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

|  |
| --- |
|  |

C30 |

团体标准

T/CSBME 050—2022

继发性高血压人工智能辅助风险识别数据标注和质控标准

Artificial intelligence-assisted risk identification data labeling and quality control standards for secondary hypertension

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 数据标注与质控 4](#_Toc93077084)

[6 数据集构建 4](#_Toc93077084)

[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：曾锐、李东泽、贾禹、李芳卉、李文韬、童瑶、公奥博。

1. 引言

高血压是心血管病最为常见，也是临床常见的慢性非传染性疾病之一。高血压带来严重的疾病负担。其中，继发性高血压占高血压患者的5%-15%。继发性高血压除高血压本身对机体的影响外，与之伴随的内分泌紊乱、低钾血症、肾功能不全、低氧血症等还可导致独立于血压之外的心脑血管损害，因此早识别、早诊断、早治疗尤为重要，然而目前对继发性高血压的诊断存在一定的困难。本数据标准规范以《中国高血压健康管理规范（2019）》作为主要参考依据。继发性高血压主要包括肾实质性高血压、肾血管性高血压、库欣综合征、原发性醛固酮增多症、阻塞性睡眠呼吸暂停综合征、嗜铬细胞瘤等。

本继发性高血压数据标准规范旨在形成共识性的继发性高血压领域数据集标准体系，为医疗数据的采集、标注和质控等提供一致性的行业标准，实现同质化的数据存储和使用，并以高质量的数据库支撑包括临床医学研究及人工智能医疗器械等研发及应用的需求。

继发性高血压人工智能辅助风险识别数据标注和质控标准

* 1. 范围

本文件主要规定了继发性高血压人工智能辅助风险识别所使用的高血压数据集（包裹基本信息、体格检查、病史、实验室检查等数据）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。

本文件适用于继发性高血压人工智能辅助风险识别的研发、生产、检测过程中的数据集建立及其评估，以及继发性高血压的初步筛查与风险评估。

* 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 医学与医疗器械 质量要求和评价 第 1 部分：术语

YY/T 1833.2 医学与医疗器械 质量要求和评价 第 2 部分：数据集通用要求

* 1. 术语和定义

下列术语和定义适用于本文件。

术语

（1）人工智能医疗器械 artificial intelligency medical device; AIMD

 采用AI技术实现其预期用途的医疗器械

注1：如采用机器学习、模式识别、规则推理等技术实现其用途的独立软件。

注2：如采用内嵌AI算法、AI芯片实现其医疗用途的医疗器械。

[YY/T 133.1-2022]

（2）继发性高血压风险识别人工智能辅助诊断技术

继发性高血压风险识别人工智能辅助诊断技术是指医务人员采用人工智能(AI)技术用于辅助继发性高血压风险识别的一项医疗技术。该技术基于患者基本特征、生命指征、血压波动特征、常规血液学检查结果（包括血常规、生化1、生化4等）此类易于采集的数据，协助临床医生尤其是基层医院医生，在患者首次确诊高血压后进行继发性高血压风险识别，筛选出继发性高血压高风险患者，建议进一步进行继发性高血压病因诊断。

（3）继发性高血压

继发性高血压是病因明确的高血压，当查出病因，并有效去除或控制病因后，作为继发症状的高血压可被治愈或明显缓解。对继发性高血压的筛查、鉴别成为高血压诊断评估和治疗的重要内容。临床上，当患者有相应临床表现、体格检查线索，或是难治性高血压时，建议筛查、排除继发性高血压。

本AI产品采集患者基本特征、生命指征、血压波动特征、常规血液学检查结果，并综合考虑以下因素： （1）血压中、重度升高的年轻高血压患者；（2）舒张压高于100 mmHg 的老年高血压患者（年龄≥65岁）；（3）症状、体征或实验室检查有怀疑线索，例如肢体脉搏搏动不对称性减弱或缺失，腹部听到粗糙的血管杂音，既往患有肾脏疾病史等；（4）不明原因的高血压伴低血钾。基于先前构建的继发性高血压风险识别算法，筛选出继发性高血压高、中、低危患者。

定义

(1)血压：血压（blood pressure，BP）是指血液在血管内流动时作用于单位面积血管壁的侧压力，它是推动血液在血管内流动的动力。在不同血管内被分别称为动脉血压、毛细血管压和静脉血压，通常所说的血压是指体循环的动脉血压。

(2)偶测血压：被测者在没有任何准备的情况下测得的血压。

(3)动态血压：使用动态血压记录仪测定一个人昼夜24小时内，每间隔一定时间内的血压值。动态血压包括收缩压、舒张压、平均动脉压、心率以及它们的最高值和最低值等项目。

