

中华人民共和国医药行业标准

YY/T ×××××—××××

人工智能医疗器械 脑机接口脑电数据专用 要求

Artificial intelligence medical device—Special requirements for EEG data used in
brain computer interface

（工作组讨论稿）

（2024 年 9 月）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

××××—××—××发布

××××—××—××实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前 言 IV

引 言 V

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

 3.1 1

 3.2 1

 3.3 1

 3.4 1

 3.5 1

 3.6 2

 3.7 2

 3.8 2

 3.9 2

 3.10 2

 3.11 2

 3.12 2

 3.13..... 2

 3.14 2

 3.15 2

4 数据采集要求..... 2

 4.1 电极布局要求 2

 4.2 接触阻抗要求 3

 4.3 背景脑电要求 3

 4.4 脑机接口系统噪声要求 3

 4.5 采样率要求 3

 4.6 信号分辨率 4

 4.7 滤波器设置 4

 4.8 信号分辨率 4

 4.9 信噪比 4

 4.10 刺激数据 4

 4.11 事件标记 4

 4.12 其他记录数据 4

 4.13 数据标签要求..... 4

5 数据传输要求..... 5

 5.1 脑电数据传输时间延迟 5

 5.2 丢包率 5

 5.3 传输稳定性信息 5

 5.4 传输实时性信息 6

5.5 传输同步性信息	6
5.6 传输速率	6
6 数据存储要求	6
6.1 数据存储格式	6
6.2 匿名化和去隐私化	7
6.3 存储管理	7
6.4 数据安全	7
6.5 数据版本	7
6.6 数据压缩	7
6.7 数据存储信息	7
6.8 数据命名规范	7
6.9 数据备份	7
7 数据使用要求	7
7.1 数据访问与权限	7
7.2 数据处理	8
7.3 数据评估指标计算	8
8 数据集质量要求	11
8.1 完整性	11
8.2 唯一性	11
8.3 一致性	12
8.4 确实性	12
8.5 时效性	12
8.6 可访问性	12
8.7 依从性	12
8.8 资源利用性	12
8.9 可追溯性	12
8.10 可理解性	13
8.11 可得性	13
8.12 可移植性	13
8.13 可恢复性	13
8.14 代表性	13
8.15 评价方法	13
9 数据集说明要求	13
9.1 实验范式信息	13
9.2 设备配置信息	14
9.3 实验对象信息	14
9.4 数据采集信息	15
9.5 数据传输信息	16
9.6 数据存储信息	16
9.7 数据使用说明	16
9.8 数据集描述	17
9.9 数据集质量特性描述	17
10 数据处理	20

10.1 数据脱敏处理	20
10.2 数据去噪处理	20
10.3 数据滤波处理	20
10.4 数据标注	21
10.5 数据对齐处理	21
10.6 数据筛选	21
10.7 数据分段	21
参考文献	22

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。

本标准由

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。

本标准起草单位:

本标准主要起草人:

引 言

人工智能医疗器械 脑机接口脑电数据专用要求

1 范围

本标准规定了脑机接口脑电数据采集、存储、使用的要求和评价方法。
本文件适用于脑机接口研发、测试、临床试验等环节数据流的质量评价。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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。

GB 9706.1—2020 医用电气设备 第1部分：基本安全和基本性能的通用要求

GB 9706.226—2021 医用电气设备 脑电图机的基本安全和基本性能专用要求

GB 9706.227—2021 医用电气设备 心电监护设备的基本安全和基本性能专用要求

GB 9706.225—2021 医用电气设备 心电图机的基本安全和基本性能专用要求

GB/T 30524—2014 科技平台 元数据注册与管理

GB/T 39725—2020 信息安全技术 健康医疗数据安全指南

YY/T 1833.1—2022 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—2022 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

GB/T 25000.12-2017 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第12部分：数据质量模型

3 术语和定义

3.1

脑机接口 **brain-computer interface**

一种能够在大脑和外部设备间发送和接收信息的系统，从而实现二者的直接通讯。

注1：脑机接口允许在大脑和外部设备间单向或双向通讯，从而实现控制或反馈功能。

注2：脑机接口系统可在人与人之间或人与动物之间建立。

3.2

数据 **data**

信息的可再解释的形式化表示，以适用于通信、解释或处理注:可以通过人工或自动手段处理数据。

[来源：GB/T 5271.1-2000，01-01-02]

3.3

数据集 **data set**

具有一定主题，可以标识并可以被计算机化处理的数据集合。

[来源：WS/T 305-2009，3.1.2]

3.4

数据元 **data element**

用一组属性规定其定义、标识、表示和允许值等的数据单元。

[来源：WS/T 305-2009，3.1.6，有修改]

3.5

元数据 metadata

定义和描述其他数据的数据。

[来源：GB 18391.1-2009，3.2.16]

3.6

采样率 frequency of sample

单位时间从脑电信号中提取并组成离散信号的采样个数。

3.7

电极 electrode

用于头皮或插入脑部区域用于检测点活动的传感器。

[来源：GB/T 9706.225-2021，定义201.3.208，有修改]

3.8

导联 lead

电极间的电压。

[来源：GB/T 9706.227-2021，定义201.3.206]

3.9

标签 label

在脑电数据中反映特定行为、事件与状态的标记。

3.10

交互 interaction

通过测量、解码用户的脑电活动，以实现用户对计算机软件或其他外部设备的控制或信息交流的过程。

3.11

范式 paradigm

脑机接口范式是指在脑机接口系统中用于诱发特定神经活动或指导使用者主动产生特定神经活动的预设执行流程，在脑机接口系统中用于从脑信号中提取和传输信息的特定方法或策略。

