

# 联合成对学习和图像聚类的无监督肺癌亚型识别



任雪婷<sup>1</sup> 赵涓涓<sup>1</sup> 强彦<sup>1</sup> Saad Abdul RAUF<sup>1</sup> 刘继华<sup>2</sup>

<sup>1</sup> 太原理工大学信息与计算机学院 太原 030024

<sup>2</sup> 吕梁学院计算机科学与技术学院 山西 吕梁 033000

(964325345@qq.com)

**摘要** 基因诊断是近年来提高肺癌治愈率的一种新型且有效的方法,但这种方法存在基因检测时间长、费用高、侵入式取样损伤大的问题。文中提出了基于成对学习和图像聚类的无监督学习的肺癌亚型识别方法。首先,采用无监督卷积特征融合网络用于学习肺癌 CT 图像的深度表示,有效地捕捉被忽略的重要特征信息,并使用包含不同层次抽象信息的最终融合特征来表征肺癌亚型。然后,使用联合成对学习和图像聚类的分类学习框架进行建模,充分利用学习到的特征表示,确保有效的聚类学习,以取得更高的分类精度。最后,利用生存分析和基因分析对肺癌亚型进行多角度验证。在合作医院和 TCGA-LUAD 数据集上的实验结果表明,该方法通过可靠无创的影像分析和放射成像技术,发现了 3 种具有不同分子特征的肺癌影像亚型,在降低基因检测问题的同时可有效辅助医师进行精准诊断和个性化治疗,进而提高肺癌患者的治愈生存率。

**关键词:** 肺癌亚型识别;成对学习;图像聚类;无监督学习;深度表示

中图法分类号 TP391.41

## Lung Cancer Subtype Recognition with Unsupervised Learning Combining Paired Learning and Image Clustering

REN Xue-ting<sup>1</sup>, ZHAO Juan-juan<sup>1</sup>, QIANG Yan<sup>1</sup>, Saad Abdul RAUF<sup>1</sup> and LIU Ji-hua<sup>2</sup>

<sup>1</sup> School of Information and Computer Science, Taiyuan University of Technology, Taiyuan 030024, China

<sup>2</sup> School of Computer Science and Technology, Lvliang University, Lvliang, Shanxi 033000, China

**Abstract** In recent years, gene diagnosis has been one of the new and effective methods to improve the cure rate of lung cancer, but it has the problems of time-consuming, high cost and serious damage from invasive sampling. In this paper, an unsupervised learning method of Lung cancer subtype recognition based on paired learning and image clustering is proposed. Firstly, the unsupervised convolution feature fusion network is used to learn the deep representation of lung cancer CT images and effectively capture the important feature information that is ignored, and the final fusion features containing different levels of abstract information is used to represent lung cancer subtypes. Then, the classification learning framework of combined paired learning and image clustering is used for modeling, and the learnt feature representation is fully utilized to ensure effective clustering learning, so as to achieve higher classification accuracy. Finally, survival analysis and gene analysis are used to verify lung cancer subtypes from multiple perspectives. Experiments on the data sets of the cooperative hospital and TCGA-LUAD show that, through reliable and non-invasive image analysis and radiological imaging technology, three subtypes of lung cancer with different molecular characteristics have been found by this method. It can effectively assist doctors in accurate diagnosis and personalized treatment while reducing problems in gene detection, so as to improve the survival rate of lung cancer patients.

**Keywords** Lung cancer subtype recognition, Paired learning, Image clustering, Unsupervised learning, Deep representation

### 1 引言

肺癌是成人中最常见、最致命的原发性恶性肿瘤。肺癌

中晚期(II、III、IV 期)患者的 5 年平均生存率在 5% 以下,晚期肺癌患者中位生存期为 6~8 个月。因此,如何提高肺癌患者的生存率,甚至实现治愈,成为医学领域的重大难题。而近

收稿日期:2019-09-10 返修日期:2019-11-15 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61872261);北京航空航天大学虚拟现实技术与系统国家重点实验室项目(VRLAB2018A08);山西省重点研发计划国际科技合作项目(201803D421036)

This work was supported by the National Natural Science Foundation of China (61872261), Open Funding Project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2018A08) and International Science and Technology Cooperation Project of Shanxi Key R&D Program (201803D421036).

