

20241021 组会大纲

3

目的：利用12导联心电图推测在马萨诸塞总医院（MGH）接受长期初级保健的患者5年内新发房颤的风险。

方法：我们训练了一个卷积神经网络（ECG-AI）。

创新点

1. 基于人工智能的12导联心电图分析与已建立的房颤（AF）临床风险因素模型具有相似的预测有用性，两者是互补的。
2. AF的ECG人工智能模型在独立研究样本中具有预测有用性，可区分心力衰竭和中风患者的风险，并适用于单导联ECG描记。

4

ECG-AI的输入是单个12导联ECG（5000 x 12 维的统一输入张量）

ECG-AI 采用的不是二元分类法^{10,11}，而是编码和损失函数²¹，既考虑了结果（即房颤）发生的时间，也考虑了普查（定义为最早的死亡或随访丢失）带来的遗漏。为此，ECG-AI 编码将时间划分为离散的时间段，在这些时间段内，房颤事件或剔除事件均有可能发生，而损失函数优化了每个时间段内预测房颤发生的负对数可能性。

5, 6

编码将总随访时间划分为n个时间段，在我们的案例中，n=25，每个时间段跨度约为72天。每个个体由两个长度为n的二进制向量表示，其中一个表示删除日期的二进制掩码，另一个表示AF事件状态的one-hot编码。删失向量 V_{censor} 在个体被删失的时间段中为0，在他们处于风险集中的时间段中为1。事件向量 V_{AF} 如果在该时间段内发生AF诊断则为1，否则为0。ECG-AI输出长度为n的向量 V_{predict} ，代表在每个时间段内AF存活概率。我们的损失函数最小化了ECG-AI预测的负对数似然。似然函数分为时间段内幸存个体和发生事件的时间段中的贡献。具体来说，我们最小化以下公式：

$$L = - \sum \log(L_{\text{survival}} + L_{\text{event}})$$

其中，

$$L_{\text{survival}} = 1 - (V_{\text{AF}} * V_{\text{predict}})$$

$$L_{\text{event}} = (V_{\text{censor}} * V_{\text{predict}}) + (1 - V_{\text{censor}})$$

通过这种方式，删失个体在删失后的时间段中不会对损失做出贡献（例如，死亡或随访最后时间）。第一个时间段被保留用于记录随访开始前的事件。该模型为每个时间段预测独立的生存概率（通过sigmoid激活函数），因此不假设比例风险，并有助于处理不连续的生存曲线建模。

7

简而言之，该模型将 10 秒钟的连续 12 导联心电图波形数据作为第一卷积层的输入。全连接层只接收卷积的心电图波形数据，以生成心房颤动发生时间的估计值（主要）以及年龄、性别和诊断声明中是否存在心房颤动的预测值（次要）。箭头表示各层之间的连接。Conv1D = 一维卷积，MaxPooling1D = 一维最大池化。

目标：本研究旨在评估一种深度学习模型，以预测PVC患者中的心肌病。

方法：选择了最大的可用预训练ResNet模型（ResNet-152）

心电图数据由XML文件组成，其中包含I、II和V1-V6导联的波形数据。其余导联（III、aVF、aVL和aVR）被称为“衍生导联”，因为它们只包含其他导联的信息。应用巴特沃斯带通滤波器对心电图波形进行降噪处理，然后再进行中值滤波。结果波形数据被绘制成图像，以便使用二维卷积神经网络。

9, 10

我们选择了最大的可用预训练ResNet模型（ResNet-152）作为分析的起点。使用在自然图像上预训练的模型可以以更少的数据获得更好的性能，同时还需要更少的时间来实现最佳解决方案。

数据基于组洗牌分割进行分割，通过确保训练组和测试组中没有患者来消除数据泄漏的可能性。我们选择使用Adam优化器，学习率为 $3e-4$ ，OneCycle学习率计划。模型训练了35个epoch，并在内部测试数据上报告了性能最佳的epoch。所得模型也分别在外部验证数据上进行了评估。我们使用学习率调度以及持续记录损失和性能指标，以实现模型的最佳拟合。当模型性能开始恶化时，停止训练（发生过拟合）内部队列是指西奈山医院的患者，而外部队列由外部队列组成。验证数据集从其他4家附属医院的合并患者中获得。

