



中华人民共和国国家标准

GB/T XXXX-XXXX

采用脑机接口技术的医疗器械 神经电信号采集与解码性能评价通用方法

(Medical device using brain-computer interface technology —General Evaluation Method for the Performance of Neural Signal Acquisition and Decoding)

标准草案稿

20XX-XX-XX 发布

20XX-XX-XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

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专利的责任。

本文件由国家药品监督管理局提出。

本文件由全国外科植入物和矫形器械标准化技术委员会有源植入物分技术委员会归口。

本文件起草单位：

本文件主要起草人：

采用脑机接口技术的医疗器械 高通量神经电信号采集与解码性能评价通用方法

1 范围

本文件规定了以神经电生理信号为输入的脑机接口医疗器械的神经电信号采集性能评价方法、数据质量评价方法、特征提取与解码模型性能评价方法、数据与性能评价方法。

本文件适用于以神经电生理信号为输入的脑机接口医疗器械。

2 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。

3 术语和定义

下列术语和定义适用于本文件。

4 神经电信号采集性能评价方法

4.1 信号采集基本要求

4.1.1 通道数与采样频率

4.1.1.1 多通道配置基本要求

4.1.1.1.1 通道数设计原则

侵入式/半侵入式设备：

- a) 单病例同步采集通道数 ≥ 32 导，支持高密度电极阵列（如 8×8 网格电极、 10×10 深部电极阵列）；
- b) 通道布局需遵循立体定向脑图谱（如MNI坐标系），覆盖目标功能区（如运动皮层、前额叶、海马），相邻触点间距 $\leq 5\text{mm}$ ，通道密度按正四边形排列 ≥ 4 触点/ cm^2 ，按正六边形排列 ≥ 4.6 触点/ cm^2 （提升局部场电位LFP的空间分辨率）。

非侵入式设备（高密度EEG、fNIRS）：

- a) 通道数 ≥ 64 导（符合扩展10-20/10-10电极布局），关键脑区（前额区、中央区、枕区）通道间距 $\leq 3\text{cm}$ ；
- b) 支持国际标准电极位点（如Fz、Cz、Oz）及功能扩展位点（如M1、TP10），位点覆盖率 $\geq 95\%$ 。

4.1.1.2 采样频率技术规范

信号类型与采样率匹配

信号类型	生理频率范围	最低采样率要求	信号幅值范围
Spike	500-5000Hz	≥10kHz	30-500 μV
LFP	0.1-300Hz	≥1kHz	≤1mV
ECoG	0.1-300Hz	≥1kHz	≤10mV
EEG	0.1-100Hz	≥500Hz	≤10mV
fNIRS	0.01-0.1Hz	≥1Hz	N/A

4.1.2 采集精度

4.1.2.1 信号分辨率要求

4.1.2.1.1 模数转换（ADC）精度

侵入式设备：

- a) ADC 有效位数 ENOB ≥10 位，输入端等效噪声 ≤10μVrms；
- b) 信噪比 ≥2.5 倍，满足单细胞动作电位的幅值分辨需求。

非侵入式设备：

- a) ADC 位数 ≥16 位（医用级设备推荐 24 位），最小可检测信号 ≤1.5μV；
- b) CMRR ≥100dB，抑制工频及生理噪声对低幅脑电的干扰。

4.1.2.1.2 幅值测量误差

对 100μV 标准信号的幅值测量误差 ≤15%（侵入式）或 ≤10%（非侵入式）。

4.1.2.2 噪声控制要求

4.1.2.2.1 噪声基线稳定性

静息态噪声有效值（注 1:截止频率 150Hz 的高通滤波后）

信号类型	静息态噪声有效值阈值（μVrms）
Spike	≤10（高频噪声 ¹ ≤5μVrms）
LFP	≤5
ECoG	≤3
EEG	≤50

基线漂移：连续采集 30 分钟，未经数字高通滤波前基线偏移幅度 ≤5μV/min（侵入式）或 ≤20μV/min（非侵入式）。

2. 工频噪声抑制

50Hz/60Hz 陷波滤波器，输入 ±500μV 正弦信号，衰减 ≥30dB。

4.1.2.3 非线性失真与频率响应

4.1.2.3.1 非线性失真度

输入 10Hz、100 μ V 正弦信号，输出信号总谐波失真 $\leq 0.2\%$ 或 $\leq 1\%$ （非侵入式）。

2. 频率响应范围。

侵入式：150-500Hz（-3dB 截止）内，平坦波动 ≤ 0.5 dB；

非侵入式：0.1-100Hz（-3dB 截止）。

4.1.2.4 电极定位与接触精度（侵入式/半侵入式）

4.1.2.4.1 三维定位精度

立体定向系统定位误差 ≤ 0.5 mm；

术后 MRI/CT 影像验证电极触点与目标脑区解剖位置吻合度 $\geq 95\%$ ，误差 ≤ 1 mm。

4.1.2.4.2 接触阻抗控制

植入时电极-组织接触复阻抗（@1kHz）幅值 ≤ 1 M Ω ；

长期植入后复阻抗（@1kHz）幅值波动不超过 10M Ω 。

4.1.3 信号动态范围

4.1.3.1 幅值范围技术要求

4.1.3.1.1 宽频信号覆盖能力

侵入式设备：

- 最小可检测信号： $\leq 50\mu$ V（Spike 信号幅值）， $\leq 100\mu$ V（局部场电位）；
- 最大不失真信号： $\geq \pm 200\mu$ V（Spike 信号幅值）， $\geq \pm 1000\mu$ V（ECoG）；
- 动态范围： ≥ 20 dB。

非侵入式设备：

- 最小可检测信号： $\leq 5\mu$ V；
- 最大输入信号： $\geq \pm 5$ mV；
- 动态范围： ≥ 60 dB，支持脑电信号（ μ V 级）与生理伪影（mV 级）的共存采集。

4.1.3.2 动态增益与过载保护

4.1.3.2.1 过载恢复能力

- 1)侵入式设备：输入信号超过量程时，恢复至正常采集状态的时间 ≤ 50 ms；
- 2)非侵入式设备：过载恢复时间 $\leq 200\mu$ s，且恢复后 10 秒内基线漂移 $\leq 10\mu$ V。

4.1.3.3 频率响应与幅值线性度

4.1.3.3.1 带通特性

- 1) 侵入式：高通滤波器，250Hz 同时直流耦合，保留神经元静息电位；低通滤波器 7500Hz，-3dB 截止；通带（500-5000Hz 频段）波动 ≤ 1 dB，阻带衰减 ≥ 20 dB。
- 2) 非侵入式：高通滤波器 0.1Hz，去除直流漂移；低通滤波器 100Hz，-3dB 截止；陷波滤波器