3.12

通道 channel

从头皮上记录 EEG 信号的单个电极放置位置。每个通道对应一个特定位置，并提供反映该点大脑电活动的独特信号。

3.13

电极位置规范 electrode position specification

电极位置规范是一套标准化的系统，用于确保在神经电生理学研究中，电极的放置具有一致性和可重复性

3.14

脑电信号 EEG signal

脑电信号大脑神经元放电产生的，反应大脑活动的生物电信号

4 数据采集要求

4.1 电极布局要求

数据集应详细描述电极的类型、数量、放置位置、遵循的电极位置规范、采集时的参考与接地方案与对应电极位置，一般情况下建议遵循国际通用的10/20或10/10等规范。

4.2 接触阻抗要求

数据集应提供脑机接口实验前后所用通道的头皮电极接触阻抗，及测量接触阻抗所用的电流频率和峰峰值。应包括采集时电极阻抗的测试方法与测试结果，采集开始前应确保电极与皮肤接触的阻抗在可接受范围内。其中，接触阻抗应 $\leq 10\text{k}\Omega$ ，测量接触阻抗所用的电流频率 f_A 应满足 $5\text{Hz} \leq f_A \leq 35\text{Hz}$ ，电流的峰峰值 I_p 应满足 $10\text{ }\mu\text{App} \leq I_p \leq 50\text{ }\mu\text{App}$ 。

4.3 背景脑电要求

数据集应提供脑机接口实验前后被试处于睁眼、闭眼和眨眼状态时脑电图机采集的原始背景脑电数据。其中，睁眼时长应 $\geq 10\text{s}$ ，闭眼时长应 $\geq 10\text{s}$ ，眨眼时长应 $\geq 10\text{s}$ ，眨眼间隔约为2s。

背景脑电测量方法如下：

关闭脑电图机软件滤波器以记录原始背景脑电。被试处于静息状态，并按图2所示的时序进行睁眼、闭眼和眨眼动作，不同状态的切换可以通过声音或触觉等方式进行提示

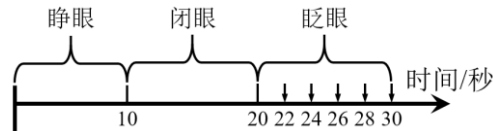


图1 背景脑电记录时序图

4.4 脑机接口系统噪声要求

数据集应提供脑机接口系统的噪声数据和有效值指标，噪声有效值应 $\leq 0.5\text{ }\mu\text{Vrms @ } 0.5\text{-}100\text{ Hz}$ 。

系统噪声测量方法如下：

将脑电电极全部浸泡于0.5%浓度的NaCl溶液中，开启包括刺激设备和脑电图机等在内的脑机接口相关设备，对于电刺激设备而言，需将刺激电极短接以提供刺激电流回路。关闭频率为50Hz的脑电图机软件陷波器，并开启带宽为0.5-100Hz的脑电图机软件带通滤波器，记录时长应 $\geq 10\text{s}$ 的脑机接口系统噪声。噪声有效值指标计算方法见公式（1）。

$$V_{n_rms} = \sqrt{\frac{1}{N} \sum_{n=1}^N |V_n|^2} \quad (1)$$

式中：

V_{n_rms} ——噪声有效值；

N ——采样点个数；

V_n ——噪声采样值。

4.5 采样率要求

数据集应提供采集脑电的采样率，采样率应 $\geq 1000\text{ SPS}$ 。

采样率测量方法如下：

脑电图机采集10 s数据，获得某一通道的总采样点数 N ，采样率 f 计算方法见公式（2）。

$$f = \frac{N}{10} \quad (2)$$

式中：

YY/T ×××××—××××

f ——采样率；
 N ——采样点个数。

4.6 信号分辨率

数据集应说明采集时信号模数转换的位数。

4.7 滤波器设置

数据集应明确采集系统中所采用的滤波器类型、方法、重要参数。

4.8 信号分辨率

对于高精度的 BCI 系统其分辨率不少于 12 位，对医用级 BCI 系统其分辨率不少于 24 位。

4.9 信噪比

为了实现高质量的脑机接口采样系统，尽量减少噪声源（肌电信号、电源干扰）对 EEG 信号的影响。数据集应记录采集环境的信噪比，应该明确描述采集数据的信噪比计算方法、公式，建议在数据使用与共享前提供信噪比结果作为信号质量的参考信息。

4.10 刺激数据

数据集应记录和保存刺激数据，数据应包括刺激类型、刺激通道、刺激时间、刺激强度等信息。

4.11 事件标记

应保存并描述数据采集过程中特殊事件的标记数据，并对标记类型进行必要的说明。

4.12 其他记录数据

应描保存并述数据采集过程中其他模态的同步数据文件，包括但不限于行为记录视频、音频、眼动、眼电、心电等数据文件。

4.13 数据标签要求

4.13.1 标签内容

标签内容由标签值、标签说明、标签时刻组成，能够反映脑电数据在特定时刻的关键行为、事件与状态。其中，标签值用于数据记录过程中标定不同关键行为、事件与状态的出现，其数据类型可以为数字、字母、符号、文本（推荐在数据采集记录过程中使用数字或字母的数据类型来记录标签）。标签说明用于解释标签值对应的关键行为、事件与状态的具体情况，其数据类型可以为文本。标签时刻用于标定不同关键行为、事件与状态出现的时刻，其数据类型可以为数字、时间、文本（推荐在数据采集记录过程中使用对应采样点索引的数字来记录标签时刻）。

示例：

对于一个包含两种稳态视觉诱发电位的脑电数据集，一个完善的标签内容可以如下表所示：

表1 标签内容示例

标签值	标签说明	标签时刻
1	对应事件为：由 10Hz 正弦波调制的视觉刺激开始闪烁	10285

2	对应事件为：由 15Hz 正弦波调制的视觉刺激开始闪烁	12285
1	对应事件为：由 10Hz 正弦波调制的视觉刺激开始闪烁	14285
2	对应事件为：由 15Hz 正弦波调制的视觉刺激开始闪烁	16285
...	...	...