我们使用梯度加权类激活映射（GradCAM）方法生成类激活映射。这些映射显示了ECG的哪些区域是最负责将模型推向预测的区域。为了进一步评估算法预测，我们提取了模型预测随后LVEF受损的患者，其概率为80%，通过检查急性失血性贫血进行消融术的特异性。然后，由电生理学家进行手动图表审查。我们排除了没有随访的患者-我们通过采集心电图、监护仪和设备询问（包括电描记图），在随访时确定是否存在室性早搏。我们使用配对双尾Student t检验评估接受PVC消融术患者的EF变化。

梯度加权类激活映射（Grad-CAM）是一种用于深度学习模型可解释性的方法，特别适用于卷积神经网络（CNN）。它通过生成热图来帮助理解模型在做出预测时关注的图像区域。

Grad-CAM的工作原理如下：

1. **梯度计算：**对于给定的输入图像和目标类别，计算模型最后一层卷积层的梯度。这些梯度表示了该层特征图对目标类别预测的影响。
2. **加权特征图：**将梯度进行全局平均池化，以获得每个特征图的权重。这些权重表明了特征图在特定类别预测中的重要性。
3. **生成热图：**将权重与相应的特征图相乘并进行加和，得到一个热图。这个热图通常会经过ReLU激活，以突出显示对目标类别有正贡献的区域。
4. **可视化：**将生成的热图叠加在原始图像上，可以直观地观察模型关注的区域，帮助理解模型的决策过程。

Grad-CAM广泛应用于图像分类、物体检测等任务，能够提高深度学习模型的透明度和可解释性。在心电图分析等领域，它同样有助于识别模型关注的ECG特征。

11

本研究的主要贡献如下：

1. 首先，研究强调了信号分割在训练和测试数据集中的重要性。本研究将每位患者的信号分组，以确保它们只分配到一个分区（即训练、验证或测试），从而使分类器在训练阶段不会部分检查患者数据，提高模型的鲁棒性。

2. 在传统神经网络模型中，采用了多种创新技术。尽管已知特征向量大小与分类器获得的准确率之间存在权衡，本研究构建了一个从心电图信号中提取较少特征的高准确率模型。为此，采用了主成分分析（PCA）和离散小波变换（DWT）的不同组合，提取心电图信号的独特特征。使用PCA，将MI检测中的信号数量从12个减少到7个，而在MI定位阶段减少到8个。结果表明，分类器在检测和定位MI时的准确率分别超过99%和98%。
3. 最后，通过利用端到端深度神经网络，我们能够构建高效的模型，只需少量步骤就能从原始信号中检测和定位MI。这一优势使得这些分类方法可以在大数据集上应用，并简化所需的硬件设计。我们提出的端到端MI检测和定位分类器在MI患者中取得了完美的结果。

12

表1展示了按研究特征、信号转换、特征提取和分类方法分类的最新MI检测和定位研究综述。

13

为了检测和定位心肌梗死（MI），实现并比较了两种自动方法（即经典的多导联心电图处理系统和端到端深度神经网络）。每种方法的第一步是信号预处理，如图1所示。接下来的步骤根据技术类型的不同而有所区别。在经典方法中，需要执行多个变换和特征提取层；而在端到端深度学习方法中，预处理后的信号直接输入深度网络，而无需额外处理。以下部分将进一步探讨这些方法的细节。

14

在预处理阶段之后，进行一系列离散小波变换（DWT）和主成分分析（PCA）变换，以提取信号的独特特征。

在对心电图（ECG）信号去噪后，信号被分解为小波系数。小波变换（WT）是信号与特定小波函数的数学卷积，用来表示信号的时频特性。

DWT（离散小波变换，Discrete Wavelet Transform）是一种信号处理技术，用来将信号分解为不同的频率分量，并分析每个分量在不同时间尺度上的特征。与傅里叶变换不同，DWT能够同时提供时间和频率信息，因此特别适合处理非平稳信号，如心电图（ECG）或语音信号。