50Hz/60Hz, 衰减 ≥ 40 dB, 残留噪声 $\leq 0.5\mu\text{Vrms}$ 。

4.1.3.3.2 非线性失真度要求

输入 1000Hz、100 μV 至 500 μV 范围内的正弦信号, 输出信号 THD $\leq 1\%$ 。

4.1.3.4 直流与低频特性

4.1.3.4.1 直流漂移抑制

输入直流电压 $\pm 100\text{mV}$ 时, 输出信号未经数字高通滤波前基线漂移 $\leq 5\mu\text{V}/\text{min}$ (侵入式) 或 $\leq 20\mu\text{V}/\text{min}$ (非侵入式);

支持交流耦合模式与可选的用于神经元静息电位监测的直流耦合模式。

4.1.3.4.2 低频响应

0.1Hz 信号幅值衰减 $\leq 3\text{dB}$ (非侵入式), 0.01Hz 信号幅值衰减 $\leq 5\text{dB}$ (侵入式), 保障慢波信号的有效采集。

4.1.4 采集设备兼容性

4.1.4.1 电极接口兼容性

4.1.4.1.1 侵入式设备 (ECoG/sEEG/Spike 电极)

标准化接口: 支持国际主流电极阵列接口 (如 Utah 阵列接口、NeuroNexus 探针接口), 单接口通道数 ≥ 16 导, 兼容电极触点间距 0.5-2mm 的高密度阵列;

阻抗自适应: 内置阻抗校准模块, 自动匹配电极-组织接触阻抗 (侵入式 $\leq 10\text{k}\Omega@1\text{kHz}$, 半侵入式 $\leq 5\text{k}\Omega@10\text{kHz}$), 支持电极阵列热插拔, 更换延迟 ≤ 20 秒;

生物相容性: 接口材料符合 ISO 10993 生物相容性标准, 支持长期植入 (≥ 30 天)。

4.1.4.1.2 非侵入式设备 (EEG 头帽/fNIRS 阵列)

电极类型兼容: 支持湿电极 (Ag/AgCl 电极, 阻抗 $\leq 5\text{k}\Omega$)、干电极 (阻抗 $\leq 20\text{k}\Omega$)、柔性电极混合使用;

头型适配: 头帽尺寸覆盖儿童 (头围 45-52cm) 至成人 (52-60cm);

快速连接: 采用磁吸式/卡扣式接口, 单通道连接时间 ≤ 2 秒, 支持 30 分钟内完成 64 导电极安装。

4.1.4.2 多设备同步兼容性

4.1.4.2.1 时间同步机制

1) 硬件同步:

a) 侵入式: 支持有线传输精密时钟协议, 同步精度 $\leq 50\mu\text{s}$; 支持无线传输时钟同步协议, 同步精度 $\leq 50\text{ms}$; 配备 TTL 触发输入/输出接口, 与刺激器、行为记录系统的触发延迟 $\leq 1\text{ms}$;

b) 非侵入式: 时钟同步精度 $\leq 1\text{ms}$; 跨设备时间戳误差 $\leq 1\text{ms}$ 。

2) 多模态触发: 可接收视觉、听觉刺激器、机械反馈装置、振动触觉器的触发信号, 支持事件标记与信号采集的毫秒级对齐。

4.1.4.2.2 跨设备数据交互

- 1) 影像学兼容：与 fMRI、PET 设备联动时，支持梯度伪影抑制算法；
- 2) 外部控制器：脑控场景中，与机械臂、外骨骼等设备的控制信号延迟 $\leq 1s$ 。

4.1.4.3 数据格式与软件兼容性

据格式支持通用格式：兼容 NWB、BDF、EDF、HDF5 等标准化格式，支持电极坐标、采样率、增益参数等元数据结构化存储；提供开发包，支持用户自定义数据解析算法，确保第三方软件的对接。

软件兼容性支持 Windows、Linux、macOS，提供跨平台数据采集处理软件。支持云端存储，兼容 DICOM 3.0 标准。

4.1.4.4 电磁与环境兼容性

4.1.4.4.1 电磁兼容性

1) 抗干扰能力：符合医疗电气设备电磁兼容，在 1V/m 电磁场下，信号噪声无显著增加，在 5V/m 电磁场下，信号噪声增幅 $\leq 10\%$ （侵入式）或 $\leq 15\%$ （非侵入式）；

2) 射频抑制：无线传输设备支持频段隔离，无明显对脑电采集的干扰（噪声增量 $\leq 2\mu V_{rms}$ ）。

4.1.4.4.2 环境适应性

1) 植入式设备：工作温度 35-40°C，湿度 95% \pm 5%，支持 MRI 环境（1.5T 磁场，梯度切换率 $\leq 200T/m/s$ ）；

2) 外置设备：工作温度 10-40°C，湿度 10%-90%，满足医疗设备抗振动、抗冲击要求。

4.1.4.5 临床场景适配兼容性 意见同 4.1.1.1.2

4.1.4.5.1 癫痫监测系统

- 1) 支持与视频监测系统实时并网，同步记录脑信号与临床视频，时间误差 $\leq 30ms$ ；
- 2) 兼容医院与实验室信息系统，自动上传患者基本信息、采集参数至电子病历，符合健康信息学数据交换标准。

4.1.4.5.2 便携式脑电设备

- 1) 支持终端链接，实时显示脑电波形、计算 α/β 波功率等基础指标；
- 2) 内置电池续航 ≥ 8 小时。

4.2 采集性能评价指标与方法

4.2.1 信号完整性评价

4.2.1.1 指标定义

1. 数据缺失率

1) 定义：采集过程中因传输故障、存储错误等导致的有效数据丢失比例，

$$\text{数据缺失率} = \frac{\text{丢失采样点数}}{\text{总预期采样点数}} \times 100\%$$

a) 预期采样点数：标称采样率² \times 采集时间 \times 通道数

- b) 丢失采样点：连续 3 个及以上采样点幅值为 0 或超出量程的异常数据
(注 2:可根据帧格式确定采样率)

2)技术要求:

- a) 侵入式设备：有线传输≤0.001%，无线传输≤0.01%；
- b) 非侵入式设备：有线传输≤0.01%，无线传输≤0.1%。

2. 信号失真

- 1)定义：采集信号与原始输入信号的波形差异程度，通过总谐波失真和均方根误差量化：

a) 总谐波失真： $THD = \frac{\sqrt{\sum_{n=2}^{\infty} A_n^2}}{A_1} \times 100\%$

(A_1 为基波幅值， A_n 为 n 次谐波幅值)

b) RMSE： $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}$

(x_i 为原始信号， \hat{x}_i 为采集信号)

2)技术要求:

- a) 1000Hz、500μV 正弦信号 THD≤1%（侵入式）/≤1%（非侵入式）；
- b) RMSE≤5%峰峰值（侵入式）/≤10%峰峰值（非侵入式）。

3. 噪声抑制比（Noise Suppression Ratio, NSR）

- 1)定义：去噪后噪声功率与原始噪声功率的比值，反映设备固有噪声水平：

$$NSR = 20 \log_{10} \left(\frac{\text{原始噪声 RMS}}{\text{采集噪声 RMS}} \right)$$

- a) 原始噪声：信号发生器输入 0V 时的设备本底噪声；
- b) 采集噪声：实际采集信号中的噪声有效值。

4. 2. 1. 2 评价方法

1. 数据缺失率测试

- 1)测试设备：高精度信号发生器、数据采集设备。

2)测试步骤:

- a) 输入 1000Hz、500μV 正弦信号至所有通道，持续采集 24 小时（侵入式）或 8 小时（非侵入式）；
 - b) 解析数据文件，标记幅值为 0 或超出量程的异常段，计算丢失采样点数；
 - c) 代入公式计算缺失率，要求侵入式≤0.001%、非侵入式≤0.01%（无线传输放宽至 0.1%）。
- 3)合格判定：连续 3 次测试缺失率均达标，且异常数据段无规律性（如非周期性丢包）。

2. 信号失真度测试

- 1)测试设备：高精度信号发生器、数据采集设备或频谱分析仪。

2)测试步骤:

- a) 输入 1000Hz、500μV 正弦信号至所有通道；
- b) 采集 5 分钟信号，分析频谱，计算 THD；
- c) 对比原始信号与采集信号的时域波形，计算 RMSE。

3. 噪声抑制比测试

- 1)测试设备：高精度信号发生器、数据采集设备。

2)测试步骤:

- a) 断开电极输入，采集 5 分钟本底噪，计算原始噪声 RMS；

- b) 接入工频干扰源（50μVrms），采集 5 分钟信号，计算采集噪声 RMS；
- c) 代入 NSR 公式，要求侵入式 NSR≥30dB、非侵入式≥60dB（抑制工频噪声能力）。

4.2.1.3 多通道一致性验证

测试方法：

- 1) 同步输入相同 1000Hz 信号至所有通道，检查各通道波形一致性（相位差≤5°，幅值差异≤2%）；
- 2) 统计异常通道数（失真度>2 倍均值的通道），要求异常率≤1%或≤1 个。

4.2.2 通道同步性能评价（时间同步精度评价方法）

4.2.2.1 指标定义

1. 时间同步偏差

多通道间信号采样时刻的时间差异，通过以下参数量化：

- 1) 最大时间差（ Δt_{max} ）：所有通道中最早与最晚采样时刻的差值，反映同步系统的最坏情况偏差；
- 2) 时间标准差（ σ_t ）：各通道采样时刻与平均时刻的标准差，反映同步系统的稳定程度。
- 3) 技术要求：
 - a) 侵入式设备： $\Delta t_{max} \leq 1\mu s$ ， $\sigma_t \leq 0.5\mu s$ ；
 - b) 非侵入式设备： $\Delta t_{max} \leq 10\mu s$ ， $\sigma_t \leq 2\mu s$ 。

2. 相位一致性

相邻通道在相同频率信号下的相位差，计算公式：

$$\text{相位差} = \arctan \frac{\text{Im}(X_1^* \cdot X_2)}{\text{Re}(X_1^* \cdot X_2)}$$

（ X_1, X_2 为两通道信号的 FFT 结果，单位：度）

- 1) 技术要求：10Hz 信号下，≥95%的通道对相位差≤5°。

4.2.2.2 评价方法

1. 硬件同步精度测试（侵入式设备）

- 1) 测试设备：
 - a) 高精度同步信号发生器；
 - b) 多通道数据采集设备。
- 2) 测试步骤：
 - a) 触发模块生成同步信号，同时发送至所有同步信号采集设备；
 - b) 持续采集 10 分钟，记录每个通道上升沿时间戳和真实上升沿时间戳（分辨率≤1μs）；
 - c) 计算 Δt_{max} （最大时间差）和 σ_t （时间标准差），公式：

$$\Delta t_{max} = \max(t_i) - \min(t_i), \quad \sigma_t = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - \bar{t})^2}$$

- 3) 合格判定： $\Delta t_{max} \leq 1\mu s$ 且 $\sigma_t \leq 0.5\mu s$ ，异常通道数（偏差>2 σ ）≤1%。

2. 软件同步精度测试（非侵入式设备）

- 1) 测试设备：
 - a) 32 导 EEG 电极帽；
 - b) 时间同步信号发送计算服务器。

2) 测试步骤:

- a) 各通道接入同一 10Hz 正弦信号 (幅值 100 μV) ;
- b) 采集 30 分钟, 通过互相关函数计算相邻通道信号的相位差;
- c) 统计相位差>5° 的通道对比例, 要求≤5%;

4.2.3 信号质量评价 (信噪比、噪声基线评价)

4.2.3.1 指标定义

1. 信噪比 (Signal-to-Noise Ratio, SNR)

1) 定义: 信号有效功率与噪声功率的比值, 反映信号中有效成分占比, 计算公式:

$$SNR = 20 \log_{10} \left(\frac{\text{信号 RMS 值}}{\text{噪声 RMS 值}} \right)$$

- a) 信号 RMS 值: 目标信号 (如 Spike、EEG 节律) 在有效频段内的均方根值;
- b) 噪声 RMS 值: 静息态或无刺激状态下的背景噪声均方根值。

2) 技术要求:

信号类型	设备标准
侵入 Spike	≥5:1 (500-5000Hz)
侵入 LFP	≥4:1 (0.1-300Hz)
侵入 ECoG	≥3:1 (0.1-300Hz)
非侵入 EEG	≥2:1 (8-30Hz)

2. 噪声基线稳定性

1) 静息态噪声有效值 (RMS): 无刺激状态下的噪声幅值波动, 单位 μVrms:

- a) 侵入式设备: ECoG/LFP 通道≤5 μVrms, Spike 通道≤10 μVrms;
- b) 非侵入式设备: EEG 通道≤50 μVrms。

2) 基线漂移: 连续采集 30 分钟内, 信号未经数字高通滤波前基线偏移的最大幅度, 单位 μV/min:

- a) 侵入式≤5 μV/min, 非侵入式≤20 μV/min (抑制直流漂移影响)。

3. 工频噪声抑制比

1) 定义: 工频噪声衰减能力, 计算公式:

$$\text{工频噪声抑制比} = 20 \log_{10} \left(\frac{\text{输入工频噪声 RMS}}{\text{输出工频噪声 RMS}} \right)$$

- a) 技术要求: 侵入式≥30dB, 非侵入式≥60dB (残留工频噪声≤0.5 μVrms)。

4.2.3.2 评价方法

1. 信噪比 (SNR) 测试

1) 测试设备:

- a) 高的精度信号发生器;
- b) 频谱分析仪。

2) 测试步骤: (注 3, 根据不同类型信号调整正弦信号频率)

a) 侵入式设备：接入 0.5% NaCl 溶液中的 ECoG 电极，模拟脑组织环境；输入 100 μ V、10Hz 正弦信号³，采集 1 分钟；计算信号 RMS 值（10Hz \pm 2Hz 频段）与噪声 RMS 值（0.5-100Hz 扣除信号频段），代入 SNR 公式，要求 \geq 4:1。

b) 非侵入式设备：受试者静息闭眼，采集 30 秒 EEG 信号（Oz 通道）；通过带通滤波器（8-30Hz）提取 α / β 波信号，计算信号 RMS；噪声 RMS 为扣除信号频段的剩余部分，要求 SNR \geq 2:1。

