

基于 U-Net 和随机森林的直肠癌淋巴结转移的智能诊断

摘要

直肠癌是消化道最常见的恶性肿瘤之一，易向肠外浸润并发生淋巴结及远处转移，直肠癌肿瘤本身的特性和周围淋巴结转移存在一定的关联性，本文建立了基于病人的直肠癌动脉期 CT 影像评估其肿瘤是否发生了淋巴结转移的模型。

首先，这是一个典型的影像组学问题，可分解为三个串联的子任务：直肠肿瘤分割、直肠肿瘤影像组学特征的提取、肿瘤影像特征与淋巴结转移的相关性验证。

针对任务一（肿瘤分割），采用 U-Net 训练肿瘤分割网络，Dice 系数的相反数作为 Loss 函数，为最大程度提升肿瘤分割模型的性能，在训练网络前，对数据集中 CT 影像做窗口调节、直方图均衡化、数据增强等必要的预处理操作，用训练好的网络进行预测时，对生成标记肿瘤区域的掩模图像做基于统计的多余连通域的过滤操作，避免一些明显错误的分割结果。最终训练得到的模型在验证集上的肿瘤分割的 Dice 系数为 0.8956。

针对任务二（特征提取），对原始 CT 影像标注肿瘤区域做相关滤波操作，获得 Wavelet、LoG、Gradient 等 9 种生成图像类型，对每种类型图像提取器肿瘤强度、形状、纹理等特征，这一过程用专门用于医学影像提取的 python 第三方库 pyradiomics 做特征的自动提取，最终对每个病人的 CT 影像提取得到 1323 个影像组学特征。

针对任务三（建立分类器），采用包裹式特征选择的方法建立随机森林分类器，将病人的性别、年龄信息纳入原始特征集，用 RFE 方法递归地从原始特征集中移除最不重要的特征。为避免过拟合，在分类器性能不明显降低的前提下选择尽量少的特征数目，最终选取得到 20 个特征，交叉验证平均准确率为 0.7300，查全率为 0.6860，查准率为 0.7799，计算得到 $F-score = 0.6878$ ，验证了肿瘤影像特征与淋巴结转移的相关性。

最后，讨论当肿瘤分割结果并不准确时，是否仍然能构建有效的模型来判断肿瘤是否已发生淋巴结转移。基于任务一中预测得到的肿瘤掩模仍按任务二、三中的流程提取特征、选择特征并训练随机森林分类器。通过十折交叉验证，平均准确率为 0.7275，查全率为 0.7003，查准率为 0.6726，计算得到 $F-score = 0.6523$ ，说明用任务一中分割不完全准确的肿瘤掩模仍能构建有一定效果的淋巴结转移分类模型。

关键词： U-Net pyradiomics 特征提取 包裹式特征选择 随机森林

Abstract

Rectal cancer is one of the most common malignant tumors in the digestive tract. It is easy to infiltrate into the intestine and develop lymph node and distant metastasis. The characteristics of rectal cancer itself and the lymph node metastasis of the rectum are related. This paper establishes patient-based rectal cancer. An arterial phase CT image is used to assess whether a tumor has a lymph node metastasis model.

First, this is a typical imaging omics problem that can be broken down into three sub-tasks: rectal tumor segmentation, extraction of imaging features of rectal tumors, and correlation between tumor imaging features and lymph node metastasis.

For task 1 (tumor segmentation), U-Net is used to train the tumor segmentation network. The inverse of the Dice coefficient is used as the Loss function to maximize the performance of the tumor segmentation model. Before training the network, the window image of the dataset is adjusted. Histogram equalization, data enhancement and other necessary pre-processing operations. When using the trained network for prediction, the filtering operation of the redundant connected domain based on statistics is generated on the mask image of the generated tumor region to avoid some obviously erroneous segmentation results. . The Dice coefficient of the tumor segmentation on the validation set of the final trained model was 0.8956.

For task 2 (feature extraction), the original CT image is labeled with the tumor area for correlation filtering operation, and 9 types of generated image types such as Wavelet, LoG, and Gradient are obtained, and the characteristics of each type of image extractor such as tumor intensity, shape, texture, etc. A process uses the python third-party library pyradiomics, which is specially used for medical image extraction, to automatically extract features, and finally extracts 1,323 imaging ensemble features for each patient's CT image.

For task 3 (establishing the classifier), a random forest classifier is established by using the method of parcel feature selection. The gender and age information of the patient is included in the original feature set, and the least important feature is recursively removed from the original feature set by RFE method. . In order to avoid over-fitting, select the minimum number of features under the premise that the performance of the classifier is not significantly reduced, and finally select 20 features. The average accuracy of cross-validation is 0.7300, the recall rate is 0.6860, and the precision is 0.7799. Calculated, the correlation between tumor imaging features and lymph node metastasis was verified.

