arXiv:2603.02616v1 [stat.AP] 3 Mar 2026

通过广义加性模型和可解释的基础模型预测因子，从心电图中检测结构性心脏病

Ya Zhou[1, #, *]、Zhaohong Sun[1, #]、Tianxiang Hao[2, #] 和 Xiangjie Li[3, *]

[1]中国医学科学院、北京协和医学院阜外医院信息中心部，北京，*100037*，中国[2]清华大学清华大学深圳国际研究生院，深圳，*518055*，中国[3]中国医学科学院、北京协和医学院阜外医院国家临床研究中心（心血管疾病），国家心血管病中心，北京，*100037*，中国

摘要

结构性心脏病（SHD）是一种常见疾病，许多病例尚未确诊；而超声心动图（ECHO）成本高且可及性受限，常常妨碍早期检测。近期研究表明，基于人工智能（AI）的心电图（ECG）分析能够检测结构性心脏病，可提供一种可扩展的替代方案。然而，现有方法均为黑箱模型，限制了其可解释性和临床应用。为应对这些挑战，我们提出了一种兼具可解释性和有效性的框架，将具有临床意义的心电图基础模型预测因子纳入广义加性模型，在保持强大预测性能的同时，实现透明的风险归因。基于包含超过80,000 组心电图—超声心动图配对数据的 EchoNext 基准，该方法相较最新的最先进深度学习基线，在 AUROC、精确率-召回率曲线下面积和 F1 分数上分别提升 $+0.98\%$、$+1.01\%$ 和 $+1.41\%$；即使仅使用 30% 的训练数据，性能也略胜一筹。亚组分析证实，该方法在不同人群中表现稳健；估计的逐项函数则揭示了传统心电图诊断风险与结构性心脏病之间关系的可解释见解。本研究展示了经典统计建模与现代人工智能之间互补的研究范式，为实现可解释、高性能且具有临床应用价值的心电图结构性心脏病筛查提供了路径。

关键词：结构性心脏病、心电图、基础模型、广义加性模型、多项式样条

1 引言

结构性心脏病（SHD）是全球性的健康挑战，早期发现对于预防不良结局至关重要（Poterucha 等人，2025年）。SHD 包括心脏

[#]这些作者贡献相同。[*]通讯作者：Ya Zhou (zhouya@fuwai.com)、Xiangjie Li (ele717@163.com)

瓣膜、心肌和心腔异常，例如心脏瓣膜病和心力衰竭，影响全球数以千万计的人群（Mensah etal.，2023年；Tsao et al.，2023；Ulloa-Cerna et al.，2022；Wang et al.，2023）。尽管患病率很高，仍有相当一部分病例未能确诊。例如，大型社区筛查研究的估算结果表明，进行系统性超声心动图检查（ECHO）后，具有临床意义的心脏瓣膜病患病率会增加一倍以上，这说明常规诊疗中漏诊疾病的负担相当沉重（d'Arcy et al.，2016）。尽管超声心动图仍是诊断金标准，但其费用高昂且依赖专业技术，限制了其广泛应用（Diao et al.，2025；Kwon etal.，2020）。此外，许多结构性心脏病亚型的症状并不特异，或仅在疾病晚期出现，缩短了有效干预的时间窗口（Poterucha et al.，2025）。这些挑战凸显了开发可扩展、准确且易于使用的工具以识别结构性心脏病患者的必要性。

心电图检查（心电图）是一种普及度高、成本低廉的检查，可记录心脏电活动，为人群筛查提供了一种可扩展的方案。然而，常规临床解读通常不足以检测结构性心脏病，因为心电信号中的相关模式可能十分细微，肉眼难以察觉（Siontis 等人，2021年）。深度学习领域的最新进展表明，利用大规模心电图数据集及配套的超声心动图标签训练的模型，能够识别这些细微模式，并取得较高的预测性能（Attia 等人，2019年；Cohen-Shelly 等人，2021年）。这项研究证明了基于心电图的结构性心脏病检测切实可行，并推动了人们对利用心电信号进行更早期、可扩展识别的研究兴趣（Diao 等人，2025年；Poterucha 等人，2025年）。

然而，现有的大多数方法都是完全的黑箱模型（Attia 等人，2019年；Cohen- Shelly 等人，2021年；Diao 等人，2025年；Poterucha 等人，2025年；Siontis 等人，2021年；Somani 等人，2021年），这限制了模型的可解释性和临床应用。主流范式构建端到端深度学习模型，将心电信号直接映射为超声心动图得出的标签。不同于心房颤动、窦性心动过速或窦性心律不齐等常规心电图任务——在这些任务中，心电图是金标准，且信号变化清晰可见，并已在临床指南中明确描述——与结构性心脏病相关的模式可能较为细微，也不符合传统的心电图知识（Siontis 等人，2021年）。因此，尽管这些模型的预测性能出色，却仍然难以解释，并引发了对其透明度、验证和现实世界部署的重大担忧。

