

组会讲稿0906

3

以一个两类预测问题（二元分类）为例，其中结果被标记为正（p）或负（n）。

4

ROC 空间由 FPR 和 TPR 分别定义为 x 轴和 y 轴，它描述了真阳性（收益）和假阳性（成本）之间的相对权衡。

可以调整阈值，这反过来又会改变假阳性率。

5

给定阈值参数 T ，如果 $X > T$ ，则实例被归类为“正”，否则被归类为“负”。如果实例实际上属于类“正”，则 X 遵循概率密度 $f_1(x)$ ，如果实例实际上属于类“负”，则遵循 $f_0(x)$ 。因此，真阳性率由下式给出

6

在医学研究中，疾病在全体人群中的发病率往往极低。

精确度-召回率是在类别非常不平衡的情况下预测成功的有用度量。在信息检索中，精确度是衡量返回了多少真正相关结果的指标，而召回率是衡量检索到了多少真正相关结果的指标。

精确度-召回率曲线展示了不同阈值下精确度和召回率之间的权衡。曲线下的高面积代表高召回率和高精确度，其中高精确度关联于低假阳性率，而高召回率关联于低假阴性率。两者的高分数表明分类器返回准确的结果（高精确度），同时返回了大多数所有正面结果（高召回率）。

一个具有高召回率但低精确度的系统会返回许多结果，但其预测的标签大多数与训练标签不一致。一个具有高精确度但低召回率的系统正好相反，它返回的结果很少，但其预测的大多数标签与训练标签一致。一个理想的系统在高精确度和高召回率下会返回许多结果，所有结果都被正确标记。

精确率（P）定义为真阳性数（TP）与真阳性数加上假阳性数（FP）之比

召回率（R）定义为真阳性数（TP）与真阳性数加上假阴性数（FN）之比

这些数量还与 F_1 分数有关，该分数是精确度和召回率的调和平均值。

7

在图的阶梯区域可以观察到召回率和精确度之间的关系 - 在这些阶梯的边缘，阈值的微小变化会大大降低精度，召回率只有很小的提高。

平均精度

其中 P_n 和 R_n 是第 n 个阈值处的精度和召回率。

8

人工智能驱动的心电图已成为左心室射血分数估算的替代工具。然而，由于预测的不确定性，直接使用人工智能驱动的心电图可能不可靠。

虽然低 EF 的患者可能在 ECG 上表现为心房颤动，但大多数心房颤动患者的 EF 是正常的，这导致在预测心房颤动患者的 EF 时存在较高的不确定性。

因此，该研究旨在建立一种具备置信度的**人工智能驱动心电图**，以**识别左心室功能障碍**。

9

标准的12导联心电图（ECG）信号包含12个N个数的序列（在我们的数据库中， $N = 1,250$ ），心电图信号序列 $X = [x1,1, x1,2, ..., x1,N; x2,1, x2,2, ..., x2,N; ...; x12,1, x12,2, ..., x12,N]$ 被用作输入

10

我们定义了一个“密集单元”作为一个神经组合，包括：(1) 一个批量归一化层来规范输入数据，(2) 一个修正线性单元（ReLU）层进行非线性化，(3) 一个 1×1 卷积层，具有4K过滤器以减少数据的维度，(4) 一个批量归一化层进行规范化，(5) 一个ReLU层进行非线性化，(6) 一个 3×1 卷积层，具有4K过滤器以提取特征，(7) 一个批量归一化层进行规范化，(8) 一个ReLU层进行非线性化，以及(9) 一个 1×1 卷积层，具有K过滤器以提取特征。

11

使用密集单元提取特征后，我们使用了从任何层到所有后续层的直接连接所产生的密集连通性来构建一个“密集块”。我们设计了一个模型，包含5个密集块，分别包含3、3、6、6和三个密集单元。

当特征图的大小发生变化时，密集块不能被连接起来。因此，在我们的架构中，使用了池化块来连接每个密集块以进行下采样。

每个密集块通过池化块连接起来，以整合前面块的特征。

输入数据首先通过一个批量归一化层，然后是一个卷积层，又一个批量归一化层，一个ReLU层和一个池化层。初始卷积层包括K个卷积滤波器，其核大小为 7×1 ，步长为 2×1 。接下来，数据通过一系列密集块和一个池化块，结果是一个 $16 \times 1 \times 864$ 的数组。最后一个密集块后面是一个ReLU层，一个批量归一化层和一个全局池化层。最后，创建了一个具有k输出的全连接层，用于后续使用。

12

ECG12Net包含12个ECG导联块，对应于导联序列。我们基于分层注意力网络设计了一个注意力机制来连接这些块，增加了ECG12Net的解释力。注意力块包括一个批量归一化层，接着是一个全连接层，然后是两个批量归一化层、ReLU层和全连接层的组合。

为每个ECG导联计算注意力得分，然后通过线性输出层进行标准化整合。标准化的注意力得分用于通过简单乘法对12个ECG导联输出进行加权。12个加权输出被求和并转换为softmax输出层，以提供最终的预测值。

13、14

本研究提出的DLM的输出包括**实际射血分数（EF）的点估计和估计的方差**，该方差基于具有两个隐藏单元的全连接输出层。由于方差值是一个正实数，我们在第二个隐藏单元应用了指数变换作为估计方差的输出。我们**基于正态分布的最大概率密度函数**优化了DLM，具体如下：

$$f(x|u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-u}{\sigma}\right)^2}$$