(4)高血压：动脉血压超过正常值的异常升高。本指南定义为收缩压≥140mmHg和(或)舒张压~>90mmHg。

(5)收缩压：心室收缩时，主动脉压急剧升高，在收缩期的中期达到最高值，这时的动脉血压值称为收缩压，也称为“高压”。

(6)舒张压：心室舒张时，主动脉压下降，在心舒末期动脉血压的最低值称为舒张压，也称为“低压”。

(7)脉压：收缩压减舒张压的差值。

(8)平均动脉压：一个心动周期中每一瞬间动脉血压的平均值，大约等于舒张压加1／3脉压。

(9)KPa：千帕，通常用于表示血压数值。

(10)mmHg：毫米汞柱，人们用水银血压计来测量血压时用水银柱的高度“毫米汞柱”来表示血压的水平。

1mmHg(毫米汞柱)=0．133kPa(千帕斯卡)

7．5mmHg(毫米汞柱)=1kPa(千帕斯卡)

(11)理想血压：收缩压<120ramHg和舒张压<80mmHg。

(12)正常血压：收缩压应<130mmHg，舒张压<85mmHg。

(13)血压正常高限或高血压前期：收缩压在130～139mmHg和／或舒张压在85～89mmHg。

(14)临界高血压：收缩压在140～160mmHg(18．6～21．3kPa)，舒张压在90～95mmHg(12．0～12．6kPa)。

(15)肌酐：人体内肌酸的代谢产物。肌酸量与肌肉容量呈正比，以稳定的速度产生，并以稳定的速度向血液中释放肌酐，由血循环带到肾脏，从尿中排出体外。长期高血压损害肾功能，使肾脏对血中肌酐的排出减少，使血肌酐升高。

(16)心脏的超声波检查：又叫超声心动图，可以反映出各瓣膜的形态、活动是否正常，各心房心室的大小及形态，心脏周围出入心脏的大血管的情况，有无先天性心脏病或风湿性心脏病造成的心脏损害、心包积液、各种心肌病，检出高血压所致的左心室肥厚和评价心功能。

(17)原发性高血压：原因不明的高血压，占90％以上，目前尚难根治但能被控制。也叫高血压病。

(18)继发性高血压：也叫症状性高血压。由于某些疾病(肾脏疾病、内分泌疾病如肾上腺肿瘤或增生和其他原因所致)在其发展过程中产生的，原发疾病治愈后，血压也会随之下降，占高血压患者的5％～10％。

(19)肾血管性高血压：指各种原因引起的肾动脉或其主要分支的狭窄或闭塞性疾病，引起肾血流量减少或缺血所致的高血压。

(20)肾实质性高血压：肾脏实质疾病，如急、慢性肾小球肾炎，肾盂肾炎，多囊肾等引起的高血压。

(21)库欣综合征：是由于长时间接触皮质醇或其分泌量过高而造成的一系列临床表现与体征。其临床表现与体征可能包括：高血压、向心型肥胖（肚腩有赘肉但双臂与双腿较细）、紫纹（红色或紫色的短纹）、满月脸（脸又圆又红）、脂失养（水牛肩，两肩中间有一较大肿块）、肌肉无力、骨质疏松、粉刺、皮肤脆弱难以愈合等。

(22)原发性醛固酮增多症：是由于肾上腺的皮质肿瘤或增生，醛固酮分泌增多所致。原醛症是一种继发性高血压症，约占高血压症中0.4～2%。发病年龄高峰为30～50岁，女性较男性多见。引起本病最常见的原因为醛固酮瘤，约占原醛症的60%～80%，大多数为单个腺瘤，左侧多见。其次有双侧肾上腺皮质增生，又称为特发性醛固酮增多症，约占20%～30%。少见原因有地塞米松可抑制性醛固酮 增多症，醛固酮癌、异位分泌醛固酮的肿瘤。

(23)阻塞性睡眠呼吸暂停综合征：是一种病因不明的睡眠呼吸疾病，临床表现有夜间睡眠打鼾伴呼吸暂停和白天嗜睡。由于呼吸暂停引起反复发作的夜间低氧和高碳酸血症，可导致高血压，冠心病，糖尿病和脑血管疾病等并发症及交通事故，甚至出现夜间猝死。

(24)嗜铬细胞瘤：为起源于神经外胚层嗜铬组织的肿瘤，主要分泌儿茶酚胺，根据肿瘤是来自交感神经或副交感神经将副神经节瘤分为副交感神经副神经节瘤（包括化学感受器瘤、颈动脉体瘤等）及交感神经副神经节瘤（包括腹膜后、盆腔及纵隔后的副神经节瘤）。某些患者可因长期高血压致严重的心、脑、肾损害或因突发严重高血压而导致危象，危及生命，但如能及时、早期获得诊断和治疗，是一种可治愈的继发性高血压病。