其中标签时刻的数据类型为数字，表示刺激开始闪烁时对应的脑电数据采样点索引。

4.13.2 标签说明的准确性要求

数据集应提供能够准确反映关键行为、事件与状态的标签说明。不同标签值对应的标签说明不应完全一致。

4.13.3 标签种类的详细性要求

标签种类应满足描述与研究目标相关的所有关键行为、事件与状态。

示例：

在一个包含40种闪烁频率的稳态视觉诱发电位数据集中，若无其他研究目标（例如探究实验时间对受试者状态的影响），则标签数量应不少于40种。

4.13.4 标签的一致性要求

在整个数据集中使用的反映同种行为、事件与状态的标签应该是前后一致的，即在实验过程中采用的标签值不应改变。

4.13.5 标签时刻的锁时性要求

标签时刻应精确对应于行为、事件与状态产生或存在的时间点。实验范式产生标签的实际时刻与脑电图机记录到标签的时刻之间的误差应≤1ms。

5 数据传输要求

5.1 脑电数据传输时间延迟

从脑电图机传输至上位机的时间延迟应<1ms。时间延迟过大的情况下应拒绝系统启动请求。

测试方法：采集设备应置有内部时钟，则可记录数据发送时间，并在上位机记录数据接收时间，二者之差即为时间延迟。

5.2 丢包率

数据集应记录脑电采集过程中的丢包情况并计算丢包率。丢包率定义为上位机接收到的数据量与脑电图机发送的数据量之比。对于有线传输的情况，丢包率应为0；对于无线传输的情况，丢包率应≤0.01%。

5.3 传输稳定性信息

建议明确系统在数据传输过程中的稳定性指标及采用的提升传输稳定性的方案，方便评估稳定性对数据使用的影响。BCI系统传输数据应具有抗干扰的能力，尤其是在无线传输的场景下。BCI系统应

设计有效的重传机制，确保在发生数据包丢失时，能够及时重新传输，确保BCI系统数据稳定传输。BCI数据传输应又备用线路，系统应能够在断线后迅速选择其他线路重连，以恢复数据传输。

5.4 传输实时性信息

应明确系统的数据传输方案、系统延迟等参数，保证数据传输的实时性。传输延迟应尽可能低，通常要求在X ms以内，以保证系统能够实时响应。系统应有实时性保障，使用诸如数据缓冲策略来平衡传输时的突发流量，保障BCI系统实时性

5.5 传输同步性信息

应明确数据传输过程中多设备/信号之间时钟是否对齐，记录数据同步机制、同步精度等信息

5.6 传输速率

速率要求：对于多通道EEG数据传输，传输速率应能够支持至少X Mbps速率。

6 数据存储要求

6.1 数据存储格式

数据集文件应包含脑电数据、元数据和其他数据，推荐以BDF、EDF、CNT等封装格式打包。

6.1.1 脑电数据

数据集中脑电数据的存储方式应为连续数据，即在实验任务中硬件设备直接采集得到的整段脑电信号，推荐表示为二维数组（导联数量，采样点数）。

6.1.2 元数据

数据集中元数据提供了数据集使用的必要补充信息，推荐以结构体的形式存储。

元数据应包含标签信息、采样率、导联名称与位置信息，推荐包含滤波范围、采集设备信息。其符号、数据形式应如表3所示。

表2 元数据内容

关键字名称	符号	数据形式
标签	Label	标签值：字符串 对应的采样点位置索引：数字
采样率	Fs	数字
滤波范围	Fcut	带通:(数字, 数字); 低通:(None,数字); 高通:(数字,None)
导联名称与位置	Loc	导联名称：字符串 导联位置：（x:数字, y:数字,z:数字）
采集设备	Device	字符串

元数据的收集、录入、存储、查询、共享等应符合GB/T 30524-2014要求。

6.1.3 其他数据

数据集文件应包含采集过程中记录的背景脑电数据、噪声数据和指标，允许包含眼电、肌电等其他模态生理数据，以及额外的数据补充信息。

6.2 匿名化和去隐私化

面向医用的脑机接口脑电数据集在数据访问权限方面应遵循严格的规范，以确保相关人员的隐私数据不被泄露。数据集应确保所有涉及受试者身份的个人信息都经过匿名化处理，无法直接或间接识别受试者身份。此过程应遵循相关法律法规，如《中华人民共和国网络安全法》和《中华人民共和国个人信息保护法》。可能包含受试者身份的个人信息包含：姓名、可标识身份的证件号码、住址、详细的出生日期等。

6.3 存储管理

应对数据集存储文件进行管理和维护，包括备份、更新和优化等方面。

6.4 数据安全

数据集存储文件的安全保护应符合GB/T 39725—2020要求，确保数据存储和使用过程的合法性和合规性。

6.5 数据版本

应记录数据文件、数据架构等变更历史，及时将数据的变更内容与数据使用方同步，版本更新前应经过严格的检查和质量控制

6.6 数据压缩

如对数据进行了压缩，应说明压缩和解压缩方法与工具，一般情况下建议使用无损压缩方式

6.7 数据存储信息

应描述数据存储的内容、存储方式与存储路径等信息，原始数据文件应当保留、处理后的数据文件应当说明处理方法与参数

6.8 数据命名规范

应提供数据命名的规则、字段代表的具体信息

6.9 数据备份

应规划数据的备份方式并定期对数据进行备份，准备数据恢复的流程方法，保护数据不受意外删除、丢失、损坏或其他灾害影响

7 数据使用要求

7.1 数据访问与权限

数据集管理过程中应根据使用者的角色和需求设置不同的访问权限。研究人员、开发人员、医生或其他相关人员应仅能访问与其工作相关的数据和信息。数据集应拒绝任何未经授权访问请求。同时应当建立数据库访问控制列表，设计明确的机制，使用合适的工具定义用户的角色与权限并对数据访问和操作进行记录和监督。