其中， u 是实际的EF， y 是DLM预测的点估计值， σ 是基于DLM预测的估计方差的平方根的**标准差（SD）**。为了与仅提供点估计的先前DLM进行比较，我们还训练了一个**基于均方根误差的传统损失函数**的具有一个隐藏单元的全连接输出层的DLM。使用MXNet软件包（版本1.3.0）实现了上述**两个DLM**。

为了应用DLM对中至轻度左心室功能障碍患者进行诊断，我们基于两个DLM使用了三种方法：方法1（M1）是由训练均方根误差的DLM预测的点估计；方法2（M2）是由训练正态分布的概率密度函数的DLM预测的点估计；方法3（M3）是由训练正态分布的概率密度函数的DLM预测的点估计和估计的SD的整合，基于正态分布的累积分布函数。累积分布函数描述了具有给定概率分布的变量 X 在小于或等于 x 的值时被发现的概率。此函数如下所示：

$$F_X(x) = p(X \leq x) = \int_{-\infty}^x f_X(u) du$$
$$f_X(x) = \frac{dF_X(x)}{dx}$$

其中 u 是点估计值， x 设置为40以计算 $EF \leq 40\%$ 的概率，从而诊断出严重功能障碍的患者。如果 f_X 在 x 处是连续的，则概率密度函数是累积函数的导数。因此，M1和M2提供的值范围从10到90，类似于实际的EF，而M3提供的概率范围从0到1，用于描述低射血分数的可能性。上述方法的详细执行情况见补充材料。

15

基于方法1的DLM预测的散点图（M1，由均方根误差训练的DLM）和方法2（M2，通过正态分布的概率密度函数训练的DLM）如图S2所示。关于M1预测的点估计，平均差（SD）、Pearson相关系数和MAE为1.49（10.72）/1.57对于M2预测的点估计值，平均差（SD）、Pearson相关系数和MAE分别为1.65和1.65（9.81）/1.54（9.75）、0.61/0.58和7.56/7.51，两种预测结果与实际EF值相似，因此，正态分布损失概率密度函数训练的DLM在实际EF预测中是可行的。

16

我们使用ROC曲线评估了DLM基于M1-M3检测LVD的性能，如图S3所示。

对于检测重度LVD，基于M1、M2和M3的DLM的截止点分别为47.6、49.1和0.212，AUC值、灵敏度和特异性在内部/外部验证集中如下：AUC值分别为0.9520/0.9395、0.9578/0.9409和0.9549/0.9364；敏感性分别为85.4%/79.3%、85.4%/77.3%和84.5%/76.2%；特异性分别为92.5%/92.4%、92.6%/93.0%和92.5%/92.9%，结果表明：两种方法的预测结果相似，但在临床实践中的可解释性有显著差异，医生可能不接受将47.6和49.1作为诊断实际 $EF \leq 40\%$ 的患者的临界点。

总之，我们认为使用正态分布的累积分布函数计算概率（M3）是临床实践中最可行的方法，并在此基础上进行了进一步分析。

17

在临床实践中，对急性胸痛患者进行急性冠状动脉综合征（ACS）的心电图（ECG）诊断是一项长期挑战，临床专家对心电图图像的目测检查并不理想，导致心电图判读存在很大差异。

心脏生物标记物，包括常规或高敏肌钙蛋白（hs-cTn），在达到峰值水平之前无法区分 OMI，而这对挽救心肌来说为时已晚。肌钙蛋白阳性结果（大于第 99 百分位数）的假阳性率很高。约 25\% 的急性心肌梗死病例初始 hs-cTn 为阴性。

18

每个分类器都通过10倍交叉验证来优化超参数。在选择了最佳超参数后，模型在整个训练子集上重新训练，以推导最终的权重，并创建一个锁定模型以在保留测试集上进行评估。我们校准了我们的分类器以产生可以解释为置信水平（概率风险分数）的概率输出。

19

随机森林分类器在训练集上取得了高准确率（低偏差），在测试集上的性能下降相对较小（低方差），表明了可接受的偏差-方差权衡和低过拟合风险（扩展数据图8）。虽然支持向量机模型在测试集上的方差较小，但与随机森林模型相比，在AUROC（Delong's检验）或二元分类（McNemar's检验）方面没有显著差异。此外，在Kolmogorov-Smirnov拟合优度（0.716与0.715）或基尼纯度指数（0.82与0.85）方面，随机森林模型和支持向量机模型之间也没有差异。由于其可扩展性和直观的架构，我们选择了随机森林模型的概率输出来构建我们的衍生OMI评分。

20

指南建议使用固定的心脏肌钙蛋白阈值来诊断心肌梗死，仅根据肌钙蛋白阈值将患者分为低、中、高风险心肌梗死，但肌钙蛋白浓度会受到年龄、性别、合并症以及症状发作时间的影响。

在排除心肌梗死方面表现良好，但识别出患有心肌梗死的患者更具挑战性。

21

梯度提升通过采用集成技术来迭代地提高回归和分类问题的模型准确性。该集成算法通过创建决策树作为学习者的序列模型实现，后续模型试图纠正前一个模型的错误。在提升方法中，被前一个模型错误分类的个体会被赋予更高的权重，以增加他们在后续模型中被选中的机会。每个模型随后以逐步的方式进行拟合，以最小化损失函数，如绝对误差或平方误差（即预测值与真实值之间的差异）。XGBoost 是通过重新设计梯度提升来显著提高算法速度，推动计算资源有限的模型。