2. 噪声基线稳定性测试

a) 设备断开所有电极，采集 30 分钟本底噪声；计算噪声 RMS 值（0.5-100Hz），要求侵入式 \leq 5 μ Vrms、非侵入式 \leq 50 μ Vrms；绘制基线漂移曲线，计算最大偏移量，要求侵入式 \leq 5 μ V/min、非侵入式 \leq 20 μ V/min。

3. 工频噪声抑制测试

1) 测试设备：高精度信号发生器。

2) 测试步骤：输入 50Hz 噪声信号至采集通道，开启陷波滤波器；采集 1 分钟信号，计算输出噪声 RMS 值；代入公式，要求侵入式 \geq 80dB（输出 \leq 0.1 μ Vrms）、非侵入式 \geq 60dB（输出 \leq 1 μ Vrms）。

4.2.4 长期采集稳定性评价（设备长期连续运行稳定性评价方法）

4.2.4.1 指标定义

1. 信号漂移率（Signal Drift Rate）

1) 定义：连续采集过程中，信号基线幅值随时间的变化程度，计算公式：

$$\text{漂移率} = \frac{|\text{终点基线值} - \text{起点基线值}|}{\text{起点基线值}} \times 100\%$$

a) 基线值：侵入式静息态信号在 0.5-10Hz 频段的均值或非侵入式 EEG8-30Hz 频段均值。

2) 技术要求：

a) 侵入式设备：8 小时漂移率 \leq 10%，24 小时 \leq 15%；

b) 非侵入式设备：1 小时漂移率 \leq 20%。

2. 设备故障率（Failure Rate）

1) 定义：单位时间内设备异常（如通道失效、数据丢包、过载报警）的发生次数，单位：次/小时。异常判定：单通道数据缺失 $>$ 1 秒、噪声 RMS $>$ 3 倍基线值、阻抗波动 $>$ 50%。

2) 技术要求：

a) 侵入式设备： \leq 0.1 次/小时（支持 72 小时连续监测）；

b) 非侵入式设备： \leq 1 次/小时（支持 8 小时连续采集）。

3. 阻抗稳定性（Impedance Stability）

1) 定义（侵入式/半侵入式）：电极-组织接触阻抗的波动范围，计算公式：

$$\text{阻抗波动率} = \frac{\text{最大阻抗} - \text{最小阻抗}}{\text{平均阻抗}} \times 100\%$$

a) 技术要求：波动率 \leq 20%（1kHz 测试频率下，深部侵入式阻抗 \leq 1M Ω ，颅内半侵入式阻抗 \leq 5k Ω ）。

4.2.5 评价方法

1. 侵入式设备稳定性测试

1) 测试场景：模拟临床长期植入场景。

2) 测试步骤：

a) 植入前校准：电极在 0.9%生理盐水浸泡 24 小时，测量初始阻抗（要求 $\leq 10\text{k}\Omega @ 1\text{kHz}$ ）；输入 10Hz、100 μV 信号，记录初始基线值。

b) 持续采集：接入模拟脑组织环境，连续采集 72 小时，每小时使用 1kHz 恒流法记录各通道阻抗；静息态信号基线值（0.5-10Hz 均值）；设备过载、丢包日志。

c) 数据分析：计算漂移率（终点 vs 起点基线值），要求 $\leq 10\%$ （8 小时）、 $\leq 15\%$ （24 小时）；统计故障率，异常事件 ≤ 0.1 次/小时；阻抗波动率 $\leq 20\%$ ，且无单通道阻抗突增 $>50\%$ 。

2. 非侵入式设备稳定性测试

1) 测试场景：模拟最长实验时长（如脑电反馈训练 4 小时）。

2) 测试步骤：

a) 空载预热：设备开机运行 10 分钟，确保传感器稳定（如湿电极电解液扩散完全）；

b) 负载测试：接入 64 导 EEG 头帽，受试者静息闭眼，连续采集 4 小时；每 30 分钟记录各通道非 $\alpha\beta$ 波段信号 RMS 值；通过信号发生器输入参考信号监测数据缺失率；设备温度（机身表面温度 $\leq 40^\circ\text{C}$ ，避免过热导致漂移）。

c) 合格判定：噪声相对于初始值 RMS 波动 $\leq 30\%$ ；数据缺失率 $\leq 0.1\%$ （无线传输）或 $\leq 0.01\%$ （有线传输）；无因过热/接触不良导致的采集中断。

3. 临床场景验证

1) 癫痫长程监测：

a) 植入式设备在体运行 7 天，对比每日 EEG/ECOG 信号频谱，验证 α 波功率漂移 $\leq 15\%$ ；

b) 统计抗癫痫药物调整前后的信号稳定性，分析棘波识别率，确保设备噪声不影响临床判读。

2) 脑机接口训练：

a) 非侵入式设备连续运行 4 小时（模拟患者训练时长），验证运动想象信号分类准确率下降 $\leq 5\%$ ；

b) 记录电极接触阻抗变化，要求湿电极阻抗波动 $\leq 30\%$ 。

5 数据质量评价

数据质量直接影响模型训练效果、解码性能及其临床或工程应用的可靠性。为确保数据满足高质量标准，从采集、标注、存储及评估多个环节进行系统控制与量化评价。

5.1 数据质量评价基本要求

5.1.1 数据采集与标注过程质量控制要求

标准化的数据采集流程，包括设备校准、设备间同步、信号质量实时监测、采集场景记录等。

标注流程依据标准化执行，包括进行传感器采集标注、多设备采集标注、审核仲裁机制，或使用半自动标注+人工触发的半自动方案。

采集与标注过程全程记录关键元数据包括以下基本信息，以便后续审计和分析：

- 采集时间、
- 人员、
- 设备型号、
- 软件版本、
- 实验参数。

质量控制流程中应包括对数据采集过程中以下内容的审核：

- 数据完整一致性，包括通道间信号时间同步，标注数据时间同步，不同采样率设备数据包同步频次，数据完整性校验，数据包连续性等；
- 信噪比（SNR）；
- 报警记录；
- 掉电、通道脱落、设备断连等异常事件记录；
- 标注一致性抽查或自动对比。

5.1.2 数据一致性与准确性要求

数据应保持时序一致性与跨通道时间同步准确性，尤其在多通道神经数据（如 EEG、ECoG、LFP、Spike）采集中需确保时钟同步误差小于指定阈值（参照本文前面章节）。