在信号分解后，PCA被用于提取ECG信号的时间偏差。PCA是一种降维方法，用于以较少的线性独立特征表示数据的基本变化。

15

图5展示了我们使用扩张卷积的端到端深度残差神经网络的结构

16（放到文章2那块去）

模型输入应该是矩阵还是图像？

针对涉及图像的任务的最流行的深度学习模型使用2维（2D）或3D卷积神经网络来执行分类或分割等任务。ECG被认为是2D图像，因为存在针对时间（x轴）绘制的幅度轴（y轴）。因此，训练用于ECG分析的2D卷积神经网络模型具有直观意义。然而，对于时间序列数据，我们可以通过将ECG数据作为矩阵输入并训练1D卷积神经网络来节省计算资源，而性能没有任何下降，因为存在于2D图像中的所有信息已经存在于1D信号中。对于ECG，一种方法是将每个样本作为输入（number of samples = sampling frequency × time）。⁹这种技术允许我们在其他地方分配计算资源，并优选地训练更大的数字以构建更鲁棒的模型。Lampert等人¹⁰确实从ECG波形数据开始，但后来将其绘制为图像。尽管这种转换有助于人类理解，这是一个无关的步骤，无意中将人工智能限制在类似人类的处理上。将人工智能从这种人类定向约束可能会导致更优化的结果。该转换以图像分辨率的形式引入了不必要的限制，并在ECG描记图周围生成多余的多余白色像素。因此，可以质疑这是否值得额外的计算量。

ECG波形数据以250 Hz采集，并提取为10秒，12 x 2500振幅值矩阵，存储为base64文本。ECG进行基线漂移校正，使用200 ms和600 ms间期的中值滤波和z评分标准化。

我们采用了一种基于先前用于从ECG预测临床表型的新架构的atrous卷积神经网络

局限性

有几个限制需要考虑。由于这是一项回顾性研究，12导联ECG的人群可能与前瞻性AF筛查人群不同。虽然我们的VA系统中的ECG是在临床访视期间常规获得的，但每个患者的平均ECG数量存在研究中心间的差异，我们可能预计本研究中接受心电图检查的患者人群心血管疾病和房颤的患病率更高。这种选择偏差可能会增加模型的阳性预测值与筛查更广泛的患者人群时使用该模型相比，减少了筛查所需的数量。尽管如此，如果选择更高风险的人群进行前瞻性筛查，前瞻性模型的性能可能相似。虽然我们使用了ECG数据库和电子健康记录中的所有数据来识别AF病例，在对照组中仍有可能有未确诊的房颤患者。这将使我们的结果偏向于零，并导致低估我们模型的性能。一些被预测为病例的患者实际上可能是未来的前瞻性研究使用我们的模型对高风险患者进行持续监测，可以证实AF预测，并澄清该方法是否改善了下游结局，如中风和血栓栓塞。

现有研究主要集中在单一标签的预测，尚未有研究开发一个预测系统，能够同时检测这些特定疾病。由于缺乏大型临床标注的医疗数据集来进行监督机器学习，这是一个公认的问题，而在人口规模上进行大规模验证，对于展示预测模型的可信度至关重要，因为这对于将模型成功应用于临床实践非常重要，在那里，早期识别和治疗可能会影响疾病并发症、医疗使用和成本。

我们采用基于ResNet的深度学习（DL）模型分析心电图波形，并采用极端梯度提升（XGB）模型分析心电图测量数据。

因此，我们使用来自单一支付者的全民健康系统的大规模人口队列，开发并验证了基于12导联心电图波形的深度学习（DL）模型，以及基于常规采集的心电图测量值的极端梯度提升（XGB）模型，以通过统一的预测框架同时预测15种常见的心血管疾病。

预测模型的目标是输出校准后的概率，针对每种疾病提供预测。这些学习模型可使用在医疗事件中的任何时间点获取的心电图。在模型训练过程中，我们使用了所有心电图数据（tip：感觉这样做不太合适）（包括同一医疗事件中获取的多份心电图），以最大化学习效果。然而，在模型评估时，我们仅使用给定医疗事件中的首次心电图，以模拟在急诊或住院期间首次获取的心电图的预测场景（详见“评估”部分）。