Finally, it is discussed whether an effective model can still be constructed to determine whether a tumor has metastasized or not when the tumor segmentation result is not accurate. The tumor mask based on the prediction in Task 1 still extracts features, selects features, and trains a random forest classifier according to the processes in Tasks 2 and 3. Through the ten-fold cross-validation, the average accuracy rate is 0.7275, the recall rate is 0.7003, and the precision is 0.6726. It is calculated that the segmentation of the lymph node metastasis with certain effect can still be constructed by using the tumor mask with incomplete segmentation in task 1. model.

Key words: U-Net; feature extraction with pyradimoics; wrapper feature selection; randomforest

目录

1 引言	4
2 问题分析	5
3 任务一：直肠肿瘤分割	6
3.1 网络结构	6
3.1.1 U-Net	6
3.1.2 Loss 函数	7
3.2 图像预处理与后处理	7
3.2.1 预处理	8
3.2.2 后处理	9
3.3 实验设置及结果分析	11
4 任务二：直肠肿瘤部分特征的提取	12
4.1 相关特征	12
4.2 特征提取	12
5 任务三：肿瘤影像特征与淋巴结转移的相关性验证	14
5.1 建立分类模型	14
5.1.1 数据平衡	14
5.1.2 训练分类器	14
5.1.3 评价指标	15
5.2 结果分析	15
6 肿瘤分割不精确的情况	17
7 总结与展望	17

1 引言

直肠癌是消化道最常见的恶性肿瘤之一。直肠癌易向肠外浸润并发生淋巴结及远处转移，对是否有淋巴结转移的准确判断是直肠癌治疗的重要步骤，但目前尚无一种方法能在术前准确地判断淋巴结转移情况。直肠癌肿瘤本身的特性和周围淋巴结转移存在一定的关联性，本问题的任务是基于病人的直肠癌 CT 影像评估其肿瘤是否发生了淋巴结转移，训练样本包含 107 个直肠癌患者的 CT 影像、已标记的直肠肿瘤掩模和患者年龄、性别及是否发生淋巴结转移等信息，任务分解如下：

- 直肠肿瘤分割
- 直肠肿瘤部分特征的提取
- 验证肿瘤影像特征与淋巴结转移的相关性

这是一个典型的影像组学问题，三个任务分别对应了影像组学的医学成像和分割、特征提取和特征选择、探索性分析与建模^[1]。但是，一般的影像组学问题在做出相关诊断时，往往不仅仅基于医学影像特征，更要综合病人的临床信息、治疗信息乃至基因信息，共同建模做出判断。因此试图仅通过直肠癌 CT 影像的相关特征做出病人是否淋巴结转移的评估，一方面给肿瘤分割和特征的提取提出了更高的精度要求，另一方面给特征选择和分类器的训练带来了极大的挑战，选择哪些特征作为判断因素对判断淋巴结转移模型的有效性有着直接的影响，针对这一难点与挑战，我们在传统影像组学分析框架的基础上，结合多种机器学习算法，对本问题做了优化，最终取得了一定的诊断效果，验证了直肠肿瘤影像特征与淋巴结是否转移的相关性。

2 问题分析

从 CT 影像出发，分割肿瘤、肿瘤特征提取、最后通过肿瘤特征判断是否发生了淋巴结转移，这是一个典型的影像组学问题，但在缺少其它辅助信息的情况下题目给我们提出了更高的精度要求。

对于任务一，分割出 CT 影像上直肠肿瘤所在的区域，生成对应的标记肿瘤区域的掩模图像。直肠肿瘤的影像学研究中，精确分割肿瘤区域一直是实践中的一个难点，传统的区域分割算法很难满足本任务的精度要求，因此考虑深度学习的方法，使用近年来在医学影像上分割效果极强的 U-Net 网络做图像分割，将 Dice 系数的相反数作为 Loss 函数。由于部分原始 CT 影像中肿瘤区域与其它区域粘连或像素值差异很小，即使人眼

也很难识别，在训练网络之前，需要对 CT 影像进行一定的预处理，如调节窗宽、窗位和直方图均衡化等操作。对于一些明显错误的肿瘤分割，需要做一定的后处理进行优化。

对于任务二，综合一个病人所有直肠 CT 影像及其标记出的直肠肿瘤掩模，提取其肿瘤特征，即用一系列数值特征代表该病人的直肠癌肿瘤状态，方便下一步基于肿瘤影像特征建立淋巴结是否转移的分类模型。肿瘤的影像组学特征一般包括肿瘤强度特征、形状特征、纹理特征、小波特征等，还可以对原始 CT 影像做一定的滤波操作后提取强度、形状、纹理特征。用专门用于医学影像提取的 python 第三方库 pyradiomics 做特征的自动提取。

