

笔记

(1)12 Lead ECG:

12导联心电图（ECG）是一种医学测试，通过附着在身体上的导线或节点记录。心电图，有时称为ECG，捕捉心脏的电活动并将其转移到绘图纸上。然后，医疗专业人员，如心脏病专家、心脏护士和技术员可以分析结果。

你可能会在医疗专业人员的办公室进行身体检查时，或在医院或诊所进行一系列测试时接受ECG。

进行心电图（ECG）的目的包括：

1. 检查心脏的电活动。
2. 找出无法解释的胸痛的原因，这可能是由心脏攻击、围心膜炎或心绞痛引起的。
3. 找出心脏病症状的原因，如呼吸短促、头晕、晕厥或快速、不规则的心跳（心悸）。
查明心脏室壁是否过厚（肥大）。
4. 检查药物的工作情况以及它们是否引起影响心脏的副作用。
5. 检查植入心脏的机械装置，如起搏器，是否有效控制正常心跳。
6. 当存在其他疾病或条件时，检查心脏的健康状况，如高血压、高胆固醇、吸烟、糖尿病或早期心脏病的家族史。

(2)ROC/AUC

[ROC曲线 - 维基百科，自由的百科全书 \(wikipedia.org\)](https://en.wikipedia.org/wiki/ROC_curve)

PRROC

[Precision-Recall — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)

这是一个评估分类器输出质量的精确度-召回率指标示例说明：

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。

精确度（P）被定义为真正例（TP）的数量除以真正例和假正例（FP）的总数。

$$P = \frac{TP}{TP + FP}$$

召回率（R）被定义为真正例（TP）的数量除以真正例和假阴例（FN）的总数。

$$R = \frac{TP}{TP + FN}$$

这些量还与F1分数相关，F1分数是精确度和召回率的调和平均值。因此，我们可以使用以下公式计算F1：

$$F1 = \frac{2TP}{2TP + FP + FN}$$

请注意，精确度可能不会随着召回率降低。精确度的定义 ($\frac{TP}{TP+FP}$) 表明，通过降低分类器的阈值可能增加返回结果的数量，如果之前的阈值设定过高，新的结果可能全部是真正例，这将提高精确度。如果之前的阈值刚好或太低，进一步降低阈值将引入假正例，降低精确度。

召回率定义为 $\frac{TP}{TP+FN}$ ，其中 $TP + FN$ 不依赖于分类器阈值。这意味着降低分类器的阈值可能会增加真正例的数量，从而提高召回率。也有可能降低阈值同时召回率保持不变，而精确度波动。

在图表的阶梯区域中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，小小的阈值变化可以显著降低精确度，同时只有微小的召回率增益。

平均精确度 (AP) 总结了这样的图表，作为在每个阈值处实现的精确度的加权平均，以从上一个阈值处的召回率增加作为权重：

$$AP = \sum (R_n - R_{n-1}) P_n$$

其中 P_n 和 R_n 是第 n 个阈值处的精确度和召回率。一对 (R_x, P_x) 被称为一个操作点。

AP 和操作点下的梯形区域是总结精确度-召回率曲线导致不同结果的常见方式。阅读用户指南了解更多信息。

精确度-召回率曲线通常用于二元分类来研究分类器的输出。为了将精确度-召回率曲线和平均精确度扩展到多类别或多标签分类，需要对输出进行二值化。可以为每个标签绘制一条曲线，但也可以通过将标签指示矩阵的每个元素视为一个二进制预测来绘制一个精确度-召回率曲线（微平均）。

(3)kappa

[Cohen's kappa - Wikipedia](#)

Definition [\[edit\]](#)

Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories. The definition of κ is

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

where p_o is the relative observed agreement among raters, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance (as given by p_e), $\kappa = 0$. It is possible for the statistic to be negative,^[6] which can occur by chance if there is no relationship between the ratings of the two raters, or it may reflect a real tendency of the raters to give differing ratings.

For k categories, N observations to categorize and n_{ki} the number of times rater i predicted category k :

$$p_e = \frac{1}{N^2} \sum_k n_{k1} n_{k2}$$

This is derived from the following construction:

$$p_e = \sum_k \widehat{p}_{k12}^{\text{ind.}} = \sum_k \widehat{p}_{k1} \widehat{p}_{k2} = \sum_k \frac{n_{k1}}{N} \frac{n_{k2}}{N} = \frac{1}{N^2} \sum_k n_{k1} n_{k2}$$

Where \widehat{p}_{k12} is the estimated probability that both rater 1 and rater 2 will classify the same item as k , while \widehat{p}_{k1} is the estimated probability that rater 1 will classify an item as k (and similarly for rater 2). The relation $\widehat{p}_k = \sum_k \widehat{p}_{k1} \widehat{p}_{k2}$ is based on using the assumption that the rating of the two raters are [independent](#). The term \widehat{p}_{k1} is estimated by using the number of items classified as k by rater 1 (n_{k1}) divided by the total items to classify (N): $\widehat{p}_{k1} = \frac{n_{k1}}{N}$ (and similarly for rater 2).

