

智能医疗

基于心电图数据的多中心心率失常研究

本赛道赛题属于总赛赛题之一，优胜者团队将获得奖金，并有机会派代表参加线下活动（具体安排另行通知）。

1、场景介绍

心律失常是指心脏电传导系统异常所引起的各种征候，泛指任何不正常的心跳或心律问题。当协调心跳的电信号不能正常工作时，就会发生心律失常问题，如心动过速、心动过缓或心动不规则。心律失常对人体的危害是非常大的。轻者仅表现出轻微的头晕、心慌等不适感，重者症状较为严重，可发生猝死。因此需要对早期预防和控制症状较轻的心律失常，避免频繁的轻症心律失常转化为重症甚至恶性的心律失常。心电图是常用的心律失常诊断手段，本赛题中参赛队伍基于隐私保护计算技术，在数据安全和隐私保护的前提下，建立心率失常分类模型，通过利用多中心心电图数据的联合分析，提高心率失常多分类模型的准确性和特异性，推动心血管疾病的早期筛查与预防体系的建立。

2、赛题任务

- **计算参与方：**发起方：某中心；数据参与方：5 个数据中心；中间方：计算协同方（根据具体的技术路线（MPC/HE/TEE）不同参赛队伍可选是否加入计算协同方）；
- **数据输入：**
 - 为参赛队伍提供 5 个数据中心的心电数据，供解决方案开发使用。数据特征见附录一。
 - 每个中心有相同的 12 导联心电图数据，数据来自不同的病人，参赛队伍根据自身情况，采用横向安全联邦学习分析方法，以下表所列的心率失常多分类结果作为最终目标。

心率失常种类	DX Code
--------	---------

Atrial Fibrillation	164889003
Atrial Flutter	164890007
Atrial Tachycardia	713422000
Sinus Bradycardia	426177001
Sinus Rhythm	426783006
Sinus Tachycardia	427084000
Supraventricular Tachycardia	426761007

- 其中样本数据的 DX 诊断码里面可能包含心率失常以外的其他心血管相关疾病的 DX 码，参赛队伍可以忽略上表以外的 DX 码。
- **目标输出：**基于多中心数据训练的心率失常多目标分类模型和验证结果。
- **技术要求：**
 - 参赛队伍根据自身技术背景选取：MPC，HE，TEE 中的一种技术，通过与横向联邦学习结合来解决赛题中的相关问题。
 - 如使用 TEE，测评采用 Intel SGX V2 TEE（可信执行环境）。中间方拥有 TEE 计算资源，模型聚合在中间方 TEE 内完成。
- **安全性要求：128bits 安全性**
 - 对于同态加密（HE）解决方案：参赛队伍请参考国际同态加密标准白皮书中推荐的参数设置。

<http://homomorphicencryption.org/wp-content/uploads/2018/11/HomomorphicEncryptionStandardv1.1.pdf>

- 对于多方安全计算：可以使用乱码电路或秘密分享算法，如使用秘密分享算法，需要明确声明（ t, n ）阈值密钥分享的设定。
https://en.wikipedia.org/wiki/Secure_multi-party_computation
- 对于可信计算环境（TEE）：参赛队伍请参考 SGX Remote Attestation 协议完成动态密钥交换。本赛题暂不考虑侧信道攻击问题。
- **安全假设：**

- 如采用可信计算环境（TEE），需要支持恶意模型假设
- 如采用 MPC/HE，需要支持半诚实模型假设
- 隐私保护目标：
 - 数据参与方均无法获得其他方的原始数据。
 - 各方之间交换的信息需要通过 MPC/HE/TEE 中的一种技术进行保护。
 - 中间方不能获得数据参与方的原始数据或者计算过程中的明文统计信息。
- 评测原则：
 - 评测中，我们会对基于半诚实模型假设的技术路线（MPC 和 HE）与基于恶意模型假设的技术路线（TEE）进行分开测评
 - 解决方案测试时，会使用另一套未公开的数据集进行测试（数据样本量在 50K 左右）。
 - 结果的正确性（训练好的模型在验证数据集上的 F1 score）
 - 联合计算所需的总时间（包括预处理、联合计算等）。
 - 计算过程中的内存使用量峰值。
 - 计算过程中的网络通信总流量。
- 评测环境：
 - 本地计算节点：提供相对高配的计算型云主机配置，暂定通过 KVM 部署，KVM 计算环境的参考配置为：4 核心 CPU，32 GB 内存，500GB 硬盘。
 - 如使用 TEE 聚合节点：支持 SGX V2 的 TEE 服务器，至少 16GB EPC 内存，可以使用多线程。（暂定通过 KVM 部署，KVM 计算环境的参考配置为 4 核心 CPU，32 GB 内存，500GB 硬盘）
 - 节点之间网络为虚拟网络带宽限制为 1G
 - 节点：数据参与方（5 个数据中心）；中间方（如使用 TEE，作为计算协同方拥有 TEE 算力）
- 解决方案提交要求：
 - Training
 - 发起方发起训练
 - 各数据参与方输入：训练数据集，按格式，带有 label，通过横向联邦学习结合多方安全计算（MPC）或同态加密技术（HE）或可信执行环境（TEE）中的一种技术完成计算。

- 。
- 中间方输出：加密后的 model，发给发起方
- 发起方输出：解密后的 model
- Testing
 - 发起方输入：model 和 测试数据集，按格式，不带 label 加密后发给中间方，进行计算。
 - 中间方输出：加密的预测 label，发给发起方。
 - 发起方输出：解密后的预测 label
- 详细的 Readme 和参赛方案设计文档 pdf
- 用于 evaluation 的 script
- 参赛选手提交到平台方案需要在 24 小时内完成运行
- 更多解决方案提交方式、日志和接口格式等待后续通知