为解决这些局限性，我们提出一种简单而有效的框架，将可解释的心电图基础模型预测因子整合到广义加性模型中。我们不训练端到端黑箱分类器，而是构建一种模型，其中每项输入都对应于从后训练的心电图基础模型中提取的、具有临床意义的预测因子。这些预测因子表示传统心电图诊断的校准风险，在临床实践中被广泛使用，也为临床医生所熟悉。重要的是，每个预测因子还反映了嵌入底层心电图信号中的特定心律或形态模式，从而提供一种基于生理学的心脏电活动表征。鉴于与结构性心脏病相关的心电图模式可能极为复杂，我们采用非参数函数估计，灵活刻画这些模式的贡献，同时保持透明度。这一建模方式使用户能够考察单个心电图衍生模式对结构性心脏病风险的影响，并评估所推断的关系是否符合既有临床知识。总之，该方法在保持可解释性和生理相关性的同时，实现了出色的预测性能。

我们使用 EchoNext 基准数据集评估所提出的方法（Elias 和 Finer，2025年），该数据集包含超过 80,000 组心电图–超声心动图配对。我们将（Zhou 等人，2025年b）中的后训练策略应用于心电图基础模型 ST-MEM（Na 等人，2024年），并使用包含 71 个传统心电图诊断标签的 PTB-XL 数据集（Wagner 等人，2020年）构建预测器提取器。对于加性建模部分，我们依据标准实践，采用通常推荐的样条阶数和基维数的 B 样条基（Fan 等人，2014年；Huang 等人，2010年）。与最先进的 Columbia mini模型（Poterucha 等人，2025年）相比，所提出的方法在 AUROC、精确率-召回率曲线下面积和 $+0.98\%$方面相对提升了 $+1.01\%$，F1分数提升了 $+1.41\%$。值得注意的是，即使仅使用 30% 的可用数据进行训练，所提出的框架仍略胜于 Columbia mini模型，这表明其数据效率更高。亚组分析表明，模型在不同的人口统计学和临床亚群中均保持稳定的性能。此外，对估计的逐项函数（即分量非线性效应）的考察揭示了传统心电图诊断风险与结构性心脏病之间的关联，为预测机制提供了可解释的结构和潜在的临床见解。

总而言之，本研究的主要贡献如下：

1. 我们提出了一种新的可解释建模框架，将心电图基础模型预测因子与广义加性结构相结合，为现有端到端黑箱结构性心脏病检测方法提供了一种透明的替代方案。2. 在包含 80,000 多份心电图的基准上，所提出的方法取得了出色的预测性能和更高的数据效率，优于当前最先进的模型，同时保持可解释性，并在不同人口统计学和临床亚组中展现出稳健性。3. 本研究在统计建模与基础模型人工智能之间确立了一种互补范式，表明可解释性与高预测性能能够兼得。

本文其余部分安排如下：第 2 节介绍所提出的框架。第 3 节说明如何通过估计方法获得该框架中的预测器提取器和加性分量。我们在第 4 节基于一个大规模真实世界数据集开展实验。第 5 节提供进一步讨论，第 6 节则简要总结全文。

2 模型

我们研究固定长度的多导联心电图（ECG），例如典型时长为 10 秒的标准 12 导联记录。每份心电图均表示为多变量时间序列

$$\mathbf{X} \in \mathbb{R}^{L \times Q},$$

其中，L表示导联数，Q 表示采样时间点数。此外，人口统计学或临床协变量（例如年龄和性别）表示为向量 $\mathbf{z} \in \mathbb{R}^{p}$。二元响应 $y \in \{0, 1\}$ 表示是否存在结构性心脏病（结构性心脏病）。

We 通过广义加性公式对条件平均响应进行建模 n:

$$g\{\mathbb{E}(y \mid \mathbf{z}, \mathbf{X})\} = \boldsymbol{\gamma}^{\top}\mathbf{z} + m(\mathbf{X}), \tag{1}$$

其中， $g(\cdot)$ 是固定的连接函数（例如逻辑斯蒂连接函数）， $\boldsymbol{\gamma} \in \mathbb{R}^p$ 是未知回归参数，而 $m(\cdot)$ 是用于捕捉心电图形态影响的未知函数。

在实践中，响应变量 y 是根据超声心动图（ECHO）而非人工心电图解读得出的。由于与结构性心脏病相关的电异常可能十分细微，难以被人眼识别（Siontis 等人，2021年），从 $\mathbf{X}$ 到 y 的端到端映射往往高度非线性，且难以解读。因此，主要的方法学挑战在于以兼具准确性和临床可解释性的方式估计 $m(\cdot)$ 。

为说明所提出的方法，我们首先在第 2.1 节和第 2.2节回顾现有的建模策略及其局限性。第 2.3 节随后介绍一种简单而有效的建模框架，兼具预测准确性和临床可解释性。

