# 肺炎 CT 影像 AI 标准数据库

依托单位：

吉林大学第一医院

浙江求是数理医学研究院

## 一、数据库名称

肺炎 CT 影像 AI 标准数据库

## 二、数据库建库背景

肺炎是世界范围内发病率和病死率最高的疾病之一，是呼吸系统常见病、多发病。其中社区获得性肺炎（CAP）是全球第六大死因，在全球所有年龄组均有较高的发病率和死亡率。基于胸部 CT 图像采用医学影像大数据人工智能方法在肺部疾病筛查、诊断、和定量评估等方面已经显示出良好的效果。为更好地提高医疗服务质量和降低医生劳动强度，国家相继颁布了多项指导性文件和政策进一步推进医学影像 AI 产品研发。在政府的大力支持下，数十家影像设备厂商、AI 公司相继研发并推出了针对肺炎医学影像人工智能产品，但目前尚缺少对上述肺炎 AI 新技术产品经济有效的评价手段，因此建立肺炎 CT 影像 AI 标准数据库用于验证肺炎 AI 产品的性能，包括肺炎筛查、诊断、定量评估、质量控制等，对促进医学人工智能产品落地推广并服务于医疗健康产业具有重要意义。

## 三、数据库建库过程

数据库数据采集由具有相关资质的放射科技师严格按照肺部 CT 扫描规范采集影像，由具有药物临床试验管理规范（GCP）资格临床实验人员收集数据。数据收集后对其进行数据脱敏，抹除患者信息、医院信息、时间信息。将脱敏后的数据进行质量评估，构建合格及不合格子数据集。按照标注流程，由具有执业医师资格证的不同级别放