通信作者:赵涓涓(zhaojuanjuan@tyut.edu.cn)

些年,靶向治疗的提出为晚期肺癌患者带来了希望<sup>[1]</sup>。肺癌影像学亚型识别研究作为医学领域中重要的研究内容,具有广阔的应用前景。

尽管靶向治疗是对肺癌中晚期治疗的最有效手段之一,但是目前临床开展的靶向治疗还存在如下问题:1)基因检测时间长,基因检测分析需要的平均时间为2~8周,这对于晚期患者短暂的治疗时机来说时间过长;2)费用高,检测费用为常规治疗的10倍以上,给患者带来了较大的经济负担,部分患者因此错失治疗时机;3)侵入式取样损伤大,基因检测所需的病灶样本在取样过程中会带来癌细胞扩散的风险,加重患者病情。在这种情况下,如何找到一种方法来有效代替基因检测效果,不需要对患者进行穿刺活检,也不需要花时间等待基因测序结果就能找到靶点是一个至关重要的问题。与基因检测不同,CT成像是非侵入性的,并且可以提供关于整个肿瘤表型的信息,包括肿瘤内异质性。因此,准确地识别肺癌影像学亚型是解决这一问题的有效方法。

目前,已有关于癌症影像学分型方法的研究,但相关研究在肺部影像中并不多见。在探索基因数据的分子亚型研究方面,Young等<sup>[2]</sup>使用无监督深度学习来学习癌症基因表达数据的层次结构,发现了生存差异的多形胶质母细胞瘤亚型,并发现差异表达的基因与亚型之间存在相关性。在探索影像特征与分子亚型的关系方面,Mazurowski等<sup>[3]</sup>从乳腺癌定量磁共振图像中提取23个影像学特征,在基因组分析的基础上确定了分子亚型,利用logistic回归和似然比检验评价影像学特征与分子亚型的相关性。随着肺癌靶向治疗方法的发展,早期肺癌的诊断和鉴别越来越重要。影像组学可以从医学图像中获得自动定量成像特征,无创预测结节和肿瘤<sup>[4]</sup>。在探索影像组学在肺癌方面的研究中,Zhang等<sup>[5]</sup>从CT影像中提取了大量定量特征,通过预测建模和特征选择提高了基于影像组学的非小细胞肺癌(Non-Small Cell Lung Cancer, NSCLC)预后预测能力。Thawani等<sup>[6]</sup>通过从放射图像中提取定量特征进行析来表述与疾病发病机制相关的图像亚视觉特征,研究表明这些特征与肺部CT上结节的恶性潜能之间存在相关性。这些结果突出了非侵入性和经济有效的影像组学工具在肺癌患者治疗选择和个性化治疗方面的长期潜力。

在探索影像组学亚型识别方面,Wu等<sup>[7]</sup>通过提取肿瘤及其周围实质的定量影像学表型,识别新的乳腺癌亚型,构建了一个基于基因表达的影像亚型分类器,并利用有基因表达数据但没有影像数据的样本测试了它们的预后意义。Itakura等<sup>[8]</sup>仅通过MR成像特征区分胶质母细胞瘤(GlioBlastoma, GBM)亚型,提取每个病灶的形状、纹理和边缘锐度的定量图像特征,识别出3个新的GBM亚组。这些研究大多采用聚类的方法进行癌症的无监督分层。而对于一组较大的未标记图像,如何基于提取的视觉表示有效地将其分组仍然是一个具有挑战性的问题。Yang等<sup>[9]</sup>提出了一种联合无监督学习的循环网络,通过将深度表示和图像聚类两个过程集成到一个统一加权三重态损失的模型中,来得到更精确的图像簇。

本文提出了联合成对学习和图像聚类的无监督分类模型(Unsupervised Classification Combined with Paired Learning and Image Clustering, UC-CPLIC)的肺癌影像学亚型识别方法。为了能够准确地捕获肺癌亚型,其只依赖于CT影像就可以准确地进行靶向治疗,利用CT定量图像特征实现肺癌病灶图像的深度特征表达,构建图像深度特征表示和无监督分类模型,利用深度神经网络对大批量未标记图像进行更加高效的特征表达,并联合成对学习 and 图像聚类以有效提高聚类性能。该方法在仅利用影像学数据的情况下,能够自动而准确地捕获具有基因特征的肺癌新亚型。在合作医院数据集和TCGA-LUAD数据集上的实验结果表明,本文算法取得的识别率高于一般的无监督学习算法。

