赛题六 构建单病种电子病历结构化数据库

--以急性心肌梗死单病种为例

### ****【赛题数据】****

#### 【数据示例】

比赛训练数据集分为两部分，包括数据输入部分和数据输出部分，以下是10个样例的数据对应文件名、数据格式和主要内容：

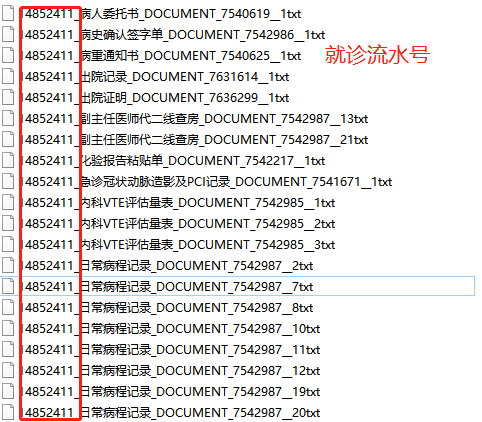
数据输入部分是电子病历文件（以TXT格式存储，见4-病历文书文件夹包含入院记录、出院记录、手术记录等相关电子病历文书）和医嘱（excel格式，见5-医嘱.xlsx），以及超声心动检查（excel格式,见6-超声心动检查.xlsx）和其他检查结果（excel格式,见7-其他检查结果.xlsx）。