标注数据应准确反映实际发生的高精度行为事件、刺激时间点或系统状态转移，与原始神经信号高度对齐，同步误差不低于最低采样率设备的 1 倍采样周期。

可以对不同实验批次的数据分布、特征空间进行统计一致性评估，评估批次间偏移可能导致的模型偏差。

5.1.3 数据溯源性要求

数据文件需具有唯一标识符（如 UUID 或哈希值），并在元数据中记录完整的产生链条信息，包括：

- 数据采集设备与配置参数；
- 操作人员；
- 数据起始 UTC 时间戳；
- 数据格式版本号；
- 每一帧数据包的连续性编号记录；
- 数据文件拆分后的分段位置用于将数据定位到完整数据整体中的精确位置；
- 若在同一文件中保存多个设备同步记录的数据，需要有关联的时间同步信息。

5.2 数据质量评价指标与方法

5.2.1 数据完整性评价（数据缺失率、冗余率）

5.2.1.1 数据缺失率

定义：单位时间内应采集的数据样本数与实际采集样本数之差占比。

计算方法：

$$\frac{\text{期望样本数量} - \text{实际样本数量}}{\text{期望样本数量}} \times 100\%$$

5.2.1.2 数据冗余率

定义：除去有效数据字节后数据中存在的重复片段、无效零值、缓冲重复写入等冗余信息占比。

计算方法：

$$\frac{\text{总数据文件字节大小} - \text{有效数据字节数}}{\text{总数据文件字节大小}} \times 100\%$$

5.2.2 数据一致性评价（跨通道、跨批次一致性评价方法）

5.2.2.1 spike 发放率

对于 SPIKE 类型信号，

定义：单位时间内单个通道有效 SPIKE 数量。

计算方法：

$$\frac{\text{统计范围内有效 Spike 数量}}{\text{统计的数据时长 (秒)}} \times 100\%$$

5.2.2.2 一致性评价

根据固定时长进行发放率的分段计算，统计获得发放率的平均值和波动的标准偏差。

平均发放率计算方法：

$$\frac{\sum \text{发放率}_i}{\text{总分段数}} \times 100\%$$

发放率波动无偏标准偏差计算方法：

$$\sqrt{\frac{\sum (\text{发放率}_i - \text{平均发放率})^2}{\text{总分段数} - 1}}$$

对于同一通道不同时期，平均发放率波动标准偏差应低于 0.6 倍平均发放率。

5.2.3 数据准确性评价（标注数据与实际信号吻合度评价）

对标注行为数据准确性，对比其时间戳与实际生理信号事件或行为触发之间的对齐程度。

评价方法参照本文 4.2.2.2 硬件时间同步对齐测试方法。

标注时间偏差均值/方差参照 5.2.2.2 中计算方法。

5.2.4 数据可复现性评价（数据可复现性及一致性评价方法）

记录的数据可复现性包含两个评价标准：

- 1) 数据记录可以模拟回放，并支持任意时间定位播放，并保证不同通道同时播放时的时间戳对齐。
- 2) 在不同批次重复实验后，数据特征保持一致性，参照 5.2.2.2 中一致性定义。

6 特征提取与解码模型性能评价

6.1 特征提取性能评价

特征提取是脑机接口系统中连接原始信号与解码模型的中间处理环节。高质量特征应具有直观可判别能力、稳定性好、抗干扰性强等要求。

6.1.1 特征有效性（分类或解码性能提升效果评价方法）

定义：特征对解码或分类性能的提升能力。

针对两种不同的解码模型种类，有以下对应的评价指标：

1) 分类任务：模型进行离散的分类种类输出预测

准确率：定义为分类正确的样本数占总样本数的比例。

计算方法：

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

其中：

TP：正确的真类别数量

TN：正确的假类别数量

FP：错误的真类别数量

FN：错误的假类别数量

精准率：定义为预测成真类别中实际正确的真类别的比例

计算方法：

$$\frac{TP}{TP + FP} \times 100\%$$

召回率：定义为实际为真类别的样本中被正确识别为真类别的比例

计算方法：

$$\frac{TP}{TP + FN} \times 100\%$$

F1-分数：定义为精确率与召回率的调和平均值

计算方法：

$$2 \cdot \frac{\text{精准率} \times \text{召回率}}{\text{精准率} + \text{召回率}} \times 100\%$$

2) 回归任务：模型进行连续型取值的预测输出的任务，如运动轨迹、力、位置等连续变量的预测任务

均方误差 MSE：定义为预测值与真实值的平方误差的平均值。

计算方法：

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

其中：

n 为统计时间范围内总的采样点个数

y_i 为实际采集的采样点 i 的标注数据值

\hat{y}_i 为模型预测对应该采样点 i 的数据值

决定系数 R^2 : 定义为预测值与真实值的平方误差的平均值。

计算方法:

$$1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

其中:

\bar{y} 为统计时间范围内实际采集的标注数据的平均值

当该值为 1 时, 代表特征预测与实际**完全符合**。

当值为 0 时, 代表特征预测与直接**使用恒定均值**作为预测输出效果一致。

当值小于 0 时, 代表特征预测**不如**直接使用恒定均值作为预测输出。

6.2 信号解码模型性能评价

6.2.1 模型实时性与计算效率评价 (延迟、处理效率等指标定义)

定义: 模型在处理新输入时的响应速度与资源消耗水平。

评价的关键指标有:

输出延迟:

定义: 对于模型输入的最小单元数据块所对应时间区间的末尾时刻, 到模型输出的预测结果的时刻的间隔。

对于模型实时进行预测的过程中, 平均输出延迟定义为一段时间区间内所有预测结果输出延迟的平均。建议平均输出延迟小于系统最大容忍延迟, 且应小于 100ms;

吞吐率:

定义: 单位时间内处理数据量(如每秒所有通道中所处理的 Spike 的个数); 建议系统吞吐率应达到 100/通道/秒。

计算资源消耗: 包括 CPU/GPU 使用率、内存占用; 实际应以所实现的**输出延迟**和**吞吐率**作为实时性的评价指标。

部署测试: 模型实时性评估需提供在实际部署平台(嵌入式设备、边缘设备、或平板电脑)上进行以上实时运行效率的评估。

6.2.2 模型稳定性与可重复性评价 (重测信度、稳定性分析方法)

定义: 模型在多次训练或不同时间点使用时对同样的数据集是否能稳定输出一致的解码结果。

评价方式:

通过将同一次数据集进行回放并调用模型进行预测输出, 除去模型启动初始化时段后, 应获得一致的解码输出。

6.2.2.1 随机初始化敏感性

定义: 通过设置不同的随机数初始化种子, 验证模型输出预测在除去模型启动初始化时段后, 是否获得一致的解码输出。

一致性计算方式参照 6.1.1 中分类任务或回归任务的评价指标。

6.2.2.2 输入扰动模型鲁棒性

定义：通过人为添加对原始输入信号的扰动，验证模型输出预测在除去模型启动初始化时段后，是否获得一致的解码输出。

添加扰动的方式包括有：

- 删除或添加 1%~5%原始通道发放率的 Spike
- 按 1%~5%的概率将 Spike 时间进行服从均值 0ms，标准偏差为 1.5ms 的方式进行扰动
- 去除某些通道