7.1.1 数据加密

数据集在存储和传输过程中应进行加密，以防止数据泄露或未经授权的访问。

7.1.2 访问可追溯性

数据集管理过程中应记录所有数据访问操作，并保留详细的日志。日志内容应至少包含：访问者信息、访问时间、访问终端、访问过程中进行的操作。

7.1.3 数据共享要求

数据提供方应明确数据的申请方式，并对数据共享的内容、使用范围、数据分享和使用双方的权利义务等信息进行说明。在数据集需要跨机构共享时，应制定明确的协议，确保数据的使用符合伦理规范和法律要求。

7.2 数据处理

7.2.1 通则

脑电数据处理过程应符合机器学习通用标准，分为模型训练、特征提取、分类预测三个部分。具体而言，有监督的识别算法应至少包含模型训练模块，特征提取模块和分类预测模块，无监督的识别算法应至少包含特征提取模块和分类预测模块方法，深度学习方法应至少包含模型训练模块和分类预测模块。数据集若包含数据处理过程，则应给出每部分的输入数据和输出数据的格式。

7.2.2 模型训练

模型训练部分，利用训练数据进行建模。输入X为多试次脑电信号，表示为三维数组（试次数量，导联数量，采样点数）；输入y为对应的标签数据，表示为一维数组（试次数量，）；输出为算法模型。

7.2.3 特征提取

利用已有模型将输入数据转换为相应特征。输入X为多试次脑电信号，表示为三维数组（试次数量，导联数量，采样点数）；输出为多试次特征，表示为二维数组（试次数量，特征数量）。

7.2.4 分类识别

用于预测新数据的标签。输入X为多试次脑电信号，表示为三维数组(试次数量，导联数量，采样点数)；输入y为对应的预测标签，表示为一维数组（试次数量，）。

7.3 数据评估指标计算

7.3.1 计算方法

若数据集中涉及对有关指标的计算，则应给出规范的计算方法，确保计算结果的可复现性。数据集应详细描述指标计算所采用的方法以及每步操作输入数据和输出数据的信息，记录该方法采用的具体参数，说明采用该方法及其参数的目的，并给出该指标的计算结果。

7.3.2 计算误差

数据集应给出有关指标的计算误差。对于具有随机性的计算方法，应进行多次计算以给出误差范围，计算次数应 ≥ 5 次。

7.3.3 可用的评估指标

7.3.3.1 分类性能指标

7.3.3.1.1 精确度

即阳性预测值（Precision），可以理解为被预测为阳性的样本中有多少是正确的。精确度尽可能保证预测的正确率。

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

其中，TP为预测为阳性且实际也为阳性的样本数量或比率，FP为预测为阳性但实际为阴性的样本数量或比率。

7.3.3.1.2 召回率

即真阳性率（Recall），可以理解为有多少阳性样本被正确地预测。召回率高则说明漏检率低。

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

其中，FN为预测为阴性但实际为阳性的样本数量或比率。

7.3.3.1.3 F1-Score

定义为精确率和召回率的调和平均值，范围为0~1，综合考虑了精确度和召回率的影响。

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

7.3.3.1.4 分类准确率、平衡准确率

分类准确率不考虑类平衡。平衡准确率针对准确率指标在不平衡数据集的一种改进。平衡准确率综合考虑了召回率和特异性这两个指标。

分类准确率P定义如下：

$$P = \frac{T}{T + N} \quad (6)$$

其中T为预测正确的样本数量或比率，N为预测错误的样本数量或比率。

平衡准确率BP定义如下：

$$BP = \frac{1}{K} \sum_i^K \frac{TP_i}{TP_i + FN_i} \quad (7)$$

其中K为类别数量， TP_i 为预测为第i类样本且实际也为第i类的样本数量或比率， FN_i 为预测为第i类样本但实际不为第i类样本的样本数量或比率。

7.3.3.2 信息传输速率

7.3.3.2.1 基本定义

表示单位时间T内能够达到的比特率B。比特率表示测量一定时间内发送的信息量：

$$B = \log_2 C + p \log_2 p + (1 - p) \log_2 \frac{(1 - p)}{(C - 1)} \quad (8)$$

C为类别总数，p为分类准确率。在离线实验中，比特率中的p表示k折交叉验证的分类精度：

$$p = \sum_{k=1}^k p_k / k \quad (9)$$

在不需要对错误的输出结果进行删除和修正的在线实验中，比特率中的 p 表示识别正确的指令数量与指令总数之间的比值：

$$p = N_{right} / N_{total} \quad (10)$$

最终的ITR按下式计算[4]：

$$ITR = B / T \quad (11)$$

$$T = (T_t + T_s) \quad (12)$$

T_t 表示任务时长； T_s 表示任务切换时长。

7.3.3.2.2 理论信息传输速率

在离线实验中，可以计算理论信息传输速率，此时 $T_s = 0$ 。

7.3.3.2.3 实际信息传输速率

在在线实验中，可以计算实际信息传输速率，此时 T_s 应与实际测试过程中的时间参数一致。在实际测试过程中，可以将在线任务分为引导指令输出的在线任务和自由指令输出的在线任务，计算ITR时的分类准确率可以按照式3计算。在引导指令输出的在线任务中，每个指令都会给出的提示，受试者只需按照提示操作脑机接口。式3中的 N_{total} 表示引导受试者完成的总指令数， N_{right} 表示受试者准确输出的指令数。而在自由指令输出的在线任务中，只向受试者提出需要完成的任务，受试者进行自由操作，直到完成给定任务。 N_{total} 表示受试者完成引导受试者完成的总指令数， N_{right} 表示完成任务所需的最小指令数。

7.3.3.3 时间最优比

时间最优比定义为完成指定任务理论最优时间和实际完成时间的比值[5]。其中，理论最优时间表示不出错地完成指定任务所耗费的最短时间。而实际完成时间表示受试者在执行任务的过程中，有犯错且有修正地完成指定任务所耗费的时间。

