|  |  |
| --- | --- |
| ICS | 11.040.99 |
| CCS | |  | | --- | |  |   C30 |

团体标准

T/CSBME 050—2022

人工智能辅助脓毒症预警数据标注和质控标准

Artificial intelligence assisted sepsis early warning data annotation and quality control standards

2022 - 01 - 15发布

2022 - 03 - 01实施

四川省生物信息学学会  发布

目次

[前言 II](#_Toc93077075)

[引言 III](#_Toc93077076)

[1 范围 1](#_Toc93077077)

[2 规范性引用文件 1](#_Toc93077078)

[3 术语和定义 1](#_Toc93077079)

[4 缩略语 2](#_Toc93077080)

[5 数据集总要求 2](#_Toc93077081)

[6 数据采集 3](#_Toc93077082)

[7 数据整理 4](#_Toc93077083)

[8 数据标注 4](#_Toc93077084)

[9 数据集构建 5](#_Toc93077085)

[附录A （规范性） 数据集分类及医学和风险级别定义 6](#_Toc93077086)

[附录B （规范性） 液基制片质量评估 7](#_Toc93077087)

[参考文献 8](#_Toc93077088)

1. 前言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由四川省生物信息学学会提出。

本文件由四川省生物信息学学会知识产权与标准化工作委员会归口。

本文件起草单位：四川大学华西医院

本文件主要起草人：康焰

1. 引言

尽管近年来国际上针对脓毒症(sepsis)做了大量的基础研究工作，并进行了有效临床干预措施的研发与推广，但遗憾的是，无论在发达国家还是发展中国家，脓毒症仍以其高发病率和病死率严重威胁着人类的健康及生命安全，消耗了大量的医疗资源。据不完全统计，全球每年有2-3千万患者罹患脓毒症，每天约有1400人死于脓毒症。因此，提高对脓毒症的认识，正确面对脓毒症所带来的巨大挑战，对改善脓毒症预后意义重大

脓毒症诊治难点在于疾病的异质性，即不同的脓毒症患者可表现为不同的临床表现以及对某种治疗的不同反应，这给临床诊疗和研究带来巨大挑战。脓毒症发生发展机制较为复杂，特别是急危重脓毒症患者，其临床诊治过程中会产生大量的数据，包括生命体征、呼吸力学、体液样本检验、影像学等数据，而人类直觉对于这些数据并不能很好地整合，因此可能无法及时发现潜在的风险。人工智能是用来研究、开发用于模拟、延伸和扩展人类智能的理论、方法、技术的一门新的技术科学。目前，人工智能已经在各行各业发挥了重要作用。在医学方面，人工智能技术在影像诊断、病理诊断中表现出了超过高年资医生的诊断水平，因此具有重要的应用前景。人工智能技术在脓毒症患者诊疗方面也具有较好的应用前景，虽然已经有大量文献报道应用人工智能技术来辅助脓毒症的诊断与治疗，但临床医师对于这类技术缺乏深刻认识。

本数据标准规范以脓毒症3.0标准作为主要参考依据，旨在形成共识性的脓毒症诊疗领域数据集标准体系，为医疗数据的处理提供一致性的行业标准。。

人工智能辅助脓毒症预警数据标注和质控标准

* 1. 范围

本文件主要规定了人工智能辅助脓毒症预警所使用的脓毒症数据集在数据采集、数据整理、数据标注、数据集构建过程中所涉及的相关要求。脓毒症数据集主要包含三大类，包含“观察记录”、“医嘱信息”、“医学检验项目”

本文件适用于人工智能辅助脓毒症预警的研发、生产、检测过程中的数据集建立、评估，以及脓毒症的诊断和风险评估。

* 1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

YY/T 1833.1 人工智能医疗器械 质量要求和评价 第 1 部分：术语

YY/T 1833.2 人工智能医疗器械 质量要求和评价 第 2 部分：数据集通用要求

* 1. 术语和定义

2014年1月，ESICM指定19位国际专家组成专家组启动了脓毒症3.0的修订，经过系统文献分析、Delphi调查法，以及脓毒症患者数据库的验证，将脓毒症定义为宿主对感染的反应失调而致的危及生命的器官功能障碍。脓毒性休克是脓毒症的一个亚型，是指脓毒症所致严重的循环障碍和细胞代谢异常，并足以使病死率显著增加。

脓毒症3.0强调致命性的器官功能障碍，专家组分析了SIRS、序贯器官衰竭评分(sequential organ failure assessment，SOFA)、Logistic器官功能障碍评分系统与感染性疾病预后的相关性；结果显示：与SOFA<2分比较，SOFA≥2分患者死亡风险增加2～25倍，同时，SOFA比Logistic器官功能障碍评分简单。因此，工作组推荐将SOFA≥2分作为脓毒症3.0的器官功能障碍标准，即脓毒症3.0的诊断标准为：感染＋SOFA≥2分。另外，符合SIRS标准，未出现器官功能障碍的感染患者将不被诊断为脓毒症。

* 1. 电子病历数据标注和质控标准
     1. 数据来源

数据集病例至少来自四川大学华西医院的六个护理单元，包括第11护理单元、第13护理单元、第15护理单元、第72护理单元、第73护理单元、温江第1护理单元。

* + - 1. 标注人员遴选

经过医院和相关部门同意的人选

* + - 1. 培训

信息部门或者大数据其他主管部门的培训，主要包含用户授权、软件安装、数据加密和反加密工具使用、数据质量监控、数据接入监控、数据统计工具使用、数据导出管理、标注内容培训等。