对于任务三，验证肿瘤影像特征与淋巴结转移的相关性，即利用上一步提取的病人 CT 影像特征数据评估该病人是否发生了直肠癌淋巴结转移，这是一个二分类问题，除 CT 影像特征外的其他帮助判断的信息仅有病人的年龄和性别。查阅其他影像组学研究时发现，研究者往往首先基于完全准确的肿瘤分割影像提取影像组学特征，再用 Lasso 回归选择与 Label（如是否转移）最相关的特征，进而对所选特征和对应权重相乘后相加获得一维影像组学标签，最后结合其他临床信息，如病人年龄、性别、癌变位置、术前癌胚抗原(CEA)的表达水平、术前糖类抗原 19—9 (CA19—9)的表达水平等，建立随机森林或逻辑回归分类器，其中 CEA 和 CA19-9 的表达水平在部分研究中影响权重甚至占到 40%以上^[2]，而病人的年龄和性别对最终的诊断几乎没有影响。因此如果我们在仅有 CT 影像特征的情况下，仍将特征选择和训练分类器分开处理显然是不合理的，需要通过包裹式选择特征的方法训练分类器，充分挖掘 CT 影像特征与是否发生淋巴结转移的相关性。

另外，同时对任务一中 U-Net 肿瘤分割结果（不完全准确）做上述特征提取、选择及训练分类器，通过与标记的直肠肿瘤掩模（完全准确）比较最后的诊断效果，讨论肿瘤分割不够精确的前提下是否仍然能够构建有效的模型来判断肿瘤是否已发生淋巴结转移。

3 任务一：直肠肿瘤分割

3.1 网络结构

3.1.1 U-Net

U-net 是近年出现的用于图像分割的一种卷积神经网络，因其能够同时结合深层和浅层信息，精确地提取复杂特征，在医学影像上的分割取得了很好的效果。U-Net 的网络结构如下^[2]：

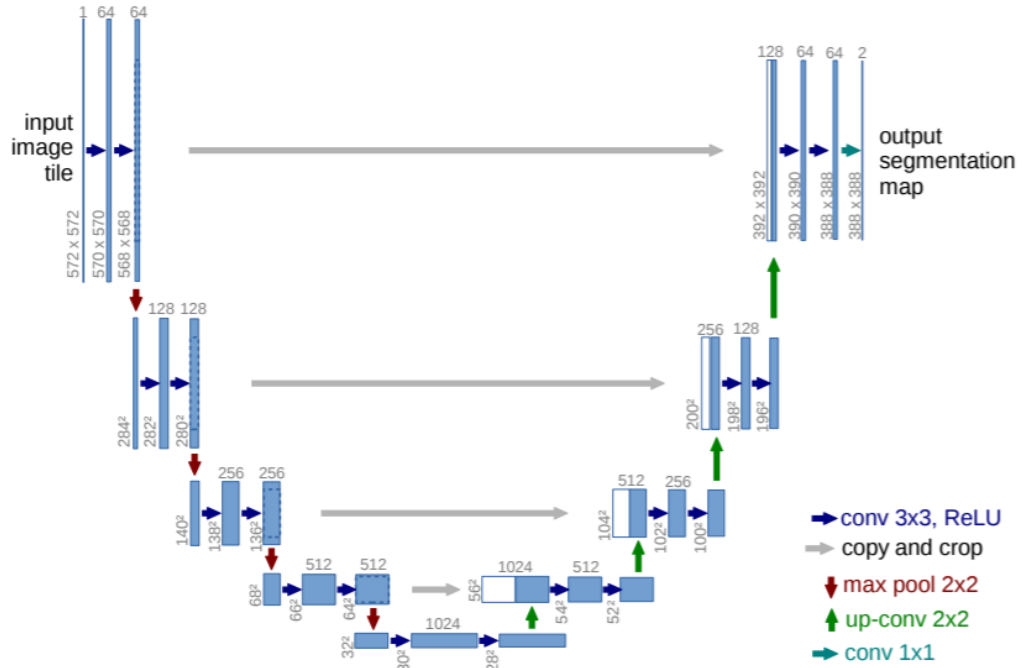


图 1 U-Net 网络结构

3.1.2 Loss 函数

图像分割采用 Dice 系数进行评价，它是一种集合相似度度量函数，通常用于计算两个样本的相似度：

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

其中 A 是医生勾画的直肠肿瘤区域， B 是 U-net 分割得到的直肠肿瘤区域。Dice 系数的取值范围是 $[0, 1]$ ，取值越接近 1 表明直肠肿瘤分割的结果与医生给出的结果越接近。