## 2 联合成对学习和图像聚类的无监督影像学亚型识别方法框架

在肺癌影像学亚型识别问题中,存在少量标记数据样本和大量未标记数据样本,若要利用深度神经网络对有少量标记的图像进行更加高效的特征表达,提高聚类性能,并在仅利用影像学数据的情况下准确地捕获具有基因特征的肺癌新亚型,使用UC-CPLIC模型可能是一个较好的选择。本文给出的系统框架包括3个模块,即深度表示和聚类学习模块、配对学习模块和肺癌影像学亚型验证与评价模块,如图1所示。

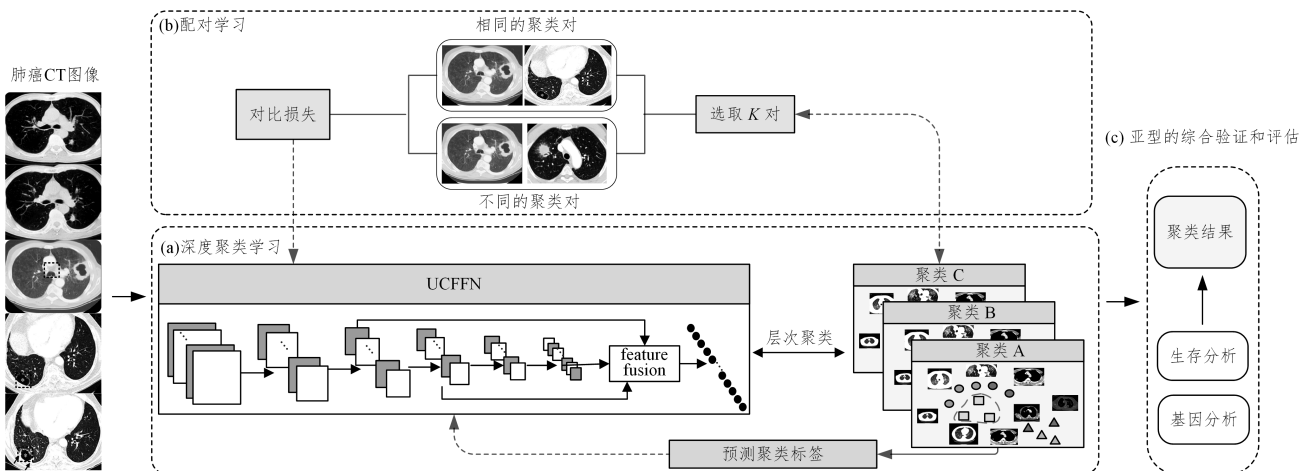


图1 UC-CPLIC总体框架图

Fig. 1 UC-CPLIC general frame diagram

本文方法的主要步骤如下:

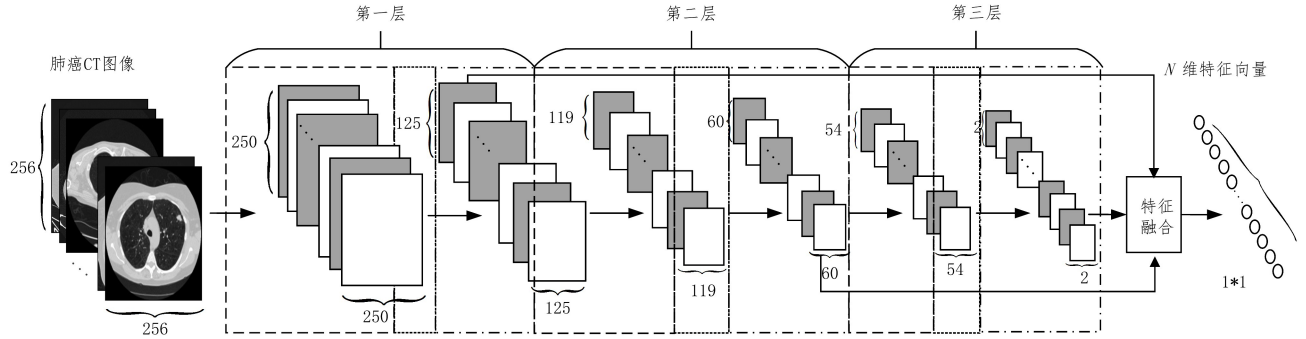
(1)通过无监督卷积特征融合网络(Unsupervised Convolutional Feature Fusion Network, UCFFN),较全面地表征图像特征的隐含信息,为聚类学习和配对学习做好准备。