鲁棒性的评价计算指标参照 6.1.1 中分类任务或回归任务的评价指标。

7 数据与性能评价规范化要求

为确保脑机接口系统性能评价的可比性、可复现性，制定统一的数据集规范、评价环境与报告格式。

7.1 评价数据要求

7.1.1 评价数据集标准化要求

数据集应覆盖目标应用的典型场景，包括：

- 侵入式脑机系统
- 非侵入式

数据集应按不同信号采集类型进行区分，有：

- Spike
- ECoG/sEEG
- LFP
- EEG

标准化数据集中应至少包含：

- 多个被试；
- 多个会话/天次；
- 多通道神经信号与对应标注；

数据应符合系统采样率、信噪比、通道数要求；

数据结构应统一，参照 5.1 数据质量要求。

评价数据集标准化要求包含 数据集任务说明文档。

7.1.2 数据标注一致性要求

用于模型解码训练的所有标注应：

- 有明确定义与类别说明；
- 遵循统一标注流程；
- 提供标注文档与代码；
- 每个数据帧单元有精确到微秒精度 1μ Sec 为单位的 UTC 时间戳；

- 建议采用固定采样率方式进行标注；
- 数值应明确是否采取了归一化系数保存和相关的对应物理量关系

7.2 性能评价方法要求

7.2.1 性能评价环境要求

明确模型测试的所需平台环境，包括：

- 操作系统、CPU/GPU 型号、内存；
- 框架版本（PyTorch, TensorFlow 等）；
- 依赖库版本；

对于实时性能，明确部署模型，并应在目标运行平台下测试，包括嵌入式硬件、边缘计算平台或平板电脑等。

7.2.2 性能指标计算方法标准化

所有性能指标应使用统一公式，除去模型启动初始化时段后，相同计算起止点，且长度不应少于 3 分钟。

性能指标结果中应明确：

- 所用的数据集；
- 所用的实时部署平台硬件规格；
- 信号起始与终止时间的定义；
- 数据划分比例（训练/验证/测试）及训练最大资源需求。
- 指标评估方式；

鼓励提供开源的性能指标计算工具或脚本。

7.3 结果报告规范要求

7.3.1 评价数据集标准化要求

7.3.1.1 数据来源规范

1. 设备与采样参数明确

- a) 记录采集设备型号、厂商、操作系统与软件版本；
- b) 明确电极配置，包括但不限于电极类型（皮层/深部/皮下）、电极材料（铂金、银氯化银等）、阵列布局（二维/三维排列）、通道数及电极编号规范；
- c) 记录电极阻抗范围及测量值，用于信号质量评估；
- d) 采样率不得低于 1 kHz，明确是否采用了抗混叠滤波、去伪影等处理方式。

2. 实验与采集流程标准化

- a) 实验任务应标准化设计，实验数据应在规范流程下采集，明确实验任务设计（如刺激类型、呈现时长、被试操作提示（如反应要求）、采集环境控制（噪声、温度、光照）等；
- b) 记录全部实验关键事件的时间戳（如刺激呈现、按键反应、语音起始等），确保事件和神经信号的时序同步；
- c) 实验过程应保留原始日志、提示记录、故障注释等，支持后期审计与重构。

7.3.1.2 数据预处理标准

1. 处理流程记录要求

明确每一步处理操作，包括：

- a) 信号滤波方式（如带通滤波范围、工频抑制方法）；
- b) 伪迹剔除方法（如信号跳变检测、人工标记）；
- c) 重参考方式（如共平均参考，单通道参考等）；
- d) 信号重采样参数（如原始 2000Hz → 500 Hz）；
- e) 时序校正（如刺激同步修正 ± x ms 等）；

所有处理过程应配套参数设置文件与日志输出。

2. 版本控制与存储分离

- a) 原始数据与预处理数据应该分开保存，采用不同文件夹/路径并加注版本信息；
- b) 建议命名规范如 SubjectID_raw.edf, SubjectID_preproc_v1.edf，并在元数据中标注处理链描述。

3. 工具与可重复性要求

- a) 所有预处理建议使用自动化脚本完成，记录执行环境（如 Python/MATLAB 版本，库依赖）；
- b) 若使用 GUI 商业工具（如 BrainVision Analyzer），应附设配置文件、批处理脚本或操作记录；
- c) 建议封装为容器（如 Docker）便于环境复现。

7.3.1.3 人群多样性与代表性要求

1. 样本多样性

数据集应涵盖目标使用人群的多个维度，包括：

- a) 性别（如男女比例近似平衡）；
- b) 年龄段（如儿童、青年、老年）；
- c) 疾病状态（如癫痫、意识障碍、脑瘤等）；
- d) 文化背景或语言类型；
- e) 电极植入区域/方式（Spike/ECOG/SEEG/EEG 等）；

每位被试应提供脱敏的人口学基本信息。

2. 代表性与覆盖性评估

- a) 建议报告数据集中各类个体的比例分布；
- b) 若应用系统声称适用于某特定人群，应提供该人群样本不少于总数的 20%。

7.3.1.4 数据格式与结构化规范

1. 信号数据格式

1) 推荐采用标准格式，如：

- a) Brain Imaging Data Structure with intracranial EEG (BIDS-iEEG)；
- b) European Data Format (EDF/EDF+)；
- c) 若使用自定义格式，必须提供详尽的格式说明文档与元数据字段定义，以及读取示例脚本；

2) 信号数据应包含：

- a) 通道名称、顺序和位置；
- b) 采样率 (Hz)、单位 (μV、mV)；
- c) 参考电极说明；

d) 信号时间起点与持续时长。

2. 元数据文件要求

- 1) 提供 JSON/YAML/XML 格式的独立元数据文件；
- 2) 包含字段：
 - a) 采集参数（设备、电极类型、滤波范围）；
 - b) 被试信息（去标识化）；
 - c) 任务结构与刺激信息；
 - d) 文件版本与更新时间戳；

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

3) 文件命名建议采用统一规范，如 BCI-YYMMDD-SubjectID-SessionID-iEEG.edf, BCI-YYMMDD-SubjectID-SessionID-iEEG.json。

7.3.1.5 数据合法合规性要求

1. 伦理与隐私遵循

- 1) 所有采集工作必须获得本地伦理委员会审批（附伦理批号）；
- 2) 被试需签署书面知情同意书（Informed Consent），并允许数据用于科研/注册目的；
- 3) 所有数据在使用前须进行匿名化与脱敏处理，移除身份证号、姓名、医疗号、住址、详细的出生日期等直接识别信息。

2. 使用授权与申明

- 1) 建议在元数据中标注许可范围（如 CC-BY-NC、仅限科研用途、仅限项目注册）；
- 2) 商业用途数据需明确授权路径与限制条件。

7.3.1.6 数据存储规范

1. 存储结构与数据类型分类

- 1) 明确区分以下数据类别，按模块化组织与存储：
 - a) 脑电原始信号（Spike/ECOG/SEEG/EEG）；
 - b) 标注文件（事件触发、信号片段）；
 - c) 元数据文件；
 - d) 多模态同步信号（EMG、EOG、ECG、行为响应）；
 - e) 背景脑电/静息状态；