射科医生对合格的数据进行标注和审核。根据数据命名原则将数据进行命名，并纳入不同的子数据库。（图 1）

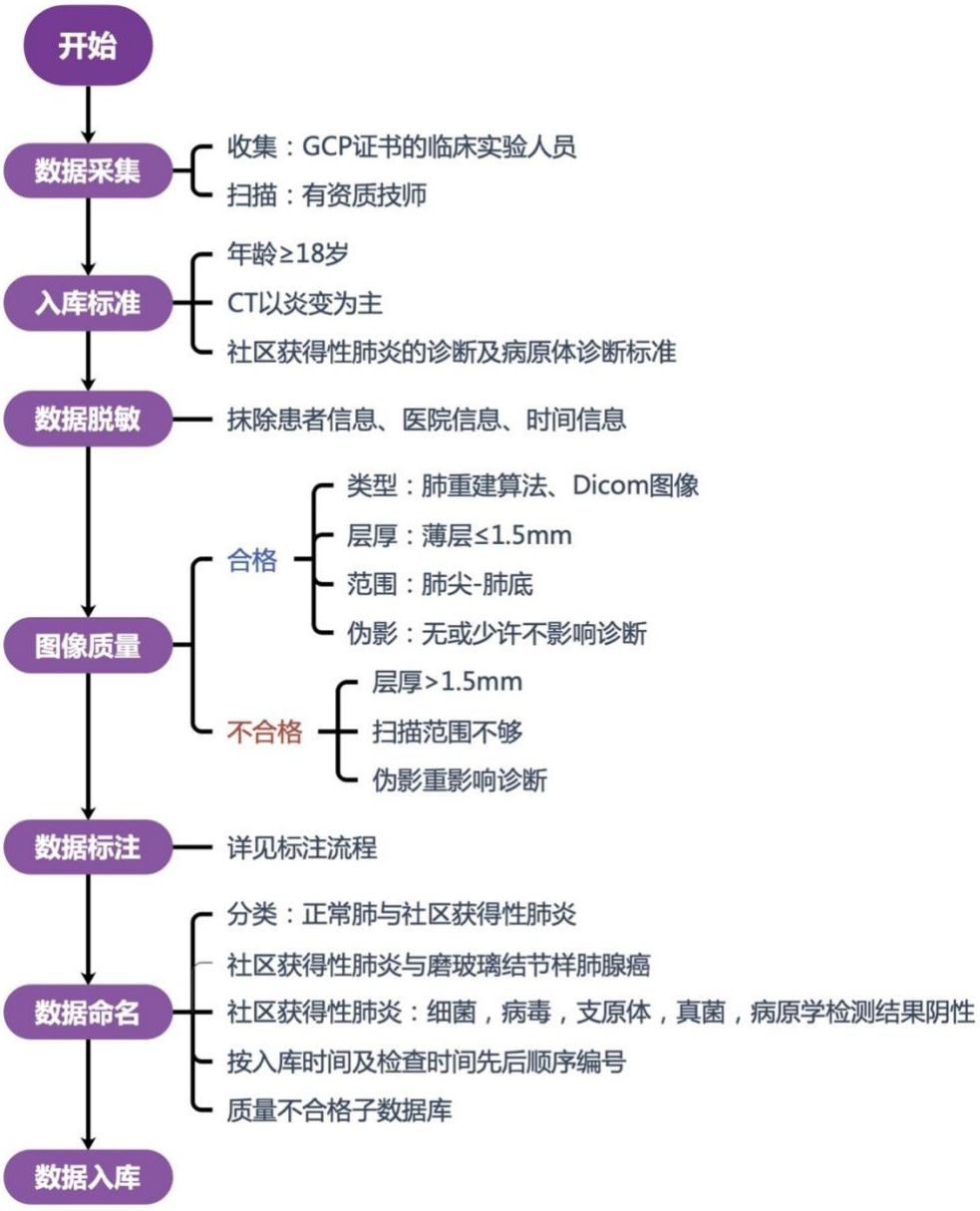


图 1. 数据库建库过程示意图。数据采集、入库标准等制定参照《胸部 CT 扫描规范化专家共识》、《中国成人社区获得性肺炎诊断和治疗指南(2016 年版)》

## 四、数据库数据分布

数据库数据涵盖吉林省、湖北省、浙江省多个省份的胸部CT影像及临床数据，病种覆盖社区获得性肺炎（CAP）、易与肺炎混淆表现为磨玻璃密度结节的肺腺癌和正常肺，总计含2754例数据。社区获得性肺炎均有病原学检查结果，所有磨玻璃密度结节均由病理结果证实为肺腺癌。社区获得性肺炎共2104例CT图像，图像质量合格共1904例（占比90%），约67万余张CT切片，涵盖细菌性肺炎（493例，占比26%），其他病毒性肺炎（197例，占比10%），支原体性肺炎（153例，占比8%）、真菌性肺炎（52例，占比3%）及病原学检查结果阴性（1009例，占比53%）。目前已标注500例社区获得性肺炎影像的炎变区域。此外数据库还包括正常胸部CT平扫图像500例，病理证实为肺腺癌的磨玻璃密度结节150例，包括纯磨玻璃结节97例，亚实性磨玻璃结节53例。

上述 CT 影像数据来自不同供应商，涵盖了目前市场上主流的大部分 CT 设备（见表 1）。

表 1. 肺炎数据库涵盖设备目录

| 规格型号 | 生产厂家 |
| --- | --- |
| Discovery CT750 HD | 美国 GE Medical Systems，LLC |
| 64 层 CT | 德国西门子公司 |
| 双源 CT | 德国西门子公司 |
| 16 层螺旋 CT OMATOM Emotion16 | 西门子 |
| Brilliance iCT | 飞利浦 |

| | |
|---|---|
| Brilliance CT 64 Slice | 飞利浦 |
| Revolution CT | GE Medical Systems，LLC |
| NeuViz 128 | 沈阳东软医疗系统有限公司 |
| NeuViz Prime | 沈阳东软医疗系统有限公司 |

## 五、数据标注及审核

数据的标注包括图像质量判定和病灶勾画，图像质量判定标准见附件《CA40-SOP-04数据标注操作规范》，由通过标注资质认定的标注医师按照标注要求完成病灶的勾画，然后由一名经过标注资质认定的评估医师对标注结果进行评估和修改。

数据库的审核包含内部审核和外部审核，内部审核人员资质为本院放射科副主任医师及以上资质，外部审核人员资质为来自于全国三家三甲医院的具有丰富阅片经验的放射科影像诊断专业副主任医师或主任医师。内外审专家两人一组，每组包括一名内审医师和一名外审医师，共同对评估结果进行审核。把待审核的数据随机分配给内外审专家进行审核，内外审专家对已标注的数据进行审核，若内外审医师审核结果一致，则完成审核；若内外审医师均认为标注数据需要修改，则由内审或外审医师进行修改后完成审核；若内外审医师意见不一致，则引入仲裁专家对图像进行仲裁，根据仲裁结果由内审或外审医师对标注进行修改，经过内审和外审合格的标注数据方可录入数据库。