例如，在字符拼写任务中完成“Hello”这个单词，理论最优时间即为相应脑机接口识别输出5个指令所耗费的时间。而受试者在犯了N次错误的情况下，还需要进行N次修正，因此实际完成时间为脑机接口识别输出5+2N个指令所耗费的时间。

7.3.3.4 菲茨吞吐量

菲茨吞吐量定义为任务难度系数和完成时间的比值。任务难度系数如何定义？其中，任务难度系数（TD）的计算公式为：

$$TD = \log_2 \left(\frac{D}{W} + 1 \right) \quad (13)$$

其中D为给定的距离，W为目标的宽度。时间（T）的单位为秒。

例如，在控制鼠标从A点移动到B点的任务中，任务难度系数与A、B两点之间的距离正相关，与被控制物体（鼠标）的尺寸反相关。

7.3.3.5 时间指标

7.3.3.5.1 信号处理时间

脑机接口脑电数据集应记录在线数据处理过程的信号处理时间。起始时刻为开始进行数据处理的时刻，结束时刻为得到处理结果的时刻。根据数据集的应用需求不同，信号处理时间应满足以下要求：

——在交互场景中（例如：神经反馈训练等辅助康复场景，轮椅、外骨骼等外部硬件设备控制场景，脑控打字等软件控制场景），脑机接口应即时处理脑电数据并反馈结果，形成自然的交互过程，因此信号处理时间应 $\leq 50\text{ms}$ ；

——在非交互场景中（例如：疾病辅助诊断场景，脑状态监测场景），脑机接口只需出具相关结果，因此不限制信号处理时间。

7.3.3.5.2 结果反馈时间

脑机接口在计算得到脑电的识别结果后，结果将被反馈至外部设备实现控制或生成提示。脑电数据集应记录在线结果的反馈时间，起始时刻为得到数据处理结果的时刻，结束时刻为外部设备或反馈系统接收到处理结果的时刻。根据数据集的应用需求不同，结果反馈时间应满足以下要求：

——在交互场景中，结果反馈时间应 $\leq 10\text{ms}$ ；

——在非交互场景中，结果反馈时间不做限制。

7.3.4 第三方测试

数据集中所有指标均应通过第三方的合规测试，并开具经过第三方审核的测试大纲和测试结果报告。

8 数据集质量要求

8.1 完整性

8.1.1 准确性

脑机接口脑电数据集应符合数据集说明文档中有关准确性的陈述，如：

- 记录脑电数据的准确性；
- 文字描述准确、清晰、无歧义；
- 标签内容的准确性；
- 采集设备的有效性；
- 人员操作的有效性；
- 数据形式的合理性。

8.1.2 完备性

脑机接口脑电数据集应包含实现数据集预期用途、满足应用场景需要的信息，如：

- 数据结构与格式应支持数据元、元数据的保存、传输与访问，避免信息空白、遗漏或丢失；
- 对不可避免的缺失数据、离群值、坏导数据，应给出明确的处理方式，证明应用场景的合理性，对相关数据元加以标识。
- 元数据应描述数据集标签信息、数据采集情况信息、数据内容信息、数据质量特性等相关信息；
- 数据包含的信息应支持数据标注与参考标准溯源的需要。

8.2 唯一性

同一数据集或子集内的数据元应是唯一的。同一数据集的各个子集应是可区分的。同一受试者的不同时间、空间的数据应明确描述处理方法。

8.3 一致性

数据集应符合数据集说明有关一致性的陈述，包括内部一致的程度和外部一致的程度。内部一致性是指相同来源数据之间的相关性，如：

- 同一数据元在数据集生存周期不同阶段应保持一致；
- 数据的特征应保持一致；
- 数据的采集、预处理、评估时采用的评估方法应保持一致；
- 原始记录、中间记录与最终记录应保持一致。

外部一致性是指不同来源数据之间的相关性，如：

- 不同来源的数据应在数据特征方面保持一致；
- 离群值应是可解释的；

不同来源的数据在采集、标注环节应遵从相同的法规、技术标准、医学规范、其他文献要求。

8.4 确实性

数据集应符合数据集说明有关确实性的陈述，如：

- 数据应来自真实的临床数据采集流程；适当时，数据采集涉及的设备、人员、方法应符合技术标准、临床规范、伦理要求或专家共识；
- 数据扩增、数据合成活动和结果应是可溯源和可解释的；
- 元数据应如实对数据集进行描述。

8.5 时效性

数据的采集、标注、流转、归档、变更等活动的时限应符合数据集说明有关时效性的陈述。动态更新的数据集应明确数据的更新周期、更新方法和更新比例。若数据涉及临床诊断、治疗中的时序过程，应证明数据在临床时效上的合理性。

8.6 可访问性

数据集应满足数据集的预期用途和应用场景范围内的访问需要。

8.7 依从性

数据集应符合数据集说明有关依从性的陈述。

8.8 资源利用性

数据集的处理和使用应符合数据集说明有关资源利用性的陈述。

8.9 可追溯性

数据集应符合数据集说明有关可追溯性的陈述，具有相关记录，如：

- 原始数据来源、元数据来源、合规性证明；
- 数据采集活动记录；
- 人员管理记录；
- 数据标注流程记录；
- 盲态管理记录；

YY/T ×××××—××××

- 数据流通记录；
- 数据质疑、审计、停用、纠正记录；
- 标注工具、平台使用记录；
- 数据集标注结果的统计信息查询，包括标注进展、标签统计、标注者进展统计、难例集等；
- 数据服务异常与故障记录；
- 数据维护与备份记录；
- 数据更新记录；
- 云服务提供商名称、联系方式、云服务类型等。