* 1. 缩略语

以下缩略语适用于本文件：

AI：人工智能（Artificial Intelligence）

AIMD：人工智能医疗器械（Artificial intelligency medical device）

SHT：继发性高血压（Secondary hypertension）

BP：血压（Blood pressure）

* 1. 数据标注与质控
		1. 数据来源

数据来源于四川大学华西医院与2011-2021年住院的诊断包含“高血压”的患者

* + 1. 标注人员

标注医师、复核医师、仲裁专家应满足以下要求。

1. 执业资质：具有病理诊断执业医师资格；
2. 参与标注前，应统一开展必要的培训，培训内容包括标注流程、术语定义、样本总体描述、相关案例等，所用案例应独立于待标注样本。培训后宜设置标注医师培训合格的测试及接受标准，宜选择与标准判读结果吻合率在90%以上的医师作为标注医师；
3. 复核医师、仲裁专家一般应由数据集建设实施单位聘请行业内知名专家担任。
	* 1. 标注原则

保证标注类别一致性。标注员在标注过程中应遵守医学伦理，确保患者隐私和数据保密。每数据例应由至少2位标注医师标注。依据标注区域的位置和类别判断标注一致性，若标注结果一致，数据方可入组；若结果不一致，数据应再由复核医师进行标注，复核结果若与任何一位医师标注结果相同，数据方可入组。若与原始标注任何一方都不相同，应再由专家进行仲裁，以仲裁结果作为最终结果入组。

* + 1. 标注方式

1. 对于一个病历资料，标注员应根据病史和检查结果，标注所有与继发性高血压相关的ICD编码。

2. 对于同一份病历资料，标注员应根据临床常识和病史分析，排除非继发性高血压的诊断，只选择与继发性高血压相关的ICD编码进行标注。

3. 所有标注结果应经过初步的质量控制，排除标注错误和遗漏。

4. 所有标注结果应经过二次审核，确保数据的准确性和一致性。

5. 对标注数据进行质量评估，评估项包括但不限于：数据完整性、准确性和一致性等。

6. 对标注数据进行统计分析，确定数据的分布情况、特征等，为后续建模和分析提供支持。

7. 对数据进行可视化，发现数据的异常情况和规律，为后续分析提供支持。

* + 1. 标注流程

1. 标注员应对所接收的数据进行审查和检查，确保数据的完整性和准确性。

2. 标注员应根据所分配的任务，对继发性高血压患者的临床资料进行标注，包括但不限于：患者基本信息、病史、生化指标、影像学检查等。

3. 标注员应根据继发性高血压患者的特点和临床表现，选择正确的标签进行标注。标签包括但不限于：高血压程度、病因、并发症、治疗方案、病历诊断记录所对应ICD编码等。

5.5 数据管理标准

5.5.1 人员管理

1. 选拔

符合数据管理标准，且经过医院信息科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。

1. 考核

通过制定相关的考核机制进行考核，内容包括如下部分：

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。

5.5.2 基础设施管理

1. 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. 软件

高血压标准数据库平台软件。

5.5.3 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，故在数据脱敏之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。

5.4 质量评估标准

1. 人员

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。

2.方法

分阶段审核。

1. 指标

数据是否按照要求进行标准化和正确性等验证。

1. 通过准则

以下8个质量维度进行评估通过：

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。

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。

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。

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。

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。

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。

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。

唯一性：记录的唯一性，如PatientID在Patient表不能重复，相同的医嘱不 应该重复。

5.5 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。

1.

参考文献

[1]国家心血管病中心. 中国心血管病报告2018 [M]. 北京: 中国大百科全书出版社,2019.

[2]国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点[Z], 2019.

[3]Wang Z, Chen Z, Zhang L, et al. Status of hypertension in China: results from the China hypertension survey, 2012‐2015[J]. Circulation, 2018, 137(22): 2344‐2356.

[4]Su M, Zhang Q, Bai X, et al. Availability, cost, and prescription patterns of antihypertensive medications in primary health care in China: a nationwide cross‐sectional survey [J]. Lancet, 2017,39

[5]刘明波，李镒冲，刘世炜，等．2010年中国人群高血压疾病负担[J].中华流行病学杂志,2014,35(6):680-683.

[6]中国高血压防治指南修订委员会．中国高血压防治指南（2018年修订版）[J]中国心血管杂志，2019,24(1):24-56.

[7]Sandström YK, Ljunggren G, Wändell P, et al. Psychiatric comorbidities in patients with hypertension‐‐a study of registered diagnoses 2009‐2013 in the total population in Stockholm County,Sweden[J]. J Hypertens, 2016, 34(3): 414‐420; discussion 420.

[8]Bautista LE, Vera‐Cala LM, Colombo C, et al. Symptoms of depression and anxiety and adherence to antihypertensive medication[J]. Am J Hypertens, 2012, 25(4): 505‐511.

[9]Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans[J]. JAMA, 2018, 320(19): 2020‐2028.

[10]国家药品监督管理局医疗器械技术审评中心. 人工智能医疗器械注册审查指导原则（征求意见稿）[Z], 2021.

[11]中国食品药品检定研究院. 人工智能医疗器械质量要求和评价 第1部分：术语(报批稿)，2020.

[12]中国食品药品检定研究院. 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(报批稿)，2020.

[13]人工智能医疗器械质量要求和评价 第3部分：数据标注通用要求(征求意见稿)