## 六、 数据库安全管理

肺炎 CT 影像 AI 标准数据库由吉林大学第一医院和浙江求是数理医学研究院联合牵头建立。根据国家药监局发布的《深度学习辅助决策医疗器械软件审评要点》，为加强数据安全管理，以满足测评数据库六个要求即权威性、科学性、规范性、多样性、封闭性和动态性，数据库安全管理遵循专人负责、分层管理、统一标准、全程可控的原则。数据库从计划执行到搭建至今，陆续形成了一套体系文件和记录，31 个文件涵盖质量手册、数据库常用影像术语规范、数据采集操作规范等多项规范和记录。该测评数据库中的所有数据由国家卫生健康委《医学图像数据库》工作组托管，按照国家最高安全等级存储于秦山核电站数据中心。

## 七、 数据库更新管理

数据库每年动态更新不少于300例（约7万张CT切片）社区获得性肺炎数据，其中病原学检测结果阳性占比不少于30%（≥2万张CT切片）。在病原学阳性组中，细菌性肺炎占比不少于30%，病毒性肺炎占比不少于15%，余为其他病原体所致肺炎，包括支原体、衣原体、真菌、军团菌等。更新数据按图像质量评价标准分别纳入合格与不合格子数据库。此外，每年新增标注数量不少于动态更新数据总量的25%（约2万张CT切片），标注内容包括肺部炎变区域、五个肺叶及全肺轮廓，以便评估各肺叶炎变区域/相应肺叶体积、全肺炎变区域/全肺体积的比值，从而监测肺炎在疾病演变以及治疗进程的动态变化。

## 八、测试使用场景

数据库可根据 AI 产品的不同预期用途，调取相应的数据集，满足以下测试需求：

1）正常肺与社区获得性肺炎。

2）社区获得性肺炎与磨玻璃结节样肺腺癌

3）社区获得性肺炎涵盖细菌，病毒，支原体，真菌多种病原体所致肺炎。

4）图像质量良好 / 质量差。

## 九、数据库管理文件

数据库从计划执行到搭建至今，陆续形成了一套体系文件和记录，文件列表如下：

| 文件名编号 | 文件名称 | 文件编号 | 文件名称 |
|---|---|---|---|
| CA40-QMS-00 | 质量手册 | - | 伦理审查批件 |
| CA40-SOP-01 | 数据库常用影像术语规范 | CA40-R-01 | 数据库医学术语与定义 |
| - | - | CA40-R-02 | 数据库建库需求分析 |
| CA40-SOP-02 | 数据采集操作规范 | CA40-R-03 | 数据接收表 |
| | | CA40-R-04 | 数据接收明细表 |
| | | CA40-R-05 | 数据质评表 |
| | | CA40-R-06 | 数据采集设备清单 |
| CA40-SOP-03 | 数据脱敏操作规程 | CA40-R-07 | 数据脱敏操作记录 |
| CA40-SOP-04<br>CA40-SOP-05 | 数据标注操作规范<br>标注培训教材 | CA40-R-08 | 数据标注培训记录 |
| | | CA40-R-09 | 标注人员档案 |
| | | CA40-R-10 | 标注数据记录 |

| | | | |
|---|---|---|---|
| | | CA40-R-11 | 仲裁审核记录 |
| | | CA40-R-12 | 标注质量评估记录 |
| | | CA40-R-13 | 标注质量报告 |
| | | CA40-R-14 | 标注数据库统计表 |
| CA40-SOP-06 | 数据安全控制规范 | CA40-R-15 | 数据备份记录 |
| CA40-SOP-07 | 数据库管理规范 | CA40-R-16 | 数据库管理与动态更新记录 |
| CA40-SOP-08 | 数据库审核规范 | CA40-R-17 | 内部审核记录 |
| | | CA40-R-18 | 内部审核报告 |
| | | CA40-R-19 | 外部审核记录 |
| | | CA40-R-20 | 外部审核报告 |

## 十、其他

欢迎您对肺炎 CT 影像 AI 标准数据库提出宝贵意见，以便于我们持续性改进和提升。

联系电话：043188782861

单位：吉林大学第一医院

邮箱：huimao@jlu.edu.cn