附录一：数据格式

数据集	备注																																																																																																									
*.hea	JS00001 12 500 5000 23-Mar-2021 20:20:47 JS00001.mat 16+24 1000/mV 16 0 -254 21756 0 I JS00001.mat 16+24 1000/mV 16 0 264 -599 0 II JS00001.mat 16+24 1000/mV 16 0 517 -22376 0 III JS00001.mat 16+24 1000/mV 16 0 -5 28232 0 aVR JS00001.mat 16+24 1000/mV 16 0 -386 16619 0 aVL JS00001.mat 16+24 1000/mV 16 0 390 15121 0 aVF JS00001.mat 16+24 1000/mV 16 0 -98 1568 0 V1 JS00001.mat 16+24 1000/mV 16 0 -312 -32761 0 V2 JS00001.mat 16+24 1000/mV 16 0 -98 32715 0 V3 JS00001.mat 16+24 1000/mV 16 0 810 15193 0 V4 JS00001.mat 16+24 1000/mV 16 0 810 14081 0 V5 JS00001.mat 16+24 1000/mV 16 0 527 32579 0 V6	12 导联心电图样本数据介绍信息																																																																																																								
	#Age: 85 #Sex: Male #Rx: Unknown #Hx: Unknown #Sx: Unknown	与本次赛题无关																																																																																																								
	#Dx	诊断码，样本数据的 DX 码里面可能包含心率失常以外的其他心血管相关疾病的 DX 码，参赛队伍可以忽略与心率失常以外的 DX 码																																																																																																								
*.mat	<div>12 x 5000</div> <table><tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>4998</th><th>4999</th><th>5000</th></tr><tr><td>1</td><td>-254</td><td>-254</td><td>-254</td><td>-254</td><td>-34</td><td>24</td><td>5</td></tr><tr><td>2</td><td>264</td><td>264</td><td>264</td><td>264</td><td>-68</td><td>-49</td><td>-34</td></tr><tr><td>3</td><td>517</td><td>517</td><td>517</td><td>517</td><td>-34</td><td>-73</td><td>-39</td></tr><tr><td>4</td><td>-5</td><td>-5</td><td>-5</td><td>-5</td><td>54</td><td>15</td><td>15</td></tr><tr><td>5</td><td>-386</td><td>-386</td><td>-386</td><td>-386</td><td>0</td><td>49</td><td>20</td></tr><tr><td>6</td><td>390</td><td>390</td><td>390</td><td>390</td><td>-54</td><td>-63</td><td>-39</td></tr><tr><td>7</td><td>-98</td><td>-98</td><td>-98</td><td>-98</td><td>-24</td><td>-15</td><td>5</td></tr><tr><td>8</td><td>-312</td><td>-312</td><td>-312</td><td>-312</td><td>595</td><td>590</td><td>605</td></tr><tr><td>9</td><td>-98</td><td>-98</td><td>-98</td><td>-98</td><td>137</td><td>132</td><td>142</td></tr><tr><td>10</td><td>810</td><td>810</td><td>810</td><td>810</td><td>-205</td><td>-200</td><td>-171</td></tr><tr><td>11</td><td>810</td><td>810</td><td>810</td><td>810</td><td>-200</td><td>-195</td><td>-166</td></tr><tr><td>12</td><td>527</td><td>527</td><td>527</td><td>527</td><td>102</td><td>93</td><td>112</td></tr></table>		1	2	3	4	4998	4999	5000	1	-254	-254	-254	-254	-34	24	5	2	264	264	264	264	-68	-49	-34	3	517	517	517	517	-34	-73	-39	4	-5	-5	-5	-5	54	15	15	5	-386	-386	-386	-386	0	49	20	6	390	390	390	390	-54	-63	-39	7	-98	-98	-98	-98	-24	-15	5	8	-312	-312	-312	-312	595	590	605	9	-98	-98	-98	-98	137	132	142	10	810	810	810	810	-205	-200	-171	11	810	810	810	810	-200	-195	-166	12	527	527	527	527	102	93	112	12 导联心电图样本心电图数据文件。 mat 文件读取方式可以参考： https://docs.scipy.org/doc/scipy/reference/generated/scipy.io.loadmat.html
	1	2	3	4	4998	4999	5000																																																																																																			
1	-254	-254	-254	-254	-34	24	5																																																																																																			
2	264	264	264	264	-68	-49	-34																																																																																																			
3	517	517	517	517	-34	-73	-39																																																																																																			
4	-5	-5	-5	-5	54	15	15																																																																																																			
5	-386	-386	-386	-386	0	49	20																																																																																																			
6	390	390	390	390	-54	-63	-39																																																																																																			
7	-98	-98	-98	-98	-24	-15	5																																																																																																			
8	-312	-312	-312	-312	595	590	605																																																																																																			
9	-98	-98	-98	-98	137	132	142																																																																																																			
10	810	810	810	810	-205	-200	-171																																																																																																			
11	810	810	810	810	-200	-195	-166																																																																																																			
12	527	527	527	527	102	93	112																																																																																																			

心率失常种类	DX Code
Atrial Fibrillation	164889003
Atrial Flutter	164890007
Atrial Tachycardia	713422000
Sinus Bradycardia	426177001

Sinus Rhythm	426783006
Sinus Tachycardia	427084000
Supraventricular Tachycardia	426761007