训练 U-net 网络的 Loss 函数定义为 Dice 系数的相反数：

$$Loss = -Dice$$

3.2 图像预处理与后处理

为最大程度提升肿瘤分割算法的精度，在训练肿瘤分割的网络前，需对图像进行

一定的预处理，在网络训练完成后，对 CT 影像上直肠肿瘤所在的区域进行预测时，需对生成标记肿瘤区域的掩模图像做后处理，避免一些明显错误的分割结果。

3.2.1 预处理

(1) 去除无肿瘤 CT 影像

在部分较浅和较深层的 CT 影像中没有检测到肿瘤，数据集中给出的肿瘤 mask 图像为全黑，如果将该部分 CT 影像进行训练，会导致 Dice 系数无法正确计算，进而导致 Loss 不收敛，因此需提前从训练集和验证集中去除无肿瘤 CT 影像。

(2) 窗口调节

通过改变 CT 影像的窗宽(Window Width)和窗位(Window Level)可以引起病灶密度的变化,显示出需要的图像信息^[3]。

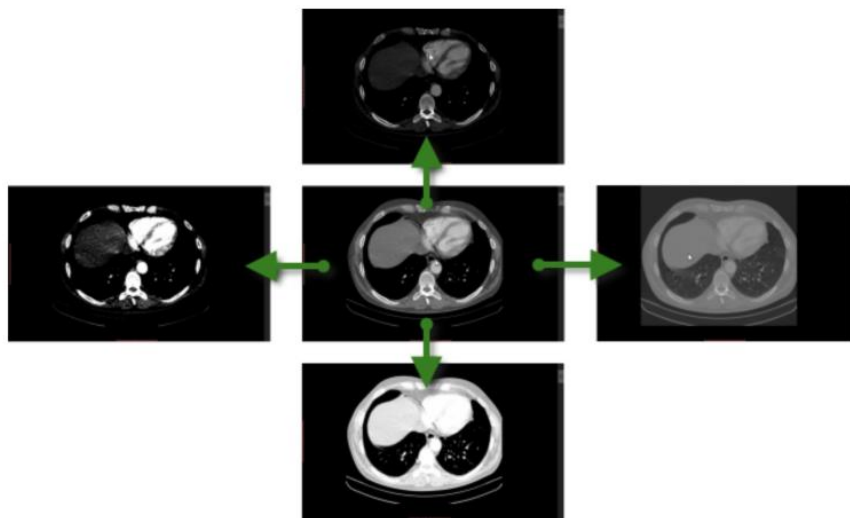


图 2 CT 影像的窗口调节

调节窗口能改变图像的亮度 and 对比度。窗宽是 CT 影像上显示的 CT 值范围，窗宽增大，则图像所示 CT 值范围加大，显示具有不同密度的组织结构增多，但各结构之间的灰度差别减少。反之小窗宽的显示图像则包含较少的 CT 值。窗位是窗的中心位置。窗宽、窗位用 CT 值表示：

$$\begin{aligned} \text{Window Width} &= CT_{\max} - CT_{\min} \\ \text{Window Level} &= (CT_{\max} + CT_{\min}) / 2 \end{aligned}$$