2. 命名与压缩规范

- 1) 所有文件命名应具备唯一标识；
- 2) 若对数据进行了压缩，应说明压缩与解压缩方法与工具，一般情况下建议使用无损压缩方式。

3. 匿名化处理

- 1) 文件名与路径不得包含被受试者真实身份信息；
- 2) 文件内部（如 EDF header）应使用代号替换真实姓名。

4. 备份与安全管理

- 1) 应规划数据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、损坏或其他灾害的影响。如使用 RAID、对象存储或副本云存储保障数据可靠性；
- 2) 数据集存储文件的安全保护应符合 GB/T 39725-2020 要求，明确数据存储有效期限、访问权限控制与加密机制，确保数据存储和使用过程的合法性和合规性。

5. 数据版本管理

- 1) 所有数据应具备版本标识（如 v1.0, v1.1），变更记录说明；

2) 对应元数据中记录生成时间与修改时间, 支持差异追踪。

7.3.1.7 数据集质量控制指标

为确保用于性能评价的数据具备高质量, 应综合考虑以下维度:

1) 完整性

数据应无缺失, 包含所有必需字段与文件; 时间序列应连续无丢包;

2) 一致性

同一类型信号格式、命名、单位应一致, 文件结构应规范统一;

3) 准确性

标注时间点误差应 $\leq 50\text{ms}$, 事件触发应与神经信号对齐; 、

4) 时效性

数据的采集、标注、流转、归档和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。

5) 唯一性

每条记录具备唯一标识符, 避免重复样本或冲突命名;

6) 代表性

被试特征应能覆盖目标适用人群的关键变量维度;

7) 可访问性

数据及元信息可在授权范围内检索、提取, 接口开放且文档完备;

8) 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。所有数据应支持从预处理结果回溯到原始记录、处理脚本与日志;

9) 资源利用性

数据体积与信息密度应合理, 避免冗余; 支持按任务、通道、标签等维度解构提取;

10) 可移植性

支持跨平台读取 (Windows/Linux/Mac), 兼容主流分析工具 (Python/MATLAB), 数据质量不应随平台和系统发生变化;

7.3.2 数据标注一致性要求

1. 标注层级要求:

1) 原始信号时间戳标注

- a) 包含所有实验事件触发点 (如刺激呈现、动作开始等);
- b) 建议采用事件触发点 $\pm 50\text{ms}$ 窗口范围, 表示对任务响应时间窗的参考;
- c) 应记录原始时间戳 (如 ms 或采样点数) 并对齐至统一时间轴 (如相对试次七点或全程零时刻)。

2) 任务类别标签

- a) 对每个任务试次或时间段标注其对应的语义类别, 如“语音/非语音”、“左手/右手运动”、“拼音音节名称”等;
- b) 标签应具备层级结构 (如任务类型>子任务>条件), 便于多粒度分析。

3) 噪声段标记

- a) 明确标注由于肌电、眼动、工频干扰等造成的高噪声片段;
- b) 标注格式应支持连续区段 (如[开始事件, 结束时间]) 或离散事件类型。

2. 标注流程规范

1) 标注工具与格式

- a) 推荐使用结构化、时序支持强的工具进行标注，如 MNE-Annotator、ELAN 等；
- b) 标注输出应为结构化文件（如 .json, .tsv, .mat, .edf+annotations），便于分析与可视化。

2) 人员培训与复核机制

- a) 所有标注人员须接受专业神经科学与任务设计相关培训；
- b) 推荐采用“双人独立标注+专家复核”机制，提升标注一致性与准确性；
- c) 标注流程应留存日志，包括标注者、版本、修改记录等元信息。

3) 语义标签管理体系

- a) 所有标签名称、含义、单位、层级关系须文档化，并提供可机读版本（如标签字典 JSON/XML）；
- b) 应支持标签与其他数据集之间的对照映射（如跨项目分析、标准共享数据集等）。

3. 标签一致性评估指标及计算方法

1) Cohen’s Kappa (科恩卡帕系数)

- a) 适用范围：用于两位标注者对同一批神经信号片段的事件类别或标签的一致性评价，适用于原始事件标注、任务标签标注等。至少选取 100 段样本事件进行双人独立标注并评估。
- b) 计算公式：

$$k = \frac{p_o - p_e}{1 - p_e}$$

(p_o 为观察一致率， p_e 为理论随机一致率)

- c) 推荐阈值（基于 Landis & Koch, 1977）

k值范围	一致性等级	技术建议
< 0.4	低	不可用于性能评估
0.4 – 0.6	中等	可用于探索性评估
0.6 – 0.8	良好	可接受
> 0.8	极好	推荐

- d) 技术要求：Cohen’s Kappa 值不应低于 0.6，建议 0.75 以上。

2) 标签覆盖率

- a) 适用范围：衡量神经数据时间轴上标注的完整性。
- b) 计算公式：
 - i) 连续时段标注覆盖率计算，适用于连续时间任务，如发音、运动持续等。

$$Coverage_{time} = \frac{T_{labeled}}{T_{total}}$$

($T_{labeled}$ 为被标注的有效时间总长度， T_{total} 为总记录时长)

- ii) 离散时间标注覆盖率计算，适用于刺激事件、字词呈现等定时任务。

$$Coverage_{event} = \frac{N_{labeled}}{N_{total}}$$

- c) 技术要求建议：对于连续任务，标签覆盖率建议不低于 90%，对于离散事件任务，标签覆盖率建议不低于 95%。两类覆盖率均应在评价报告中明确标注并分别计算。

3) 标注延迟偏差

- a) 适用范围：衡量多个标注者对同一事件的起始时间（或结束时间）标注之间的差异平均值，用于评估标注者间时间标注的一致性；
- b) 计算公式：

$$Deviation_{time} = \frac{1}{N} \sum_{i=1}^N |t_i^{(1)} - t_i^{(2)}|$$

(N 为标注者数量, $t_i^{(1)}$ 、 $t_i^{(2)}$ 为两位标注者对同一事件的起始或结束时间标注)

- c) 技术要求: 若为语音、运动任务的时间标注, 建议延迟偏差 ≤ 50 ms, 若为视觉刺激、纯事件触发类任务, 应控制在 10 – 20 ms。

7.4 性能评价方法要求

7.4.1 性能评价环境要求

1. 硬件基准要求

- 明确实验记录使用的硬件平台, 包括 CPU/GPU 型号、内存容量、存储配置;
- 对涉及实时性解码的系统, 需标注系统负载情况与延迟测量方法。

2. 软件环境记录

- 记录操作系统类型及版本(如 Ubuntu 22.04, Windows 10), 主要运行库及依赖工具(如 PyTorch 2.0, scikit-learn 1.3.0, MATLAB R2022b);
- 建议在元数据中以 YAML/JSON 格式自动导出依赖清单(如 requirements.txt 或 environment.yml)。