8.10 可理解性

数据集应符合数据集说明有关可理解性的陈述。

8.11 可得性

数据集应符合数据集说明文档有关可得性的陈述。

8.12 可移植性

数据集应符合数据集说明有关可移植性的陈述。如果数据集允许在不同的平台和系统下使用，数据质量不应随平台和系统发生变化。

8.13 可恢复性

用于保持数据集质量并抵御失效事件的措施应符合数据集说明有关可恢复性的陈述。

8.14 代表性

数据集的数据特征层次、样本来源多样性、数据多样性等应符合数据集说明有关代表性的陈述。

8.15 评价方法

参照YY/T 1833.2-2022第6章的方法，对脑电数据集开展抽样检验，形成抽样样本集，通过试验验证、操作检查的方式，判断数据集是否满足8.1~8.14的要求。

9 数据集说明要求

9.1 实验范式信息

9.1.1 实验范式类型

数据集说明应描述采用的实验范式类型。脑机接口的实验范式可以分为以下几种类型：

a) 脑电诱发范式：

- 视觉诱发范式：通过视觉刺激引发脑电活动，如闪光、图像等；
- 听觉诱发范式：通过听觉刺激引发脑电活动，如声音、语音等；
- 触觉诱发范式：通过触觉刺激引发脑电活动，如振动、触摸等；
- 其他感觉诱发范式：通过其他感觉刺激引发脑电活动。