我们使用了基于ResNet的深度学习（DL）模型处理包含丰富信息的电压-时间序列数据，并使用了基于梯度提升（XGB）的模型处理心电图测量值【25】。为了确定人口统计特征（如年龄和性别）是否对仅基于心电图的模型性能有额外的预测价值，我们开发并报告了以下三类模型：(a) 仅基于心电图（DL：ECG trace）；(b) 基于心电图、年龄和性别（DL：ECG trace, age, sex，这是本文的主要模型）；以及(c) XGB模型：基于心电图测量值、年龄和性别。

学习算法

我们采用了**多标签分类方法**，对每种疾病的存在与否进行二分类预测，以估计新患者患有这些疾病的概率。

由于使用心电图测量值的模型输入的是结构化表格数据，我们为每个疾病标签训练了独立的梯度提升树集成模型（XGB）【38】，而对于处理心电图电压-时间序列数据的模型，我们使用了深度卷积神经网络模型。对于XGB和DL模型，我们使用90%的数据进行训练，其余10%作为调优集，以跟踪性能损失并在合适的时间点“早停”，从而减少过拟合的可能性【39】。对于深度学习，我们使用了**单一的ResNet模型，进行多分类多标签任务**，将每份心电图信号映射为15个值，对应于每种疾病的概率。而对于XGB模型，我们为每个标签独立训练了15个二分类模型，分别预测每个疾病的概率。

23

目的：我们旨在基于**12导联心电图⁽¹⁾**开发一个深度学习模型（DLM），作为一种诊断辅助工具。

整合DLM可能协助一线医生及时准确地识别AMI，以防止AMI的延误诊断或误诊，从而提供及时的再灌注疗法。

评估了DLM和传统心肌肌钙蛋白I（cTnI）对STEMI和NSTEMI的诊断能力。

24

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。

25

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。**

26

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。**

27

该模型的训练**目的是识别在首次住院期间被判定诊断为1型、4b型或4c型心肌梗死的患者。**

我们首先使用四种统计方法——**逻辑回归、朴素贝叶斯、随机森林和极端梯度提升（XGBoost）**——开发并评估了模型。

XGBoost 模型的输出是一个概率分数，使用训练模型的终端节点权重的和进行反向逻辑变换得出。

梯度提升模型的数学公式可以描述为：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F,$$

其中， f 是一个将每个变量向量 $x_i (x_i = \{x_1, x_2, \dots, x_n\}, i = 1, 2, N)$ 映射到结果 y_i 的函数， K 是分类和回归树的数量 ($k = 1, 2, N$)， F 是包含所有分类和回归树的函数空间。

XGBoost 优化一个目标函数，该函数形式为：

$$Obj = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

其中，第一项是损失函数 l ，它评估模型拟合数据的效果，通过测量预测值 y_i 与结果 y_i 之间的差异。第二项是正则化项，是 XGBoost 用来避免过拟合的方法，通过惩罚模型的复杂性来实现。此外，为了最大程度发挥 XGBoost 的优势，我们通过 10 折交叉验证使用网格搜索策略对算法的超参数进行了调优（参见补充表 11）。

29

简而言之，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。

在诊断时有或没有心肌损伤的患者中，使用 XGBoost 模型表现最佳，无论是使用首次心肌钙蛋白测量值还是连续测量值（补充表 2）。这些 XGBoost 模型被整合到一个名为 CoDE-ACS 的临床决策支持系统中，该系统计算一个得分（0-100），代表个体患者发生心肌梗死的概率（<https://decision-support.shinyapps.io/code-acsc/>）。

30

确定了驱动模型分类的最重要的心电图特征

推导出的 OMI 风险评分提高了诊断和排除的准确性，有助于对三分之一的胸痛患者进行正确的再分类。

拟合了 10 种机器学习分类器：正则化逻辑回归、线性判别分析、支持向量机（SVM）、高斯朴素贝叶斯、随机森林、梯度提升机、极端梯度提升、随机梯度下降逻辑回归、k 最近邻和人工神经网络。

随机森林分类器在训练集上取得了高准确率

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。

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

31(DWT可能很合适)