通过窗口的调节凸显直肠癌肿瘤，将大大降低肿瘤分割的难度，通过医学影像查

看软件__手动调节发现，当窗宽设置在 ，窗位设置在 ，在 CT 图像上看到的直肠癌肿瘤最为明显。调节效果如图三所示。

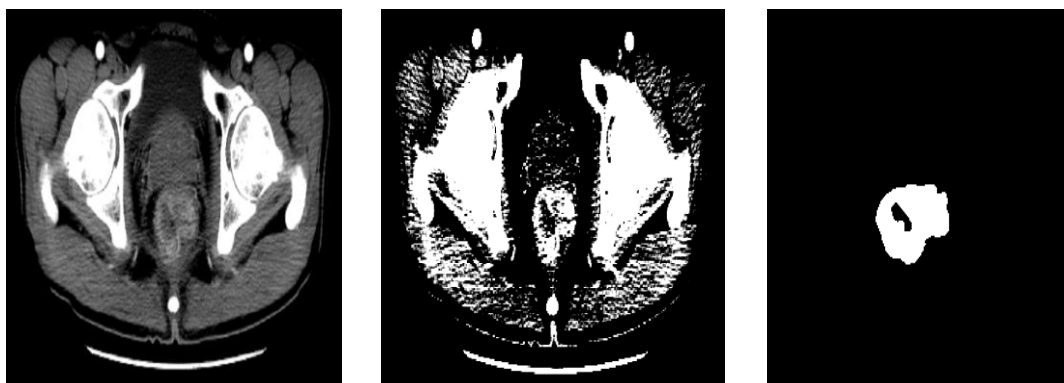


图 3 某病人含有肿瘤的 CT 影像改变窗宽、窗位前后

（左图：原始 CT 影像；中图：调节窗口后的 CT 影像；右图：标记肿瘤掩模）

（4） 图像直方图均衡化

利用直方图均衡化可以增强 CT 影像局部的对比度而不影响整体的对比度，从而进一步凸显肿瘤图像的细节。均衡化的具体操作不再赘述。

（5） 数据增强

为了使 U-Net 网络取得更好肿瘤分割效果，使用数据增强的方法扩展训练集。

- 平移：对训练集中每张原始 CT 影像随机向上、下、左、右四个方向平移 0~1%倍图像边长；
- 旋转：对训练集中每张原始 CT 影像绕图像几何中心随机旋转 0~3°；
- 缩放：对训练集中每张原始 CT 影像随机缩小或放大 0~1%；

3.2.2 后处理

在做肿瘤分割预测时发现，部分 CT 影像（特别是没有肿瘤的 CT 影像）预测出了多个肿瘤区域，需要做进一步处理，如图 4 所示。

（1） MPAT 过滤

由于每个病人拍摄 CT 影像的位置一致，并且肿瘤在直肠中出现的位置应有一定范围，因此可以统计数据集中所有肿瘤掩模图像中肿瘤出现区域的最大范围，将其作为肿瘤出现的最大可能范围（Most Possible Area of Target, MPAT），预测结果

中超出此范围的肿瘤区域视为错误分割，并被去除。MPAT 如图所示。

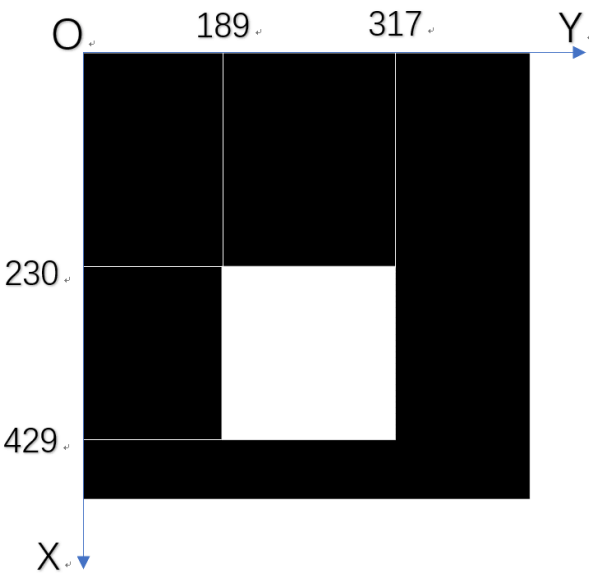


图 3 肿瘤出现的最大可能范围，MPAT

(2) 连通域阈值过滤

在上一步处理后，仍有部分肿瘤预测影像中 MPAT 中存在多个肿瘤区域，采用阈值法选择肿瘤区域。当最大区域的连通域大小（像素个数）大于某一阈值时，保留此最大区域，去除其他较小的区域，否则，去除所有区域，即视该 CT 图像中不存在肿瘤。经过多次实验调整发现，当阈值设为 100 时，过滤效果最佳。