b) 自主节律范式：

- 运动想象范式：通过受试者想想运动来引发脑电信号的变化；
- 注意力调节范式：通过冥想、放松或集中注意力来改变脑电活动；

——情绪调节范式：通过控制情绪状态（如快乐、愤怒）来调节脑电信号。

c) 任务相关范式：

——认知任务：执行特定的认知任务（如计算、记忆等）并记录脑电信号；

——运动任务：执行实际运动任务（如手部动作等）并记录脑电信号；

——语言任务：执行语言处理任务（如阅读、说话等）并记录脑电信号；

——反应任务：执行特定时间反应任务（如对特定类型的图片做出按键反应）并记录脑电信号。

d) 其他实验范式。

9.1.2 实验范式内容

数据集说明应描述采用实验范式的具体内容，包括使用的刺激类型、刺激参数（如闪光刺激的尺寸、强度、频率等）、任务要求、范式中的引导语或引导标识等。

9.1.3 范式流程

应描述范式基本流程，包括任务描述、刺激方法、用户交互方式、环境要求等信息

9.1.4 范式脑电响应特征

应提供脑电信号响应时间、响应区域等信息的描述

9.1.5 范式安全性信息

应描述脑机接口系统的使用过程中对用户产生的不良影响以及识别出的潜在风险

9.1.6 实验过程

数据集说明应描述实验过程中采集对象执行的具体任务、实验流程对应的时序信息以及实验过程中出现的其他情况。

9.2 设备配置信息

9.2.1 数据采集设备配置信息

数据集说明应描述数据采集设备的相关配置信息。

9.2.2 计算机配置信息

数据集说明应描述在实验过程中使用的（如有）以及在数据质量评估过程中使用的（如有）计算机的配置信息。

9.2.3 其他设备配置信息

数据集说明应描述在数据采集过程中和数据处理过程中使用的其他设备的信息。

9.3 实验对象信息

9.3.1 合规性陈述

数据集说明应提供数据来源的合规性陈述。涉及动物、人体实验数据集应提供审批的伦理结构名称和伦理号等信息。

9.3.2 性别

数据集说明应描述受试者的性别信息。

9.3.3 年龄

数据集说明应描述受试者的年龄信息。

9.3.4 左/右利手

根据研究目的与实验设计，相关数据集说明应描述受试者的左/右利手信息。

9.3.5 视力

根据研究目的与实验设计，相关数据集说明应描述受试者的视力情况。

9.3.6 相关病史

根据研究目的与实验设计，相关数据集说明应描述受试者的有关病史情况。

9.3.7 有/无 BCI 经验

数据集说明应描述受试者是否具有BCI经验。

9.4 数据采集信息

9.4.1 电极布局

数据集说明应描述电极布局情况，并给出图示。

9.4.2 接触阻抗

数据集说明应给出实验前后各个电极的接触阻抗值。

9.4.3 背景脑电

数据集说明应描述采集过程中背景脑电采集过程的信息，包括背景脑电采集的起止时间、采集过程中受试者的状态。

9.4.4 系统噪声

数据集说明应描述采集过程中系统噪声的信息。

9.4.5 采样率

数据集说明应给出采集过程中使用的采样率。

9.4.6 标签质量特性描述

9.4.6.1 标签说明的准确性

数据集说明应描述数据集中标签说明与实际实验一致的程度，并以书面形式提供可验证的指标。

9.4.6.2 标签种类的详细性

数据集说明应描述数据集中标签种类是否满足覆盖研究目的中设计的实验条件或种类，并以书面形式提供可验证的指标。

9.4.6.3 标签的一致性

数据集说明应描述数据集中标签之间一致的程度，考虑内部一致性和外部一致性，并以书面形式提供可验证的指标。

9.4.6.4 标签时间的锁时性

数据集说明应描述数据集中标签时间与实际行为、事件或状态产生时刻之间的误差，并以书面形式提供可验证的指标。

9.5 数据传输信息

9.5.1 脑电数据传输时间延迟

数据集说明应描述脑电数据传输的时间延迟信息，并给出测试方法和测试结果。

9.5.2 丢包率

数据集说明应描述脑电数据传输过程中的丢包情况，并给出测试方法和测试结果。

9.6 数据存储信息

9.6.1 基本结构

数据集说明应说明数据包含的内容、存储方式以及封装格式。

9.6.2 元数据

数据集说明应描述数据集元属性包含的内容、存储范式以及封装格式。

9.6.3 其他数据

数据集说明应说明所存储的其他数据包含的内容、存储方式以及封装格式。

9.7 数据使用说明

9.7.1 数据访问与权限

根据研究目的与实验设计，数据集说明应描述信息安全、数据保密相关的措施，说明数据集限定的使用范围、使用规范、使用限制，并以书面形式展示可验证的证据。

9.7.2 数据处理与分析

根据研究目的与实验设计，数据集说明应给出数据处理与分析过程中所有操作的具体过程和参数，并说明执行此操作的原因。

9.7.3 数据评估指标

根据研究目的与实验设计，数据集说明应描述数据集中所有评估指标的计算方法、计算误差和计算结果，并给出第三方检测的测试大纲和测试报告。

9.7.4 数据使用范围

9.7.5 数据载入示例

数据集说明宜提供不少于2种编程语言下的数据载入示例代码。

9.7.6 数据应用示例

数据集说明宜提供不少于2种编程语言下的典型应用示例代码。

9.8 数据集描述

9.8.1 数据集类型

数据集说明应注明数据集的类型。根据YY/T 1833.2-2022附录A中的描述，数据集类型按照以下五个维度进行划分：

——预期用途：模型训练与调优集、模型验证集、第三方性能测试集、临床评价数据集、产品质量集等；

——数据来源：公有数据集、私有数据集、混合数据集等；

——用户类型：自用数据集、他用数据集等；

——访问管理方式：开放数据集、封闭数据集等；

——更新形式：静态数据集、动态数据集等。

9.8.2 数据集形态

数据集说明应描述数据集的数据模态、数据格式、数据量、存储方式。

9.8.3 数据集层次

数据集说明应描述数据集的数据层次、子集分解和组合变化。

9.8.4 应用场景

数据集说明应描述数据集预期模拟的应用场景。

9.8.5 版本控制

数据集说明应描述数据集的版本号、存储信息、说明文档编制时间、使用状态。

9.8.6 数据集更新

对于动态数据集，数据集说明应描述新增数据的入组规则和数据退役规则。

9.8.7 文档管理

数据集说明宜通过计算机技术（如软硬件、普通网络、云服务），对文档内容进行发布、交换、管理和查询。

9.8.8 数据元属性

数据集说明宜描述数据元的共用属性、专用属性，如数据集版本、注册机构、分类模式、主管机构等。

9.9 数据集质量特性描述

9.9.1 完整性

9.9.1.1 准确性

数据集说明应陈述数据与真实值一致的程度，可按照数据采集、信息记录、数据标注等环节展开，如：

a) 记录信息的准确性；

- b) 文字描述准确、清晰、无歧义;
- c) 数据标注与数据对应准确, 并给出可验证方法;
- d) 采集设备的有效性;
- e) 人员操作的有效性;
- f) 数据形式的合理性;

并以书面形式提供可验证的指标。

9.9.1.2 完备性

数据集说明应陈述数据集包含的信息能覆盖数据集的预期用途的程度, 并以书面形式展示可验证的证据。

其中数据集应包含信息如:

- a) 数据结构与格式应支持采集信息、数据元、标注信息、元数据的保存、传输与访问, 避免信息空白、遗漏或丢失;
- b) 对不可避免的缺失数据、离群值, 应给出明确的处理方式, 证明应用场景的合理性, 对相关数据元加以标识;
- c) 元数据应描述数据集标识信息、内容信息、数据质量特性等相关信息;
- d) 数据包含的信息应支持数据标注与参考标准溯源的需要。

9.9.2 唯一性

数据集说明应陈述数据集、子集和数据在何种程度上是唯一的, 并以书面形式提供可验证的指标。

9.9.3 一致性

数据集说明应陈述数据之间一致的程度, 考虑内部一致性和外部一致性, 并以书面形式提供可验证的指标。

其中内部一致性是指相同来源数据之间的相关性, 如:

- a) 同一数据元在数据集生存周期不同阶段应保持一致;
- b) 数据的特征应保持一致;
- c) 数据的采集、预处理、标注方法应保持一致;
- d) 原始记录、中间记录与最终记录应保持一致。

外部一致性是指不同来源数据之间的相关性, 如:

- a) 不同来源的数据应在数据特征方面保持一致;
- b) 离群值应是可解释的;
- c) 不同来源的数据在采集、标注环节应依从相同的法规、技术标准、医学规范、其他文献要求。

9.9.4 确实性

数据集说明应陈述数据集在何种程度上是真实和可信的, 包括数据、元数据的获取和处理, 并以书面形式展示可验证的证据。

9.9.5 时效性

数据集说明应陈述数据集开发阶段的各个环节所需时限符合预期的程度, 考虑预处理、清洗、标注等, 并以书面形式提供可验证的指标。数据的采集、标注、流转、归档、变更等活动的时限应符合数据集说明有关时效性的陈述。动态更新的数据集应明确数据的更新周期、更新方法和更新比例。若数据涉及临床诊断、治疗中的时序过程, 应证明数据在临床时效上的合理性。

9.9.6 可访问性

数据集说明应陈述数据集可被访问的程度，并以书面形式展示可验证的证据。

数据集应满足数据集的预期用途和应用场景范围内的访问需要。

9.9.7 依从性

数据集说明应陈述数据集依从的标准规范、专家共识、操作规程或其他参考文献。

9.9.8 资源利用性

数据集说明应陈述执行数据集相关任务需要的资源消耗，并以书面形式展示可验证的依从性证据，如访问、读取数据、预览、检索等任务需要的软件、硬件、网络配置。

9.9.9 可追溯性

数据集说明应描述数据可被追溯的程度，考虑数据采集历史、数据标注历史、数据访问踪迹和数据变更踪迹，并以书面形式展示可验证的证据。

数据集应符合数据集说明有关可追溯性的陈述，具有相关记录，如：

- a) 原始数据来源、元数据来源、合规性证明；
- b) 数据采集活动记录；
- c) 人员管理记录；
- d) 数据标注流程记录；
- e) 盲态管理记录；
- f) 数据流通记录；
- g) 数据质疑、审计、停用、纠正记录；
- h) 标注工具、平台使用记录；
- i) 数据集标注结果的统计信息查询，包括标注进展、标签统计、标注者进展统计、难例集等；
- j) 数据服务异常与故障记录；
- k) 数据维护与备份记录；
- l) 数据更新记录；
- m) 云服务提供商名称、联系方式、云服务类型等。

