



# 中国科学院分子影像重点实验室

## 肺癌影像组学数据规范细则

影像组学 (Radiomics) 已经成为国际医学影像领域新兴的发展方向, 中国科学院分子影像重点实验室经过多年的发展, 在影像组学方向上取得了系列的研究成果, 形成了完整的自主知识产权体系, 并研发了一套针对肺癌的影像组学预测软件, 可以实现肺癌的良恶性、TNM 分期、生存期的预测, 并能给出最优的诊疗方案。为了推进与合作医院在影像组学上的研究工作, 分子影像重点实验室与合作医院共同制定了下面的肺癌影像组学数据规范细节, 希望后续的合作医院可以严格按照下面的格式要求提供肺癌数据, 共同推动影像组学在临床中的应用。

肺癌影像组学数据包括影像数据和预后信息两部分, 先就两部分进行分别说明:

### 1. 影像数据格式规范

#### 1) 数据来源

从螺旋 CT 扫描机上得到的肺部影像数据, 平扫和增强数据都可, 如果两者都有, 请完整提供。

#### 2) 数据格式

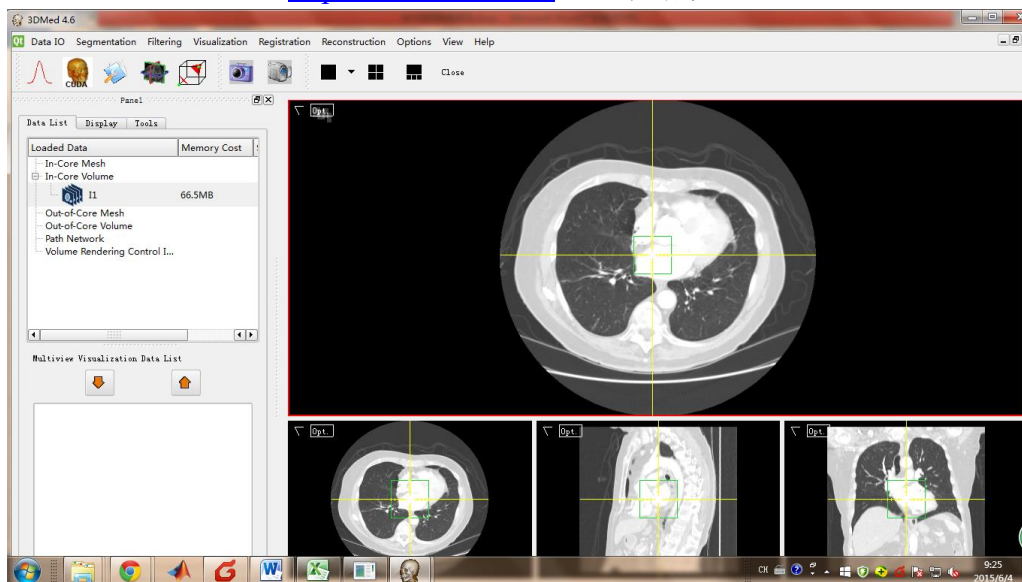
分子影像重点实验室支持 DICOM 格式的影像数据, 请勿提供压缩格式的 DICOM 数据, 以防止图像信息丢失, 数据可以使用 ITK-SNAP 或 3DMed 软件对影像数据进行读取测试, 以确定数据可否被分子影像重点实验室识别:

ITK-SNAP: 下载地址 <http://www.itksnap.org/pmwiki/pmwiki.php>, 以下为 ITK-SNAP 软件肺癌 DICOM 影像数据的图示。



ITK-SNAP 读取 DICOM 格式数据

3DMed: 下载地址 <http://www.mitk.net/>, 其数据读取截图如下图所示。



3DMed 读取 DICOM 格式数据

注意: DICOM 数据不要采用任何图像压缩方法, 以防止图像信息部分丢失, 影响预测结果。如有其它非 DICOM 格式的数据, 请联系 Radiomics 组责任老师董迪。

### 3) 数据要求

**图像大小:** 单层图像的像素数不小于 512\*512 像素, 单层影像大小一般在 512KB 及以上

**数据层厚:** 影像组学分析预测需要使用薄层的 CT 数据, 层厚需要限制在 0.625mm-2.5mm

**层数:** 对于薄层数据, 一个病人的影像层数应该在 100-600 层之间。

注意: 薄层数据的采集会增加病人的诊断费用, 多数医院可能仍采用厚层 (5mm) 的 CT 扫描, 这种情况下只需要保证同批数据的层厚相同, 例如都设为 5mm 即可, 但需要在提供数据时同时注明数据的层厚。

### 4) 数据存储

数据存储文件命名方式如下图所示 (本规范建议但不严格限制按照下面的存储命名范式):