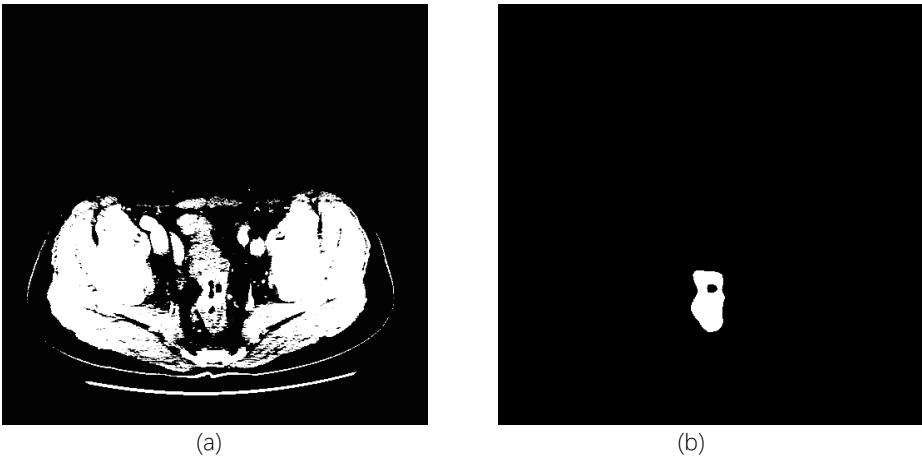




图 5 分割图像后处理效果

(a)为预处理后的 CT 影像；(b)为标记的直肠肿瘤掩模；(c)为 UNet 预测肿瘤掩模；(d)为后处理的掩模。

3.3 实验设置及结果分析

将全部数据集（107 个病人的直肠动、静脉期 CT 影像及其已标记肿瘤掩模）经过图像预处理操作后，随机选取 11 人的数据作为验证集，剩下的作为训练集进行训练。

经过多次实验调整，选取 Adam 作为优化器，训练 U-Net 网络时相关参数设置如表 1。

表 1 网络训练时的相关参数

参数	参数值
学习率	10^{-5}
epoch	200
batchsize	4
clipvalue	10^{-4}

最终实验结果表明，通过动、静脉期 CT 影像的训练模型的分割效果几乎没有差别。采用动脉期的训练模型继续完成后续特征提取、选择和直肠肿瘤是否淋巴结转移的诊断任务。最终训练得到的模型在验证集上的肿瘤分割的 Dice 系数为 0.8956，对全部 CT 影

像（包含没有肿瘤的 CT 影像）进行预测，为了同时评估模型在没有肿瘤的 CT 影像上的预测效果，对 Dice 系数计算公式做平滑处理：

$$Dice(A,B)=\frac{2|A\cap B|+1}{|A|+|B|+1}$$

全部数据集上预测结果的 Dice 系数为 0.4800，分割图像经过后处理后的 Dice 系数达到了 0.7042。在精确分割肿瘤影像作为实践中一个难点的背景下，我们的模型在直肠肿瘤的分割上取得了相当好效果。

4 任务二：直肠肿瘤部分特征的提取

4.1 相关特征

影像组学的常用的特征大致可归纳如下：

- 一阶统计量计算出量化的肿瘤强度特征，从肿瘤体素强度直方图计算求得；
- 基于肿瘤形状的特征，如肿瘤的体积、表面积等；
- 肿瘤纹理特征，在肿瘤内的所有三维方向上计算，从而得出每个体素与周围体素相对的空间位置；
- 从原始图像的小波分解中计算强度和纹理特征，从而提取肿瘤内不同频段上的特征。

4.2 特征提取

同时对任务一中得到的肿瘤分割结果和数据集中医生标记直肠肿瘤掩模做以下特征提取。

为方便地对肿瘤的相关特征进行高效率地提取，采用专门用于医学影像特征提取的 python 第三方库 pyradiomics 提取特征。由于数据集中病人 CT 影像和肿瘤掩模分别以 dicom 和 png 格式单张存储，为提取更多 3D 特征，需将每个病人的所有 CT 影像按顺序叠加并整合成一个 nrrd 格式文件，肿瘤掩模也是如此。

利用 pyradiomics 提取肿瘤区域的具体特征如表 2 所示，已覆盖 4.1 中所述所有相关特征。

表 2 CT 影像特征

类型	特征	特征数量
一阶特征 (First Order Statistics)	能量、熵、标准差、灰度值范围等	19
形状特征 (Shape-based)	表面积、体积、球形度、平坦度等	16
灰度共生矩阵 (Gray Level Cooccurrence Matrix)	对比度、相关度、联合熵等	24
灰度区域矩阵 (Gray Level Size Zone Matrix)	灰度不均匀性、区域不均匀性、灰度级方差等	16
灰度游程矩阵 (Gray Level Run Length Matrix)	灰度短（长）游程增强、灰度游程方差、灰度游程熵等	16
灰度色差矩阵 (Neighbouring Gray Tone Difference Matrix)	粗糙度、对比度、复杂度等	5
灰度相关矩阵 (Gray Level Dependence Matrix)	灰度级方差、灰度级不均匀性、依赖熵等	14

我们对每个病人原始 CT 影像分别做小波变换（3 个维度共 8 个变换图像）等 9 种滤波操作，在生成的其它 9 种类型图像上同时提取表 2 中的相关特征，进一步提高特征数量，生成的图像类型如表 3 所示。最终共 1323 个影星组学特征（均为数值型特征）被提取，由于特征数量太多且每个特征的意义及计算方式不是本问题的主要任