(2)对获得的影像学特征层次聚类,以确定最佳聚类数目。

(3)得到初始聚类结果后进行联合配对学习,根据样本之间的特征距离采集相同的样本对和不同的样本对。

(4)给定相同的簇和不同的簇输入对,使用对比损失函数对 UCFFN 进行训练。根据预测的标签来更新 UCFFN 的网络参数,并迭代该过程。

本文对得到的肺癌影像学亚型进行多角度分析,通过生



注:输入图像尺寸为  $256 \times 256 \times 1$ ,不同的方框表示不同的运算类型:  $\square$ 表示卷积运算,  $\square$ 表示池化运算,  $\square$ 表示非线性运算

图2 UCFFN流程图

Fig. 2 Unsupervised convolutional feature fusion network flowchart

如图2所示,利用多层特征是增强肺癌CT图像深层表示的一种可行方法。图2中,每一层由卷积、池化和非线性操作组成。该网络将输入的肺癌CT图像转换成不同层次的特征图,经过特征融合算子将其融合成一个  $n$  维特征向量。

设  $P$  是尺寸为  $x \times y \times z$  的CT图像,其中  $x, y, z$  分别对应通道的宽度、高度和数量。 $Conv^l(\cdot)$  表示第  $l$  个卷积层的变换,一共包含3个操作,即卷积、池化和非线性运算。 $l$  层的特征为:

$$f^l = Conv^l(f^{l-1}) \quad (1)$$

其中,  $l=1, 2, \dots, L, L$  是CNN的层数,  $f^0$  表示CT图像  $P, f^L$  表示最终的输出图像。

对于所提出的UCFFN,通过特征融合算子  $F$  不仅可以得到最后一层的输出特征,还可以得到中间层的输出特征,得到的特征可以用公式表示为:

$$f^{FN} = F(S(f^1, f^2, \dots, f^L)) \quad (2)$$

其中,  $S$  表示从  $\{f^1, f^2, \dots, f^L\}$  中选择子集的操作,而  $f^{FN}$  表示融合来自各个子集特征的总特征。

对于本文提出的肺癌CT图像深度表示方法,需要选择一种合适的融合策略进行特征融合。首先采用差值算法将图像放缩到与第三层特征图相同的大小,本文分析了图3所示的两种融合方法。第一种是将不同层的特征向量相加,如图3(a)所示,这种融合策略以一种简单的方式获得组合的特征,但很难保留来自不同层特殊的特征信息,会丢失局部信息。第二种将不同层的特征串联成一个组合的特征向量,如图3(b)所示,即将各层的信息进行组合,同时完全保留各层特

征分析和基因分析对其进行综合验证和评价。

### 3 无监督卷积特征融合网络

由于不同层包含不同级别的抽象信息,而一般的CNN框架通常使用端到端的学习方法,仅根据最后一层的特征,学习模型无法充分获取每一层的输出特征,很难在不同层的输出中捕捉到重要信息。为了提高表示性能,本文提出无监督卷积特征融合网络来构建一个高效的CNN肺癌CT图像深度表示,将不同层输出的特征进行融合。这种方法可以有效地捕捉被忽略的重要特征,并在最终的融合特征中包含不同层次的抽象特征。UCFFN流程如图2所示。