然而，这些方法大多只关注时域信息，忽略了来自其他模态或视角的信息，考虑到多模态数据之间的互补性[15]，仅依赖于单一模态的方法无法捕捉到更全面的信息，从而限制了模型的推理和判断能力。

我们提出了一种新的双峰掩蔽自编码器框架，表示为 BMIRC

主要贡献

- 我们提出了一种新颖的用于时频联合建模的双模掩蔽自编码器框架。该方法将心电图的频谱集成到掩蔽预训练过程中，使双模联合编码器能够学习全面且通用的表示

- 我们在编码器和解码器之间建立内部表示连接（IRC），并设计了一个门控表示混合器（GRM）来复用不同层次的信息，从而减轻了解码器的重构负担，同时促进编码器获得更具鉴别力的表示。

32

图2 预训练阶段的整体框架：

首先，原始ECG信号通过离散傅里叶变换（DFT）转换为频谱。随后，时间和频率模态的patch嵌入（称为tokens）通过1D卷积核提取。这些tokens根据预定比例进行掩码处理，部分tokens被丢弃。未掩码的tokens分别输入时间和频率编码器。共享编码器用于融合来自两种模态的表示，而时间和频率解码器依赖于共享编码器的输出进行重构。同时，来自时间、频率和共享编码器的某些中间层表示被送入解码器，从而建立内部表示连接（IRC）。

33

在我们的方法中，ECG和频谱被分成不重叠的patch以进行编码。一个多导联ECG及其对应的频谱表示为 $T = [t_1, t_2, \dots, t_C] \in \mathbb{R}^{L \times C}$ 和 $F = [f_1^*, f_2^*, \dots, f_C^*] \in \mathbb{R}^{\frac{L}{2} \times C}$ ，其中 C 是导联的数量。我们使用两个一维卷积层对来自两种模态的patch进行编码。卷积核的大小设为 $S \times C$ ，步长设为 S ，以确保patch的独立性。在这种配置下，每个patch的长度为 S ，表示心电图的一个片段。按照MAE的思想，patch嵌入被表示为token，每个token对应于特定patch的嵌入。T和F的token表示为：

$$Z_t = [z_t^1, z_t^2, \dots, z_t^{\frac{L}{S}}] \in \mathbb{R}^{\frac{L}{S} \times D}, Z_f = [z_f^1, z_f^2, \dots, z_f^{\frac{L}{2S}}] \in \mathbb{R}^{\frac{L}{2S} \times D}$$

其中， D 表示卷积核的数量，表示每个token的维度。"t"和"f"分别表示ECG和频谱模态。为了便于表述，ECG和频谱也分别被称为时间和频率模态。在接下来的部分中，使用了 $N_t = \frac{L}{S}$ 和 $N_f = \frac{L}{2S}$ 来表示时间和频率模态中的token数量。

在我们的方法中，Transformer 作为编码器的主要组件。鉴于自注意力机制对输入位置的天然不敏感性，可学习的位置嵌入 $PE \in \mathbb{R}^{N \times D}$ 被集成到patch嵌入中，以增强模型的token定位能力。此外，还为每个模态引入了一个额外的可学习全局token $z_g \in \mathbb{R}^D$ ，其中 "g" 表示 "全局"，从而促进全局信息的提取。最后，对于模态 $m \in \{t, f\}$ ，输入tokens $I_m \in \mathbb{R}^{N_m \times D}$ 表示为：

$$\begin{aligned}\tilde{I}_m &= Z_m + PE_m \\ I_m &= \text{Concat}(z_g^m, \tilde{I}_m)\end{aligned}$$

其中，Concat是连接运算符，"t"和"f"分别代表时间模态和频率模态。

在预训练阶段，输入到编码器的tokens需要以预定速率进行遮掩。受MAE启发，我们采用了一种随机遮掩策略，意味着每个token有相同的概率被遮掩。由于每次遮掩操作是随机的，在各个批次和训练轮次之间，重构任务的表现具有多样性，从而使预训练任务更加具有挑战性。高遮掩率，例如60%和75%，表示重构难度较高，并已被证明可以提升预训练模型的性能【7,11】。然而，由于时间模态和频率模态的数据特性不同，在联合建模中采用统一的遮掩率并不理想。通过实验，我们确定遮掩率为50%（用于ECG）和75%（用于频率谱）能帮助模型实现最佳性能。考虑到ECG相比频率谱的复杂性更高，采用相对较低的遮掩率有助于提升模型性能，这凸显了在建模能力和重构难度之间权衡的必要性。