务，因此不再赘述。

表 3 CT 影像经过不同滤波的生成图像类型

图像类型	描述
Wavelet	小波滤波，在三个维度中的每个维度中应用高通或低通滤波器的所有可能组合，生成 8 个变换图像。
LoG(Laplacian of Gaussian filter)	高斯滤波器的拉普拉斯算子，边缘增强滤波器。
Square	取图像强度的平方并将其线性缩放回原始范围。
SquareRoot	获取绝对图像强度的平方根并将其缩放回原始范围。
Logarithm	取绝对强度+ 1 的对数。在应用滤波器后，将值缩放到原始范围。
Exponential	采用 e 的指数滤波。应用滤波器后，将值缩放到原始范围。
Gradient	返回局部梯度的大小。
LocalBinaryPattern2D	在分片操作（2D）中计算局部二进制模式。
LocalBinaryPattern3D	使用球谐函数计算 3D 中的局部二进制模式。

5 任务三：肿瘤影像特征与淋巴结转移的相关性验证

5.1 建立分类模型

5.1.1 数据平衡

题目给出的数据集包含 107 个直肠癌患者的 CT 影像和标记直肠肿瘤掩模，其中肿瘤淋巴结转移阴性 62 人，阳性 45 人。为了减少建立模型时，由于不同组别数据量差异造成模型准确率减低（肿瘤淋巴结转移阴性较多，建立的模型会对淋巴结转移阴性过拟合）。在训练分类器前，平衡了不同组别患者数据量的差异，并没有采用全部患者的数据，而是截取淋巴结转移阴性 62 人中 45 人的数据，使得肿瘤淋巴结转移阴性和阳性数据量相同，从而消除由于数据量差异导致的误差。

5.1.2 训练分类器

为获得一个基于肿瘤影响组学特征判断是否发生淋巴结转移的分类器，并且最大程度提升分类器性能，用包裹式特征选择的方法直接针对给定分类器进行优化^[5]，并且将病人的性别、年龄信息纳入原始特征集，用 RFE 方法递归地从原始特征集中移除最不重要的特征。采用随机森林作为分类器，随机森林由多棵分类回归树 (CART) 构成，每一棵 CART 通过 Boot-strapping 训练子分类器，最终通过每一棵 CART 的投票得到诊断分类结果（肿瘤淋巴结转移阴性、阳性）。限制树的数目为 100。为避免过拟合，在分类器性能不明显降低的前提下选择尽量少的特征数目(最大树深)，采用十折交叉验证评价分类器的诊断效果。

5.1.3 评价指标

使用 F-Score 对分类结果进行评价：

$$F = \frac{2PR}{P + R}$$

其中 P 为查准率 (Precision)， R 为查全率 (Recall)。

5.2 结果分析

基于数据集中标注的肿瘤掩模提取特征、选择特征并训练分类器的结果如下。在分类器性能不明显降低的前提下选择尽量少的特征，共选择出了 20 个特征，主要是小波分解图像、squareroot 滤波图像的部分高阶特征，输出如下：

```
selected features: ['wavelet-LLH_ngtdm_Complexity', 'wavelet-LHH_ngtdm_Busyness', 'wavelet-
HLL_gldm_InverseVariance', 'wavelet-HLL_gldm_SumEntropy', 'wavelet-
HLL_gldm_DependenceNonUniformityNormalized', 'wavelet-HLL_gldm_DependenceVariance', 'wavelet-
HLL_gldm_LongRunEmphasis', 'wavelet-HLH_gldm_Correlation', 'wavelet-HLH_gldm_SumEntropy',
'wavelet-HLH_gldm_LongRunEmphasis', 'wavelet-HLL_firstorder_Mean', 'wavelet-
HHH_ngtdm_Complexity', 'wavelet-HHH_ngtdm_Contrast', 'wavelet-LLL_firstorder_90Percentile',
'wavelet-LLL_gldm_DifferenceVariance', 'wavelet-LLL_ngtdm_Coarseness', 'log-sigma-3-0-
mm-3D_glszm_SizeZoneNonUniformityNormalized',
'square_gldm_RunLengthNonUniformityNormalized', 'squareroot_gldm_GrayLevelNonUniformity',
'squareroot_ngtdm_Coarseness']
```

通过十折交叉验证，平均准确率为 0.7300。查全率为 0.6860，查准率为 0.7799，计算得到 $F - score = 0.6878$ 。基于以上 20 个肿瘤影像特征对肿瘤是否发生淋巴结转移做出判断，ROC 曲线如图 4 所示，ROC 曲线下面积 $AUC=0.8$ ，平均准确率明显大于 0.5，并获得了 0.6878 的 F-score，验证了两者的相关性。