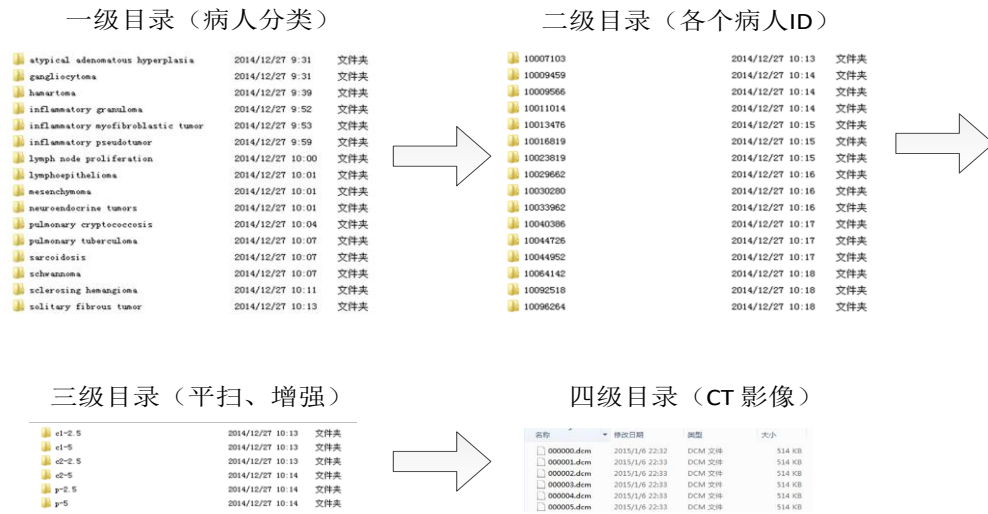
数据存储分为四级目录:

一级目录用于区分病人的病理类别;

二级目录用于区分不同的病人, 每个病人的文件夹使用 ID 号进行区分, 如果医院 ID 号系统不完善, 也可以采用 ID+病人姓名的命名方式;

三级目录用于区分同一病人不同的扫描序列, 可以是增强或者平扫、不同的层厚对应的体数据等;

四级目录为病人的影像数据文件。



## 2. 预后信息格式规范

预后信息应以表格的形式给出, 以下图示为预后信息的总览, 其格式请参考“[肿瘤临床信息规范.xlsx](#)”。合作单位应以此为参考模板, 但不局限于该模板, 请给出自己的肿瘤预后信息 (如显示不清晰, 可放大阅读):

表格中,

Order 表示需进行后续分析的肿瘤顺序;

分类表示良-恶性分类;



其中，“z\_imageposition”不是必需的，对于每一个肿瘤，需要的是 z(肿瘤所在层)，x,y (肿瘤在该层的位置)。详细格式请参考“[肿瘤位置规范.xlsx](#)”与“[肿瘤临床信息规范.xlsx](#)”。

## 2) 肺结节良恶性分类

在肿瘤预后一个关键步骤就是计算机辅助预测肿瘤的良恶性，因此医院在提供预后信息时，需要将肿瘤的良恶性分类信息明确的给出，其格式请参考“[肿瘤临床信息规范.xlsx](#)”。

## 3) 肺结节病理分型和 TNM 分期

与（2）中要求一样，需要合作单位提供每一个肺结节的 TNM 分期和病理分型。其格式如下，规范格式请参考“[肿瘤临床信息规范.xlsx](#)”。

118	lesion117	良性	硬化性血管瘤	sclerosing hemangioma	10816354	310	307	110	110				
119	lesion118	良性	硬化性血管瘤	sclerosing hemangioma	10817763	215	243	88	88				
120	lesion119	良性	硬化性血管瘤	sclerosing hemangioma	10836818	210	271	87	87				
121	lesion120	良性	硬化性血管瘤	sclerosing hemangioma	10860289	304	340	183	183				
122	lesion121	恶性	大细胞癌	large cell carcinoma	10355533	393	227	40	80	N	CT3N1M0	III A	
123	lesion122	恶性	大细胞癌	large cell carcinoma	10410228	343	304	63	126		pT3N0M1	IV	
124	lesion123	恶性	大细胞癌	large cell carcinoma	10487653	355	232	50	200		pT2N0M0	I B	
125	lesion124	恶性	大细胞癌	large cell carcinoma	10784175	338	305	176	176		T2N2M0	III A	

## 4) 肺癌病人存活期

预测生存率的关键是需要明确每一个病人的存活时间，因此合作单位在提供肿瘤预后信息时，需要给出病人的随访信息，**由于采集随访信息难度较大，因此建议院方提供一定数量可用的随访信息，随访信息需要具备：生存时间，删失标记，复发标记，病人基本信息（吸烟史等）。**其格式请参见“[随访规范.xlsx](#)”。

其中，

ID 表示院方在获取 CT 数据时的病人编号

Gender 表示性别，age 表示年龄，Smoke 表示吸烟史，statue\_death 表示删失标志，即该病人的随访结果是否为终点事件，survival 表示存活期。

A	B	C	D	E	F	G	H	I	J	K	L
ID	gender	age	Smoke	statue_de	Survival	TNM	临床分期	病理分期	病理分期	Order	分类
10016819	1	69	N	3	1049	T1N0M0	I A	腺癌	adenocarclesion326		恶性
10125291	1	51	N	3	2247	pT2N0M0	I B	鳞状细胞癌	squamous lesion266		恶性
10033962	1	78	N	3	961	T1aN0M0	I A	腺癌	adenocarclesion332		恶性
10130919	2	54	N	3	1128	T1bN0M0	I A	腺癌	adenocarclesion342		恶性
10230763	1	58	Y	3	11	T2N0M0	I B	腺癌	adenocarclesion356		恶性

## 5) 其他信息

病人血检信息能够客观反映出某些肿瘤标志物，对于肿瘤的预后具有较高的参考意义，因此建议合作单位能够提供病人的必要的**血检报告**。鉴于目前关于肿瘤基因测试的使用开始普遍，**所以建议合作院方提供肿瘤的基因靶向治疗信息**，在后续的肿瘤预后分析中，实验室会对肿瘤基因类型进行进一步的完备分析。

\*注：该规范由自动化所分子影像中科院重点实验室 Radiomics 组完成，如有任何疑问请咨询：董迪（手机：18513001536；13811833760。邮箱：di.dong@ia.ac.cn）。