Binary classification confusion matrix [\[edit\]](#)

In the traditional 2×2 [confusion matrix](#) employed in [machine learning](#) and [statistics](#) to evaluate [binary classifications](#), the Cohen's Kappa formula can be written as:^[7]

$$\kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

where TP are the true positives, FP are the false positives, TN are the true negatives, and FN are the false negatives. In this case, Cohen's Kappa is equivalent to the *Heidke skill score* known in [Meteorology](#).^[8] The measure was first introduced by Myrick Haskell Doolittle in 1888.^[9]

网络架构

我们的DLM架构基于之前用于血清K⁺浓度估计的ECG12Net【14】。假设标准的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 $X = [x_{1,1}, x_{1,2}, \dots, x_{1,N}; x_{2,1}, x_{2,2}, \dots, x_{2,N}; \dots; x_{12,1}, x_{12,2}, \dots, x_{12,N}]$ 被用作输入，输出是AMI类别的一位有效编码（STEMI, NSTEMI, 和非AMI）以及STEMI的IRA（STEMI-LMCA, STEMI-LAD, STEMI-LCx, 和STEMI-RCA）。

例如，STEMI的标签被编码为[1,0,0]，NSTEMI的标签被编码为[0,1,0]。每个输出标签对应一个输入的片段。因为心电图信息主要由具有位移不变性的形态变化提供，因此使用了共享权重的卷积层来适应这种情况并减少过拟合的风险。因此，我们开发了一个12通道序列到序列模型，作为DenseNet的一个修订，来执行这项任务。DLM的完整架构显示在补充图1中。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。K是一个模型常数，在我们所有的实验中被设置为32。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。这个块包括一个具有2×1步长的密集单元和一个具有2×1核大小和步长的平均池化层，用于下采样。每个密集块通过池化块连接起来，以整合前面块的特征。

在我们的实验中，使用了长度为864的数值序列作为输入。我们设计了一个带有80个可训练层的心电图（ECG）导联块，其架构如补充图1A所示。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16×1×864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，其中k是类别的数量，在第一个AMI检测模型中等于3，在第二个STEMI的IRA分析模型中等于4。这个ECG导联块用于从每个ECG导联中提取864个特征，基于每个导联进行基本的输出预测。补充图1B展示了ECG12Net如何整合ECG的所有信息来进行总体预测。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。第一和第二个全连接层各包含8/k个神经元。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。上述使用ECG信息的模型被命名为ECG12Net，包含82个可训练层。使用m-log-loss函数计算模型损失。仅在全连接层中添加了一个dropout层，dropout率设置为0.5。

训练细节

12导联心电图（ECG）信号序列首先被12个ECG导联分别训练。由于STEMI、NSTEMI和非AMI分布严重不均匀，采用了过采样过程以提高性能，确保罕见样本得到充分识别。我们在每个批次中采样12个STEMI ECG、12个NSTEMI ECG和12个非AMI ECG。这个过程充分考虑了罕见的STEMI和NSTEMI案例，以避免被正常案例的压倒性数量偏移。我们使用软件包MXNet版本1.3.0来实现ECG12Net。训练模型的设置如下：(1) 使用标准参数的Adam优化器（ $\beta_1 = 0.9$ 和 $\beta_2 = 0.999$ ）和36的批大小进行优化；(2) 初始学习率设置为0.001，并在验证损失在一个时代后达到平台期时降低10倍三次；以及(3) 10^{-4} 的权重衰减。因为我们机器的采样率是500 Hz，我们的12导联心电图信号包含12个数字序列，每个序列有5000个数字。然而，ECG12Net的标准输入格式是长度为1,024的数值序列。在训练过程中，我们随机裁剪长度为1,024的序列作为输入。在推理阶段，基于间隔采样的九个重叠长度1,024的序列（X₁至X₁₀₂₄，X₄₉₈至X₁₅₂₁，X₉₉₅至X₂₀₁₈，X₁₄₉₂至X₂₅₁₅，X₁₉₈₉至X₃₀₁₂，X₂₄₈₆至X₃₅₀₉，X₂₉₈₃至X₄₀₀₆，X₃₄₈₀

至X4503，和X3977至X5000）被用来生成预测，并作为最终预测进行平均。第100个时代的模型被用作最终模型，模型在验证集中的性能只验证一次。

数据增广

前一项研究报告了在心房颤动检测任务中出现的严重过拟合问题，并建议一系列数据增强方法来改善模型性能。在当前研究中，过拟合问题是由于深度学习架构中的大量参数（约300万可训练参数）相对于样本大小而言。解决这个问题的一步是通过调整心率来调整序列长度。我们随机在从0.8 HR到1.2 HR的均匀分布中重新采样了更广泛的心率范围，其中HR是每个样本的原始心率。第二步是随机裁剪长度为1024的序列作为输入。第三步是添加一个从均值为0且标准偏差为0.1的高斯分布中抽取的随机变量。第四，均匀且随机地选择时间点，并将这些点附近50毫秒内的ECG信号值设置为0。这种方法被称为dropout burst。最后，在联合训练步骤中将六个随机ECG导联序列设置为0。我们观察到，最终的DLM只使用了几个ECG导联的信息来进行预测，并推断出模型已经停止从其他ECG导联学习特征，因为它已经完美地预测了训练集中的所有数据。这种方法迫使DLM学习所有异常的ECG导联。