因此,本文选择第二种融合方法进行不同层特征融合,  $f^{FN}$  融合了所选子集  $\{f^i, f^j, \dots, f^k\}$  的特征,最终可以表示为:

$$f^{FN} = f^i \oplus f^j \oplus \dots \oplus f^k \quad (3)$$

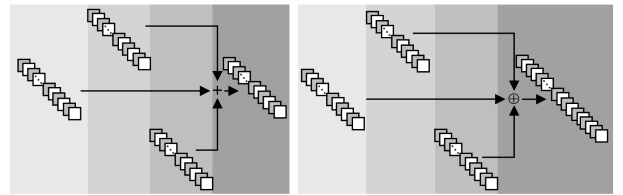


图3 不同的融合方法

Fig. 3 Different fusion operators

### 4 联合成对学习和图像聚类模型(UC-CPLIC)

目前的癌症影像学识别算法通常采用聚类的方法进行癌症的无监督分层,需要有大量被标记数据才能准确得到分类结果。而我们从医院得到的数据大多数是未标记的,为充分利用这些数据,本文采用联合成对表示学习和聚类算法,通过成对学习方案从每个聚类迭代的结果中适当选取相同聚类和不同聚类的图像对,有效地学习了判别表示。

大多数基于深度学习的图像聚类方法都是通过深度网络提取特征来预测输入图像的聚类标签。研究人员大多根据图像的显著特征进行分组,对于很多实际应用来说,图像内部的隐含信息和患者检测出的基因联系也尤为紧密。在这种情况下,利用成对学习对比损失函数对图像聚类和特征学习进行联合求解和更新,有助于得到更加准确的分类结果。



步长 $[t_p^s, t_p^e]$

输出:最终图像聚类标签 $z^*$ ,UCFFN 参数 $\theta^*$

1.  $t \leftarrow 0; p \leftarrow 0$
2. 初始化  $\theta$  和  $z$
3. for
4. 通过合并两个聚类更新  $y^i$  到  $y^{i+1}$
5. 通过两两配对学习更新  $z^i$  到  $z^{i+1}$
6. if  $t = t_p^s$
7. 通过训练 UCFFN 更新  $\theta^p$  到  $\theta^{p+1}$
8.  $p \leftarrow (p+1)$
9. end if
10.  $t \leftarrow (t+1)$
11. 直到聚类数量达到  $n_c^*$
12. end for
13.  $z^* \leftarrow z^i; \theta^* \leftarrow \theta^p$

联合成对学习 和 图像聚类学习 算法,将成对学习 与 基于 UCFFN 的层次聚类相结合,利用对比损失函数对图像聚类和特征学习进行联合求解和更新。将提取的单个特征作为样本进行聚类,分配聚类标签,并根据预测的标签选择相同的聚类对和不同的聚类对,应用成对学习 方法对 UCFFN 的参数进行更新,迭代循环直到聚类结果收敛到一个稳定点,达到提升预测准确度的效果。

## 5 实验及结果分析

在合作医院的 219 例和 TCGA-LUAD 的 186 例 CT 图像数据集上分别测试本文方法。合作医院和 TCGA-LUAD 的图像由  $256 \times 256$  的 CT 图像组成。关于候选结节区域的检测,根据实验室前期的一些工作,本文采用一种基于多尺度增强滤波器和三维形状特征的肺部病灶区域自动检测方法<sup>[10]</sup>。文献<sup>[10]</sup>利用阈值迭代和种子点的区域增长算法自适应地实现肺实质图像的分割。构建两种多尺度增强滤波器对三维肺图像中的结节和血管图像进行增强,并将结节图像中的大部分血管图像去除,得到疑似结节图像,然后使用 18 邻域生长算法提取肺结节,获得结节的精细分割。分割结果如图 5 所示。



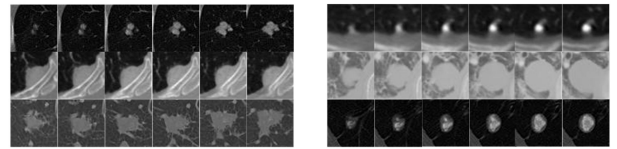
图 5 分割结果图

Fig. 5 Segmentation results

在 UCFFN 实验中采用无监督实验设置,设置小批量数据集大小为 50,卷积大小为  $7 \times 7$ ,卷积核的个数为 64,步长为 1,每一层的最大池化为  $2 \times 2$ ,步长为 2。从均值为 0、方差为 0.01 的高斯分布中采样,将其作为初始权值。将提出的 UCFFN 作为特征提取器,提取的特征将被输入联合图像聚类和成对学习模型中进行肺癌影像学亚型分类。通过亚型分类的准确性间接评价该无监督表示方法的能力。本文方法在