9.9.10 可理解性

数据集说明应使用数据集的用户可理解的术语，对数据元、元数据和标注结果的含义提供解释，并以书面形式展示可验证的证据。

9.9.11 可得性

数据集说明应陈述数据集能被授权用户使用和检索的程度，并以书面形式展示可验证的证据。

9.9.12 可移植性

数据集说明应陈述数据集能被安装、替换或从一个系统移动到另一个系统中，并保持已有质量的属性的能力，考虑数据安装、替换、移动的效率，并以书面形式展示可验证的证据。数据集应符合数据集说明有关可移植性的陈述。如果数据集允许在不同的平台和系统下使用，数据质量不应随平台和系统发生变化。

9.9.13 可恢复性

数据集说明应陈述数据集可被恢复的程度，并以书面形式展示可验证的证据。数据集说明可提供用于数据恢复的措施。数据集说明可提供防止数据集使用过程中出现中断、失效的措施。

9.9.14 代表性

数据集说明应对样本的组成、比例、人群分布特征、数据的多样性和接近应用场景的程度开展分析，并以书面形式提供可验证的指标。

9.9.15 保密性

数据集应符合数据集说明有关保密性的陈述，采取措施确保其只能被授权用户访问。隔离使用的数据集应具有数据集授权访问机制、隔离保护机制，应具有防止数据泄露、数据篡改、数据丢失的措施，如数据匿名化、物理隔离、数据审计等。

9.9.16 同步性

数据集应符合数据集说明有关同步性的陈述，如：

- a) 同步记录的数据及数据间逻辑关系
- b) 数据同步的方法，包含必要的公式、工具说明
- c) 数据同步的精度

9.9.17 有效性

数据集应符合数据集说明有关有效性的陈述，如：

- a) 产生数据集的实验目标，有效性定义
- b) 有效性评价的指标
- c) 有效性评价的方法，包含必要的公式、工具说明

10 数据处理

10.1 数据脱敏处理

应描述用于保护受试者隐私的技术手段，如数据去标识化、数据匿名化等。适当时，数据集说明文档应描述数据去标识化或者数据匿名化的规则。对参与者的身份信息进行匿名化处理，去除或加密所有能够直接或间接识别个人的字段，如姓名、出生日期、地址等。对于无法去除或加密的数据进行通过降低数据的精度来保护隐私。

10.2 数据去噪处理

应描述数据处理时使用的去噪方法及其参数设置

- 1 能进行噪声来源分析，识别和记录常见的噪声来源（如电源、肌电等）。
- 2 标明去噪算法，标明采用的去噪算法，（数据使用中预处理标准化）

10.3 数据滤波处理

应描述数据处理时使用的滤波方法及其参数设置

- 1 滤波器选择：根据研究目标选择适当的滤波器类型（如低通、高通、带通、带阻滤波器）和滤波器阶数。

2 频段设置：明确设定滤波器的截止频率。对于 EEG，可以设置低通与高通滤波器，去除低频和
高频噪声。去除特定频段的噪声或不需要的信号成分，增强目标信号的质量。

10.4 数据标注

如果对数据进行标注，应保存标注文件，并描述数据集标注的内容、方法、标准、参与人员等信
息标准化标注：使用统一的标注规范（如时间戳、事件类型、任务阶段）确保不同实验或数据集的标
注方式一致。建立标注验证流程，确保标注的准确性和一致性。

10.5 数据对齐处理

模态间数据时间上的一致性。

10.6 数据筛选

时采用的标准，包括采用标准和剔除标准，以及使用的筛选方法质量评估标准：定义数据筛选的
质量标准，如信噪比、伪影含量、采集完整性等。剔除不符合要求的数据。筛选记录：详细记录筛选
的标准和结果，提供筛选后的数据集和原始数据的对比。确保筛选正确。

10.7 数据分段

数据分段通常基于实验目的或脑电活动特点进行，应描述数据分段时数据分段的含义、分段使用
的方法与参数，分段后的数据应按照标准化格式进行存储。

参 考 文 献

- [1]国家市场监督管理总局, 国家标准化管理委员会. GB/T 9706.1—2020 医用电气设备+第1部分: 基本安全和本性能的通用要求[S]. 北京: 中国标准出版社, 2020.
- [2] 国家市场监督管理总局, 国家标准化管理委员会. GB 9706.226—2021 医用电气设备 脑电图机的基本安全和本性能专用要求[S]. 北京: 中国标准出版社, 2021.
- [3]国家市场监督管理总局, 国家标准化管理委员会. GB 9706.227—2021 医用电气设备 心电监护设备的基本安全和本性能专用要求[S]. 北京: 中国标准出版社, 2021.
- [4] 国家市场监督管理总局, 国家标准化管理委员会. GB 9706.225—2021 医用电气设备 心电图机的基本安全和本性能专用要求[S]. 北京: 中国标准出版社, 2021.
- [5] Wolpaw J R, Wolpaw E W. Brain-Computer Interfaces: Principles and Practice [M]. New York, NY, USA: Oxford University Press, 2012.
- [6] International Electrotechnical Commission. IEC 80601-2-26:2019 Medical electrical equipment - Part 2-26: Particular requirements for the basic safety and essential performance of electroencephalographs[S]. Switzerland: International Electrotechnical Commission, 2019.
- [7] 中华人民共和国, 原国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 191—2008 包装储运图示标志[S]. 北京: 中国标准出版社, 2008.
-