(1): 4006.

[12] Mukaka MM. A guide to appropriate use of Correlation coefficient in medical research.[J]. Malawi Med J, 2012, 24(3): 69-71.

[13] Osborne MR, Presnell B, Turlach BA. On the LASSO and Its Dual[J]. J Comput Graph Stat, 2000, 9(2): 319-337.

[14] Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis[J]. J Clin Epidemiol, 1996, 49(12): 1373-1379.

[15] Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach[J]. Biometrics, 1988, 44(3): 837-845.

[16] Vallières M, Freeman CR, Skamene SR, et al. FDG-PET/MR imaging for prediction of lung metastases in soft-tissue sarcomas of the extremities by texture analysis and wavelet image fusion[D]. Montreal: McGill University, 2013.

[17] Sweldens W. The lifting scheme: A custom-design construction of biorthogonal wavelets[J]. Appl Comput Harmon A, 1996, 3(2): 186-200.

[18] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer[J]. J Clin Oncol, 2016, 34(18): 2157-2164.

(本文编辑 郭征)

本文引用： 程梓轩, 黄燕琪, 黄晓媚, 吴小媚, 梁长虹, 刘再毅. 不同类型小波滤波对影像组学特征相关性和诊断效能的影响[J]. 中南大学学报(医学版), 2019, 44(3): 244-250. DOI:10.11817/j.issn.1672-7347.2019.03.003

Cite this article as: CHENG Zixuan, HUANG Yanqi, HUANG Xiaomei, WU Xiaomei, LIANG Changhong, LIU Zaiyi. Effects of different wavelet filters on correlation and diagnostic performance of radiomics features[J]. Journal of Central South University. Medical Science, 2019, 44(3): 244-250. DOI:10.11817/j.issn.1672-7347.2019.03.003