Intel Core i5-7400 CPU 和 16GB RAM 的工作站上使用 Tensorflow 实现。GPU 卡 (Nvidia GTX Titan X) 加速了训练过程。

由于合作医院数据集和 TCGA-LUAD 数据集的真实标签是可用的,本文所选取的 CT 图像在两个数据集中均可分为 3 个类别,我们给出了每个类别特征较为显著的几幅图像,如图 6 所示。对于所有算法,我们将簇的数量设置为真实类别的数量。本文研究的目的是通过提取肿瘤的定量图像表型和肺实质来发现新的肺癌亚型,表明成像亚型与不同的分子通路相关,并且每一类成像亚型中的患者在多个独立的类别中有不同的预后。通过影像特征统计分析得出,亚型 1 的特点为结节体积小,病灶边缘凹凸不平,在支气管、血管进出病灶处往往呈现凹陷;亚型 2 的特点为肿块边缘不同程度棘状或毛刺样突起;亚型 3 的特点为结节内小灶透光区直径小于 5mm,结节间含气腔隙。通过实验对比监督聚类算法(本文中采用 K-means 聚类模型)和本文提出的无监督聚类算法的分类准确率,同时对比本文算法与其他无监督经典算法。



(a) 合作医院

(b) TCGA-LUAD

图 6 CT 图像数据集

Fig. 6 CT image data set

表 1 列出了 K-means 与 UC-CPLIC 在两个数据集上的准确率、精确率、召回率和 F 值的对比结果。可以看到,在不同的数据集上,UC-CPLIC 都取得了较好的学习效果,本文方法的准确率更高。由此表明,UC-CPLIC 可以有效利用大量未标记样本提升学习系统的泛化能力,从而达到较好的识别效果。

表 1 两种方法在两个数据集上的实验结果

Table 1 Experimental results of two methods on two datasets

DataSet	Method	Cooperative Hospital	TCGA-LUAD	Average Value
Accuracy rate	K-means	0.81	0.83	0.82
	UC-CPLIC	0.94	0.94	0.94
Precision rate	K-means	0.81	0.83	0.82
	UC-CPLIC	0.95	0.91	0.93
Recall rate	K-means	0.81	0.83	0.82
	UC-CPLIC	0.94	0.92	0.93
F value	K-means	0.81	0.83	0.82
	UC-CPLIC	0.91	0.93	0.92

表 2 列出了 UC-CPLIC 与其他无监督学习方法的实验对比结果。用于对比的聚类算法有 4 个,其中前两个是自训练的学习方法。第一个为自训练 1 (Self-Training 1, ST 1),它是基于 K-means 的无监督学习算法。首先,随机选取  $K$  个对象作为初始的聚类中心,计算每个对象与各个种子聚类中心之间的距离,把每个对象分配给距离它最近的聚类中心,每分配一个样本更新一次聚类中心;然后,使用成对学习算法,根据预测的标签选择相同的聚类和不同的聚类对进行重新训

练,不断迭代重复这个过程直到取得最佳聚类结果。第二个是自训练2(Self-Training 2, ST 2),它是无监督学习算法,将深度信念网络(Deep Belief Network, DBN)作为学习器。其余两个对比算法是基于密度的聚类算法 DBSCAN 和深度嵌入式聚类算法(Deep Embedded Clustering, DEC),它们利用深度神经网络同时学习特征表示和聚类分配,学习从数据空间到较低维特征空间的映射,并在此映射中迭代优化聚类目标。从表 2 可以看出, UC-CPLIC 在两种数据集上的正确率都优于其他方法。

表 2 5 种方法在两个数据集上的正确率对比

Table 2 Accuracy comparison of five methods on two datasets

DataSet	Cooperative Hospital	TCGA-LUAD
ST1	0.75	0.71
ST2	0.83	0.85
DBSCAN	0.79	0.82
DEC	0.91	0.92
UC-CPLIC	0.95	0.93