3. 干扰控制措施

- 需明确是否启用 GPU 并行加速、随机种子设置;
- 多模型运行或多用户环境中应隔离评估进程, 避免评估数据污染。

4. 运行环境封装。推荐使用容器化工具(如 Docker)封装完整的训练/测试流程, 支持环境导出与运行脚本共享, 支撑第三方复现;

7.4.2 性能指标计算方法标准化

1. 分类任务(如音节/运动意图分类)

指标	计算公式简要	说明
准确率 Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	总体正确率
F1 score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$	类别不平衡下更稳健
混淆矩阵	各类别的 TP, FP, FN, TN 数量表格	可视化分类混淆模式
AUC-ROC	基于概率输出的接受者操作特征曲线下及其曲线下面积	二分类下性能综合评估

2. 回归/连续解码任务(如时序电极信号/语音频谱预测)

指标	计算公式简要	说明
均方根误差 (RMSE)	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	衡量预测误差绝对值
皮尔逊相关 系数	$\rho = \frac{Cov(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}}$	预测信号与目标信号相关性
解码延迟	首次有效输出相对于事件触发点的延迟(单位: ms)	实时 BCI 系统必报指标

3. 指标统计汇报要求

- 所有性能指标必须汇报均值 \pm 标准差; 若使用交叉验证, 应报告所有折结果与平均值;

- 2) 推荐同时提供置信区间（如 Bootstrap 95% 置信区间）或显著性统计；
- 3) 对比多个系统或模型应进行显著性检验，如 permutation test（随机重排法）等。

7.4.3 鲁棒性验证

通过鲁棒性评估解码模型在输入扰动、通道缺失、信号污染等异常条件下的性能保持能力。

1. 评估情景设计

- a) 通道扰动：随机遮盖或删除部分电极通道（如 10%，30%，50%）；
- b) 噪声注入：向信号中加入高斯噪声（SNR 控制）、肌电干扰模板等；
- c) 任务变化：更换任务类型或指令节奏，测试任务迁移能力；

2. 评估指标：

- a) 性能下降比率（Degradation Rate）：

$$DR = \frac{P_{clean} - P_{noisy}}{P_{clean}}$$

越低表示鲁棒性越好。

7.4.4 可重复性保障

1. 代码与模型共享

- a) 提供完整的训练/测试代码；
- b) 明确模型结构、损失函数、超参数设定等关键训练信息；

2. 数据划分与随机性控制

- a) 所有交叉验证或划分方法必须设定并报告随机种子；

3. 再训练与复现误差容差

- a) 对于神经网络等存在训练波动的模型，应说明性能上下限与可接受的偏差范围；
- b) 推荐使用标准数据集（Benchmark Dataset）+参考结果作为“复现锚点”；

7.5 结果报告规范要求

7.5.1 评价结果报告模板与内容规范

性能评价报告应包括如下关键组成部分，确保报告结构清晰、逻辑完整，便于对系统技术能力进行全面评估：

1. 摘要

- a) 简要概述系统结构、评估目的与主要结论
- b) 包含设备名称、型号、软件版本、算法框架、评价时间及报告编号。

2. 数据与实验方法

- a) 数据来源：注明采集来源、被试数量、总记录时长、采集任务类型；
- b) 数据预处理：详细描述滤波、降噪、重参考、时序校正及标注处理等步骤；
- c) 特征提取与建模流程：说明提取特征类型（时域、频域、时频域）、建模架构、超参数设定；可视化展示特征有效性（如基于 SHAP 值的归因热图）；
- d) 数据划分策略：报告训练-验证-测试集划分比例或交叉验证方式；明确是否设计随机种子以控制实验可重复性；
- e) 性能指标：列出所用评价指标及计算方式（如准确率、F1 Score、RMSE 等），并注明计算时间窗或采样单位。

3. 评价结果和系统性能展示

使用标准化图表与表格呈现关键结果，包括但不限于：

- a) 解码性能指标均值±标准差（或 95%置信区间）；
 - b) 解码延迟分布直方图（基于 1000 次采样）；
 - c) 跨日重测性能对比（间隔 24h 后的性能一致性）；
 - d) 鲁棒性测试结果（如故障注入-性能曲面）；
 - e) 长期趋势分析图（按周/月跟踪解码性能变化）。
4. 统计检验与显著性分析
- a) 说明是否进行显著性检验，采用的检验方法（如配对 t 检验、非参数检验、置换检验、Bootstrap 置信区间等）；
 - b) 报告 p 值、效应量、显著性结论；
 - c) 若进行多模型对比，应注明控制假设数量与校正方法（如 FDR、Bonferroni 校正）；
5. 局限性分析
- a) 说明模型性能下降的边界条件（如低 SNR、通道缺失，跨任务泛化）；
 - b) 报告模型失效模式（如对特定标签混淆严重，过拟合风险等）；
 - c) 对样本代表性或任务设置可能带来的偏差提出说明；
6. 附录
- a) 数据集访问方式：数据下载链接或托管平台信息；
 - b) 处理代码仓库地址与 Git commit ID；
 - c) 模型权重文件说明与链接；
 - d) 脱敏处理后的受试者人口统计信息摘要（性别、年龄分布、利手习惯等）；
 - e) 伦理审批编号与批件复印件（扫描件）；
 - f) 使用授权类型说明（科研用途、商业用途、注册用途等）；

7.5.2 报告结构化输出与可登记性

1. 提供结构化电子报告（如 JSON、XML、YAML）输出文件，字段建议包括：
 - a) 报告编号、出具单位、生成时间、设备型号与软件版本
 - b) 性能指标字段及结果数值；
 - c) 数据与模型版本信息（数据集 ID、模型配置 ID、Git Hash 等）；
 - d) 审查状态与签章信息（如数字签名/审查记录号）；
2. 结构化格式应被监管平台接收用于在线登记、证据审核或版本溯源。

7.5.3 性能评价报告审查与验证流程

1. 第一方验证（提交方自查）
 - a) 提供完整的数据采集、处理流程、日志、模型代码及原始评估结果；
 - b) 保证数据未污染测试集、模型无手动调优至测试集性能最优。
2. 第三方验证（独立审计）
 - a) 委托具有 CMA/CNAS 资质的实验室或国家科研平台开展审查；
 - b) 审计方式包括盲测测试集、随机重训练复现、性能对比分析等；
 - c) 应使用冻结模型与版本锁定数据集，确保测试集未泄露。
3. 版本冻结与追踪
 - a) 所有报告应绑定对应的数据版本、模型结构与代码哈希；
 - b) 代码及模型仓库应使用 Git 或类似版本控制工具，记录变更轨迹。

7.5.4 更新机制与持续监测

GB/T XXXX-XXXX

1. 对于已纳入注册目录的医疗器械产品，建议每年提交一次性能更新报告；
2. 更新内容应包括新增数据样本上的再训练或泛化测试结果；
3. 若存在系统升级（如模型架构更替、采集设备更换），应提供新版性能比对与显著性验证；
4. 所有更新报告需保留原始模型与新模型间差异对照，并提交完整文档记录。