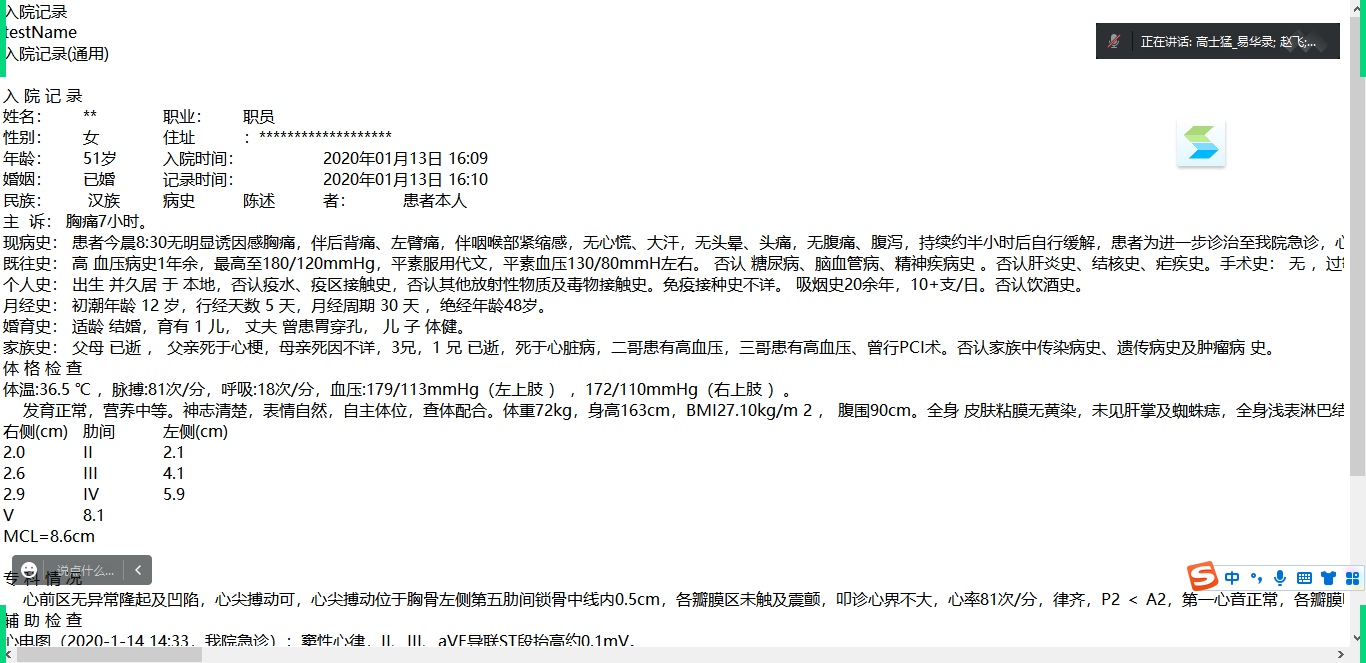
数据输出部分是急性心肌梗死单病种数据库（CSV格式，见8-输出结果.CSV）。

患者的就诊流水号是所有数据的关联id。

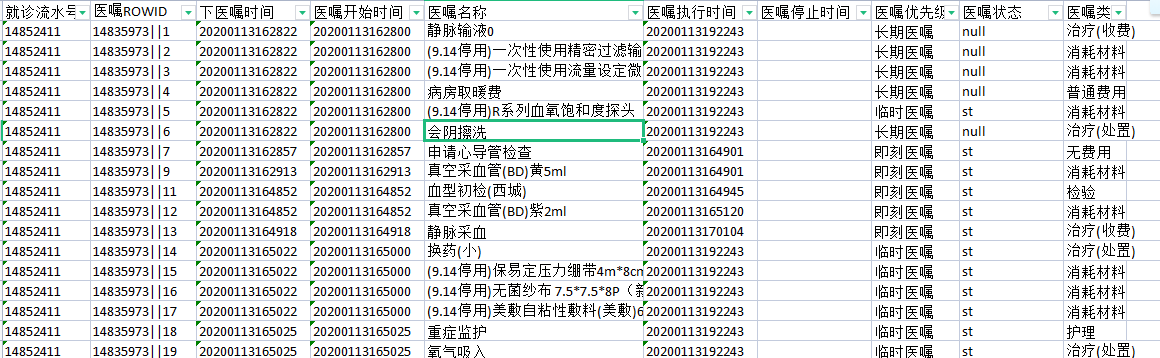
**输入部分数据示例：**

1、电子病历文件（以入院记录为例）

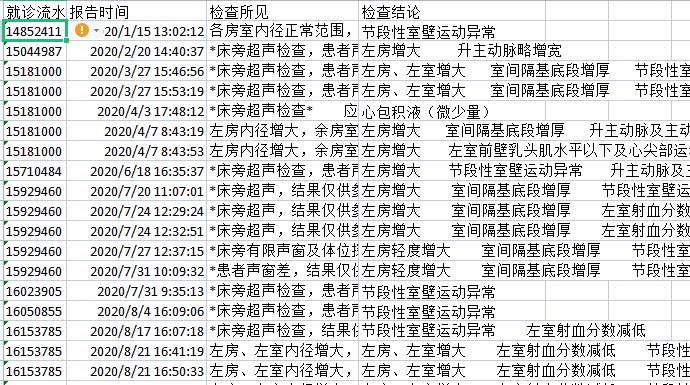




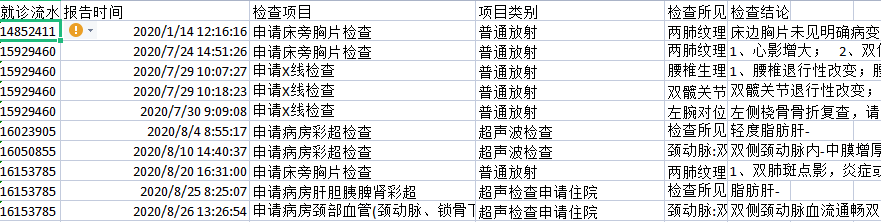
2、医嘱.xlsx



**3、超声心动检查.xlsx**



4、其他检查结果.xlsx



**输出部分数据（心肌梗死单病种数据库）示例：**

**1、输出问题举例**

预测问题举例: "某患者急性心肌梗死主要病变血管"

备选答案选项："a:LAD-左前降支",

"b:LCX-左回旋支",

"c:RCA-右主干",

"d:LM-左主干",

"UTD-无阳性病变"

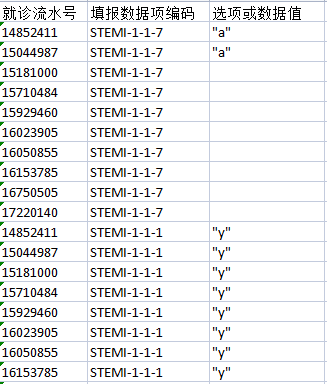
假设该患者病历中有如下描述 "行冠脉造影检查：冠状动脉粥样硬化性心脏病，三支病变（累及LAD、LCX、RCA），并于LCX置入1枚支架，RCA置入2枚支架。"

根据患者病历信息应获得问题的答案为字符串，多选选项以"\"隔开,答案示例"a\b\c"。

**具体问题以及选项说明，见附件2-STEMI数据项说明.xlsx。**

1. 输出文档（CSV）

包含患者的就诊流水号、填报数据项编码以及预测的选项或数据值。8-填报结果.xlsx



#### 【任务说明】

1. 电子病历文书（txt），医嘱.xlsx,超声心动检查.xlsx,其他检查结果.xlsx,填报结果.xlsx通过患者的就诊流水号进行关联
2. 允许使用外部数据；若使用外部数据，请在算法说明文档中对数据来源、数据特点等信息进行描述
3. （输出）识别结果与真实值必须完全一致，才算正确，不选、多选、少选都算错误
4. （输入）电子病历文书（txt），医嘱.xlsx,超声心动检查.xlsx,其他检查结果.xlsx
5. （输出）识别结果：要求参赛团队将所有识别结果整理为⼀份CSV文件，其中包含患者的就诊流水号、填报数据项编码以及预测的选项或数据值

#### 【数据文件结构】

我们将提供相关数据，具体文件夹结构如下：

|--train\_data

|--电子病历文件夹

|—各类电子病历文书（入院记录，出院记录等）

|--医嘱.xlsx

|--超声心动检查.xlsx

|--其他检查.xlsx

|--填报结果.CSV

|--test\_data

|--电子病历文件夹

|—各类电子病历文书（入院记录，出院记录等）

|--医嘱.xlsx

|--超声心动检查.xlsx

|--其他检查.xlsx

train\_data文件夹存放训练数据。

test\_data 文件夹存放测试数据。