在输入共享编码器之前，各个模态的tokens经过初步融合，经过层归一化（LN）处理。具体来说，时间和频率模态的**全局tokens相加，并插入到序列的第一个位置**，其他tokens依次串联。然后，使用共享编码器促进双模态表示的深度融合。形式上，模态特定编码器最终层的输出表示为

$O_m = [o_g^m, o_1^m, o_2^m, \dots, o_n^m]$ ，我们将其传递到共享编码器 Θ 中，按照以下方程式处理：

$$\begin{aligned}\tilde{O}_m &= LN(O_m) = [\tilde{o}_g^m, \tilde{o}_1^m, \tilde{o}_2^m, \dots, \tilde{o}_n^m] \\ O_0^s &= [o_g^t + \tilde{o}_g^f, \tilde{o}_1^t, \tilde{o}_2^t, \dots, \tilde{o}_n^t, \tilde{o}_1^f, \tilde{o}_2^f, \dots, \tilde{o}_n^f] \\ O_s &= \Theta(O_0)\end{aligned}$$

其中 O_s 是共享编码器的输出表示。

35

在解码器的深度为 H 的情况下，我们从编码器中均匀地选择 $H - 1$ 层表示 $[V_h^m, V_{h+1}^m, \dots, V_H^m]$ 进行融合，其中 m 代表模态（时间模态或频率模态），小下标代表来自较深层次的表示。为了融合这些编码器的表示，我们设计了一个叫做**门控表示混合器（GRM）**的机制。

正如图 4 (a) 所示，针对模态 m ，在解码器的第 h 层，来自编码器的第 h 层表示 V_h 通过门控机制与解码器的第 h 层表示 U_h^m 融合。层规范化-线性模块（由 P_h 表示）用于将 V_h 转换为 \hat{V}_h ，从而促进需要融合的表示之间的对齐。**由于解码器输出 U_h^m 包含可学习的掩码标记，我们使用零标记填充 \hat{V}_h 的对应位置，以保持维度的一致性。**

门控单元 G_h 包含连接和线性变换，用来控制输入表示到输出的贡献。由于只有两个输入，**Sigmoid 激活函数 σ** 被用来将贡献转换为相应的权重向量 $w_h \in \mathbb{R}^N$ ，这意味着每个 token 都有一个融合权重。在 token 层级， \hat{V}_h 和 U_h^m 经过加权后相加，**权重由 w_h 引导，这意味着可以对每个 token 应用自适应融合策略。**随后，融合后的表示 C_h^m 被输入到第 $h + 1$ 层的 Transformer 块 Λ_{h+1} 以获得 U_{h+1}^m 。

特别是，以上描述的融合仅涉及可见 token，旨在补充各种信息的不同层次，促进掩码 token 学习过程的进展。该过程在每个后续层中迭代进行，编码器的浅层表示逐步融合到解码器的更深层，进而逐渐推动重构过程的进展。上述过程通过以下公式表示：

$$\begin{aligned}\hat{V}_h &= P_h(V_h) \\ w_h &= \sigma(G_h(\hat{V}_h, U_h^m)) \\ C_h^m &= w_h * \hat{V}_h + (1 - w_h) * U_h^m \\ U_{h+1}^m &= \Lambda_{h+1}(C_h^m)\end{aligned}$$

36

普查（定义为最早的死亡或随访丢失）带来的遗漏

确保训练组和测试组中没有患者来**消除数据泄漏的可能性**

DWT（离散小波变换，Discrete Wavelet Transform）是一种信号处理技术，用来将信号分解为不同的频率分量，并分析每个分量在不同时间尺度上的特征。与傅里叶变换不同，DWT能够同时提供时间和频率信息，因此特别适合处理非平稳信号，如心电图（ECG）或语音信号。