通过实验对聚类结果进行评估,通过聚类过程识别肺癌影像学亚型,这些新发现的影像学亚型还需进一步验证,如果成功,其可能成为肺癌个体化治疗的有用生物标志物。下面从总体生存率和分子亚型组成两方面分析已识别的肺癌影像学亚型,并对其进行进一步的验证。

生存分析的主要目的是估计生存函数,采用 Kaplan-

Meier 法对肺癌亚型进行生存分析,如图 7 所示, Kaplan-Meier 曲线显示不同亚型的生存率存在显著差异(log-rank,  $p < 0.001$ )。亚型 1 的生存时间最短,亚型 3 的生存时间最长。

表 3 列出了肺癌分子亚型<sup>[11]</sup>中每一种分子亚型与影像亚型(亚型 1/2/3)的关系<sup>[12-13]</sup>, 1.1 型、1.2 型、1.3 型和 2.1 型和 3.1 型的比例分别为 23.61%, 21.32%, 15.48%, 19.59% 和 20.01%。如图 8 所示,分子组成与影像学亚型之间的关系<sup>[11]</sup>表明,影像学亚型间分子亚型的分布有显著性差异(采用卡方分布,  $p < 0.05$ )。肺癌在不同患者之间存在显著异质性,每一种影像学亚型都与不同的具有治疗意义的分子通路相关。利用成像技术对这种异质性进行非侵入性的体内表征,可为精确诊断和治疗提供帮助。

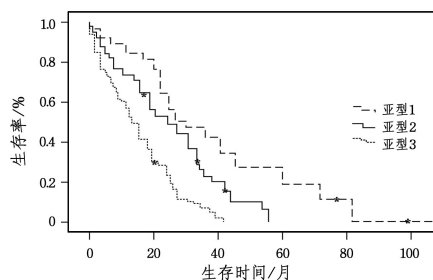


图 7 Kaplan-Meier 曲线图

Fig. 7 Kaplan-Meier curves

表 3 肺癌的分子亚型及靶向药物作用下的预后缓解发生率

Table 3 Molecular subtypes of lung cancer and incidence of prognosis remission under action of targeted drugs

聚类亚型	患者人数(n=219)	突变基因	信号通道	相关的组织学亚型	靶向药物	预后缓解发生率/%
1.1	46	EGFR 受体突变	EGFR	肺腺癌	吉非替尼	65
1.2	36	EGFR 耐药突变	EGFR	肺腺癌	阿法替尼	72
1.3	29	Veristat 蛋白质组学特征	EGFR	肺腺癌	奥希替尼	64
2.1	44	K-ras 基因突变	K-ras	肺腺癌	克唑替尼	44
3.1	64	EML4-ALK	EML4-ALK	肺腺癌	Keytruda(K 药)	89

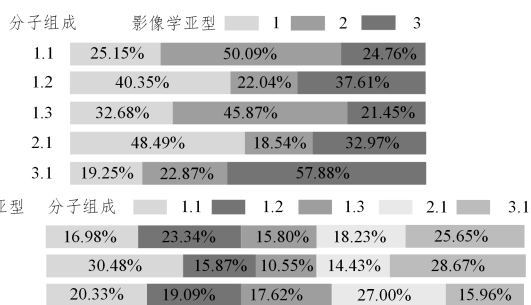


图 8 分子组成和影像学亚型的关系

Fig. 8 Relationship between molecular composition and imaging subtypes

由对肺癌亚型聚类的临床关联性研究可知,在合作医院的 219 例患者中,根据突变基因的不同,合作医院的专业医师给出了针对不同亚型患者服用的相应靶向药物,通过对每种亚型的患者用药治疗后的预后缓解发生率可知,亚型 3.1 和亚型 1.2 相比于亚型 1.1、亚型 1.3 和亚型 2.1,病程更长,耐药性更强,而且晚期肺癌的死亡率更高,由此说明肺癌亚型聚