培训由信息部门或者大数据平台主管部门组织，由数据采集方进行，培训前应准备培训环境和培训资料。

建议入选的标注医师接受相关培训，同意对标注规则的认识，熟悉标注软件操作。

* + - 1. 考核

通过制定相关的考核制度进行考核，内容包括数据安全管理制度、数据使用管理规范、使用违规惩罚制度、标注规则等四部分：

1. 数据安全管理制度： 数据安全管理从系统安全和审批管理两方面进行考虑。系统安全包括信息部门数据备份策略、用户权限、账户弱密码、账户有效期和数据导出管理等方面进行管理；审批管理主要指账户权限审批流程和数据申请审批流程。
2. 数据使用管理规范：数据使用人需严格遵守国家有关法律法规，采集数据时需注明申请数据范围、用途、使用事件等重要信息，对使用数据有保密责任，数据采集结束要立即向管理部门报备，不可随意使用数据或进行其他违规操作。
3. 使用违规惩罚制度：对于擅自传播、转让、更改数据用途、未经许可使用数据或其他违规行为，医院管理部门有权利追究法律责任。
4. 标注规范。

基础设施管理

* + - 1. 标注场所

在医院内部网络对数据进行管理，不允许连接互联网。

* + - 1. 标注软件

使用稳定的专业医学标注软件进行标注工作，可灵活配置标注参数，以不断优化标注过程。

数据标注标准

* + - 1. 标注人员

人员组成：为提高标注的客观性和一致性，标注工作需要3名主治以上级别的医师参加。每一例疗效评价任务由两名医师达成一致意见才能确定，否则由一名仲裁专家介入、进行疗效评定。仲裁专家由工作经验15年以上的副主任医师或主任医师担任。

* + - 1. 标注流程

电子病历中的标注点主要有出院小结。

* + - 1. 临床诊疗规范

参照脓毒症3.0诊断标准对脓毒症进行诊断。

* + - 1. 标注细则

采用自然语言处理，对电子病历中出院小结进行加工和提取。

* + - * 1. 分歧处理

当出现标注医师的标注结果有标准容许以外的差异时，解决方法为，由专家组进行仲裁判定。

* + - * 1. 可追溯性

为满足数据后期与其他临床资料可对照及后续跟进研究的需要，保证病例可追溯，故在数据脱敏前对所提取的患者信息进行加密备份一份，用于后续追溯使用。

标注质量评估标准

* + - 1. 人员

仲裁专家由从事重症医学领域工作经验15年以上的副主任医师或主任医师担任。

* + - 1. 方法

根据电子病历中出院小结，采用多人盲标+分阶段审核的方法进行。

* + - 1. 指标

针对脓毒症早期诊断，普通病历的评价由3名标注医师执行，由1名审核专家进行审核。遇到复杂征象时，可酌情增加审核人数，以保证标注质量，标注医师和仲裁专家可以手动修改。

* + - 1. 通过准则

以下8个质量维度进行整体评估：

1. 完整性：数据项内容的完整情况，主要考察字段数据是否存在，NULL等同于空。
2. 规范性：数据项的数据格式，数据长度，时间范围，数值范围是否正确。
3. 标准性：数据项的数据是否与字典一致，如性别的取值应该是男，女。
4. 正确性：数据项内容是否正确，如入院记录中月经史患者性别不应该为男。
5. 一致性：数据项间内容是否一致，如考察病案首页上的患者、手术、诊断、费用等信息是否同HIS或其他系统中的信息一致。
6. 时序性：时间逻辑合理性（流程时间分布），如出院时间需要晚于入院时间。
7. 整合性：相关系统对应数据项目可对照或关联，如医疗费用明细表里的医嘱标识应该在中西医处方表、药品类医嘱表、草药处方表或者非药品医嘱表里。
8. 唯一性：记录的唯一性，如Patient1D在patient表不能重复，相同的医嘱不应该重复。

数据验证方法：

1. 采集方提供数据报告，包括接入数据范围（系统数量）、数据量（需要分类，比如患者、医嘱、处方、检验检查数量等）、数据纵深（起始到终止日期）。医院信息部门工程师可进行宏观数据核验。
2. 数据质量抽样检查：比照HIS、EMR等业务系统的单个患者进行数据核验。
3. 病案数据抽样检查：随机抽取一定数量的病例数据，跟院内数据进行核对。
4. 隐私数据检查：查看是否按照要求进行去隐私化处理，比如患者姓名、地址、身份证号等，可有意识的搜索隐私数据看可否能检索到。

标注部分评估：

检出环节：3名标注医师背靠背独立标注，然后用计算机自动判断检出的一致性，以所有人标注结果的并集作为结果。

1. 审核环节：由其他标注组长和仲裁专家各自独立对检出和分类结果进行审核与修改，纠正漏诊、误诊和误判。
2. 疗效评价环节：在疗效评价与分类完成之后，由于疗效评价相对简单，建议普通病例的疗效评价由1名标注医师执行，由1名审核专家进行审核。遇到复杂病历时，可酌情增加审核人数，以保证标注质量。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **编号** | **数据域** | **数据项名称** | **值** | **说明** |
| 1 | 标注人员 | 姓名 |  |  |
| 2 | 医院 |  |  |
| 3 | 年资 | 5年 |  |
| 4 | 培训状态 | 是/否 |  |
| 5 | 考核状态 | 是/否 |  |
| 6 | 基础设施 | 标注场所 |  |  |
| 7 | 标注软件名称 |  |  |
| 8 | 标注软件版本 |  |  |
| 9 | 标注指标 | 出院小结 |  |  |
| 11 | 质控人员 | 质控方法 |  |  |
| 12 | 医院 |  |  |
| 13 | 年资 | 15年 |  |
| 14 | 培训状态 | 是/否 |  |
| 15 | 质控方法 | 质控流程 | 疗效判断 |  |
| 16 | 疗效判断质控 |  |  |