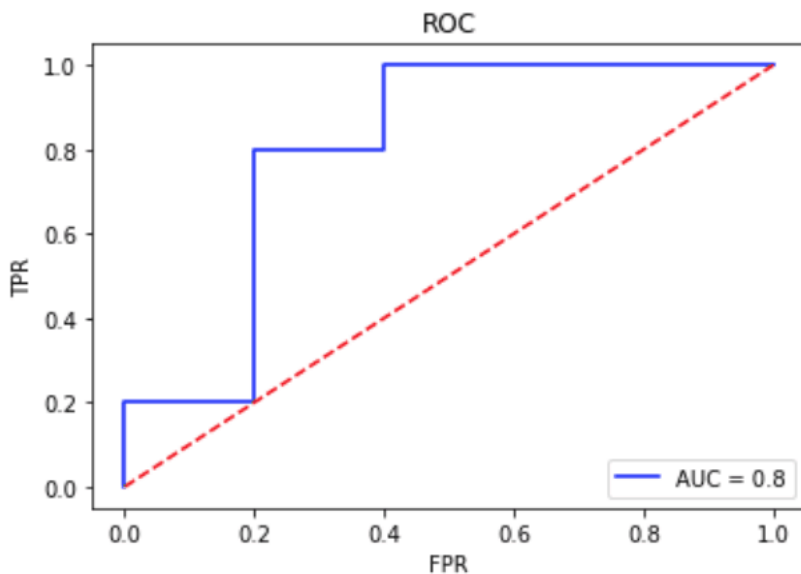


图 4 基于标记肿瘤掩模建立是否淋巴结转移分类器的 ROC 曲线

6 肿瘤分割不精确的情况

另外，为讨论当所选取区域适当扩大（例如覆盖一部分肿瘤的周边组织）或适当缩小（例如只选取肿瘤的部分核心区域）时，是否仍然能构建有效的模型来判断肿瘤是否已发生淋巴结转移。基于任务一中肿瘤分割模型预测得到的肿瘤掩模提取特征、选择特征并训练分类器，结果如下。

最终也选择出了 20 个特征，主要是小波分解图像、LoG 滤波图像、和 exponential 滤波图像的部分高阶特征，如下：

```
reduced number of features: 20
selected features: ['wavelet-LHL_firstorder_Mean', 'wavelet-LHL_firstorder_Median',
'wavelet-LHL_gldm_GrayLevelNonUniformity', 'wavelet-LHL_glrIm_GrayLevelNonUniformity',
'wavelet-LHH_gldm_GrayLevelNonUniformity', 'wavelet-HLL_firstorder_Mean', 'wavelet-
HLL_glszm_SizeZoneNonUniformityNormalized', 'wavelet-HLL_glszm_SmallAreaEmphasis', 'wavelet-
HLH_firstorder_Mean', 'wavelet-HLH_glcM_ClusterShade', 'wavelet-HLH_glcM_Idn', 'wavelet-
HHL_firstorder_Mean', 'wavelet-HHH_glszm_ZonePercentage', 'wavelet-LLL_firstorder_Kurtosis',
'log-sigma-3-0-mm-3D_glrIm_GrayLevelNonUniformity', 'log-sigma-3-0-
mm-3D_glszm_HighGrayLevelZoneEmphasis', 'square_glrIm_LongRunHighGrayLevelEmphasis',
'squareroot_firstorder_10Percentile', 'exponential_glrIm_RunLengthNonUniformity',
'exponential_glrIm_ShortRunEmphasis']
```

通过十折交叉验证，平均准确率为 0.7275。查全率为 0.7003，查准率为 0.6726，计算得到 $F-score = 0.6523$ 。说明用任务一中分割不完全准确的肿瘤掩模仍能构建有一定效果的淋巴结转移分类模型。

7 总结

综上所述，我们从直肠癌患者的 CT 影像出发，运用深度学习的方法训练肿瘤分割模型并做了预测肿瘤掩模的后处理优化，用 pyradiomics 充分提取肿瘤区域的特征，最终建立基于肿瘤区域影像特征的判段是否淋巴结转移的二分类模型，在相关临床信息缺少的情况下取得了一定的分类效果，可以自动化地完成直肠癌转移预测，为医生诊断提供一定的参考。

参考文献

- [1] 魏伟, 刘振宇, 王硕, et al. 影像组学技术研究进展及其在结直肠癌中的临床应用[J]. 中国生物医学工程学报, 2018, 37(05):4-11.
- [2] Ronneberger O , Fischer P , Brox T . U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. 2015.
- [3] 黄荔丽, 王博亮, 黄晓阳. 基于 DICOM 格式的肝脏肿瘤 CT 图像分割[J]. 计算机技术与发展, 2008, 18(1):48-51.
- [4] 郭钰, 李明洋, 刘祥春, et al. CT 影像组学对结直肠癌肝转移的诊断价值[J]. 中国临床医学影像杂志, 2018, 29(11):45-49.
- [5] Huang Y . Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer[J]. Journal of Clinical Oncology, 2016, 34(18):2157-2164.