类结果与靶向药物治疗预后结果具有强关联。

**结束语** 针对肺癌晚期患者基因检测时间长和侵入式取样损伤大的问题,本文提出联合成对学习和图像聚类的无监督肺癌亚型识别方法,没有寻找与预先定义的基因组特性相关的影像学特征,而是通过影像学表型的广泛表征和无监督聚类来发现亚型,利用无监督卷积特征融合网络对肺癌 CT 影像学特征进行深度特征表达,然后采用成对学习和图像聚类模型对 UCFFN 进行深度训练。充分利用图像特征的有用信息和大量未标记样本的潜在信息,增强肺癌亚型识别的准确率。实验结果表明,本文方法在准确率方面优于已有的聚类学习方法,因此本文方法可行且有效。但本文仍存在以下问题:1)联合成对学习和图像聚类时需要预先提取全方位多角度图像的深度表示<sup>[14]</sup>,如果图像的深度特征提取不全面,则会影响亚型分类的准确率,如何全面且不冗余地提取图像的深度表示,是下一步研究的重点;2)与以往的影像基因组学研究相似,本文方法虽然确定了影像学表型和分子特征之间的相关关系,但不一定是因果关系。为了验证这些影像基因

组关联,需要使用临床前敲除模型进行实验验证,这需要在未来的研究中进行进一步的研究。

## 参 考 文 献

- [1] SHAO L, SONG Z B, ZHANG Y P, et al. Advances of Molecular Subtype and Targeted Therapy of Lung Cancer[J]. Chinese Journal of Lung Cancer, 2012, 15(9): 545-552.
- [2] YOUNG J D, CAI C, LU X. Unsupervised deep learning reveals prognostically relevant subtypes of glioblastoma[J]. BMC Bioinformatics, 2017, 18(S11): 381.
- [3] MAZUROWSKI M A, ZHANG J, GRIMM L J, et al. Radiogenomic Analysis of Breast Cancer: Luminal B Molecular Subtype Is Associated with Enhancement Dynamics at MR Imaging [J]. Radiology, 2014, 273(2): 365-372.
- [4] WILSON R, DEVARAJ A. Radiomics of pulmonary nodules and lung cancer [J]. Translational Lung Cancer Research, 2017, 6(1): 86.
- [5] ZHANG Y, OIKONOMOU A, WONG A, et al. Radiomics-based prognosis analysis for non-small cell lung cancer [J]. Scientific Reports, 2017, 7: 46349.
- [6] THAWANI R, MCLANE M, BEIG N, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician [J]. Lung Cancer, 2018, 115: 34-41.
- [7] WU J, CUI Y, SUN X L, et al. Unsupervised Clustering of Quantitative Image Phenotypes Reveals Breast Cancer Subtypes with Distinct Prognoses and Molecular Pathways [J]. Clinical Cancer Research, 2017, 23(13): 3334-3342.
- [8] ITAKURA H, ACHROL A S, MITCHELL L A, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities [J]. Science Translational Medicine, 2015, 7(303): 303ra138-303ra138.
- [9] YANG J, PARIKH D, BATRA D. Joint unsupervised learning of deep representations and image clusters [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-

tion. 2016: 5147-5156.

- [10] HAO R, QIANG Y, LIAO X, et al. An automatic detection method for lung nodules based on multi-scale enhancement filters and 3D shape features [J]. KSII Transactions on Internet & Information Systems, 2019, 13(1).
- [11] WEST L, VIDWANS S J, CAMPBELL N P, et al. A novel classification of lung cancer into molecular subtypes [J]. PLoS one, 2012, 7(2): e31906.
- [12] WU M, MA J. Association Between Imaging Characteristics and Different Molecular Subtypes of Breast Cancer [J]. Academic Radiology, 2016, 24(4): 426-434.
- [13] RATHORE S, AKBARI H, ROZYCKI M, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1 [J]. Scientific Reports, 2018, 8(1): 5087-5098.
- [14] LAMBIN P, RIOS-VELAZQUEZ E, LEIJENAAR R, et al. Radiomics: extracting more information from medical images using advanced feature analysis [J]. European Journal of Cancer, 2012, 48(4): 441-446.



**REN Xue-ting**, born in 1994, postgraduate, is a member of China Computer Federation. Her main research interests include medical image processing and deep learning.



**ZHAO Juan-juan**, born in 1975, Ph.D., professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include intelligent information processing and image recognition.