2.1 经典统计建模：可解释，但表达能力有限

经典统计建模方法具有可解释性，但应用于心电信号时，往往缺乏足够的表达能力。这一局限通常会导致预测准确性欠佳。

一个自然的起点是假设 $m(\cdot)$ 具有线性形式，这种形式易于估计，并能提供清晰透明的风险归因。然而，对于基于心电图的结构性心脏病预测而言，线性模型的限制过于严格。为放宽这些限制，可以采用加性建模方法（Hastie 和 Tibshirani，1990年；Stone，1985年）：

$$m(\mathbf{X}) = \nu + \sum_{l,q} m_{l,q}(X_{l,q}), \tag{2}$$

其中， $\nu \in \mathbb{R}$ 是截距项， $X_{l,\ q}$ 表示 $(l,\ q)$ 中的第 $\mathbf{X}$ 个元素，而 $m_{l,\ q}$ 是未知的单变量函数。这种形式使每个预测变量都能对响应产生灵活且可能是非线性的影响。然而，对于以 250 赫兹采样、持续 10 秒的标准 12 导联心电图， $LQ = 30,\ 000$，这会导致需要估计的独立未知函数数量过于庞大。此外，单变量函数的加性结构从根本上无法对不同导联和时间点之间的非线性生理依赖关系进行建模，而理解空间和时间上协调的电生理模式则需要这种建模能力（Na 等人，2024年）。

原则上，纳入多变量交互作用的高阶模型可以解决这些局限。然而，设定此类交互作用需要大量领域知识，并且很快会导致模型组合数量激增，使这些方法难以大规模应用，尤其是在心电图数据分析中。

这些局限促使人们采用深度学习方法。这类方法能够灵活地建模传统方法无法捕捉的非线性时间结构和跨导联依赖关系。然而，尽管此类模型具有出色的预测性能，却也带来了透明度和可解释性方面的新挑战，详见下文。

2.2 纯深度学习：准确但难以解释

近期研究表明，基于原始心电图信号进行端到端训练的深度学习模型，能够在检测细微心电图异常方面取得出色的预测性能。鉴于结构性心脏病（SHD）的临床重要性，人们已广泛探索利用端到端深度学习，根据心电图数据预测复合型SHD或特定亚型。例如，Attia 等人（2019年）利用来自44,959名患者的配对12导联心电图和超声心动图测得的左心室射血分数，训练卷积神经网络（CNN）来检测心室功能障碍。Cohen-Shelly 等人（2021年）利用129,788组心电图—超声心动图配对数据，开发了一种用于识别中度至重度主动脉瓣狭窄的CNN。Poterucha 等人（2025年）利用来自大型综合医疗系统的一百多万组心电图—超声心动图配对数据训练CNN，以检测结构性心脏病。近期，Diao 等人（2025年）提出了一种基于466,149组心电图—超声心动图配对样本训练的自蒸馏预训练Transformer模型，用于检测三尖瓣反流及其亚型。

尽管这些模型实现了很高的预测准确性，但它们仍然难以解释。临床医生无法轻易理解为何一份心电图被判定为阳性或阴性，这限制了信任和临床应用。归因图或 Grad-CAM（Hata 等人，2020年）等事后可解释性工具可以突出影响模型决策的输入区域；然而，被突出显示的波形片段往往并不对应于可识别的临床特征，尤其是在潜在疾病特征较为细微时。

这些挑战促使我们开发能够保留深度学习预测优势、同时确保决策透明的建模策略。为弥补这一差距，我们提出一种建模框架，将高容量的深度心电图表征与结构化加性模型相结合，从而在不牺牲准确性的前提下实现临床可解释的潜在预测变量，如下文所述。

2.3 具有可解释潜在预测因子的混合建模框架 s

我们提出一种混合建模框架，将深度学习得到的心电图表征与加性统计建模相结合。其核心思想是将原始心电图波形转化为一组结构化、可解释的潜在预测变量，以反映具有临床意义的形态学特征和节律特征。未知函数 $m(\cdot)$ 使用这些预测因子的加性函数进行建模，表述如下：

$$m(\mathbf{X}) = \sum_{j=1}^{J} f_j[\sigma\{h_j(\mathbf{X})\}], \tag{3}$$

其中，J 表示预定义的预测因子数量，$h_j : \mathbb{R}^{L\times Q} \to \mathbb{R}$ 表示从心电图中提取的潜在预测变量，$\sigma(\cdot)$是Sigmoid函数，$f_j(\cdot)$是一个未知的平滑单变量函数，用于捕捉预测变量与结构性心脏病之间可能存在的非线性关联。与传统心电图建模框架不同，在传统框架中，$h_j(\cdot)$由专家根据规则手动设计；在这里，每个 $h_j(\cdot)$都对应深度神经网络的一个Logit输出。实际应用中，可以选取与传统心电图诊断类别对应的一组对数几率作为可解释预测变量，例如心房颤动、窦性心动过缓或窦性心律不齐。经过Sigmoid变换后，这些对数几率会得到代表临床定义的传统心电图诊断风险的数值，这些诊断

基于成熟的波形模式和临床指南。对于这类诊断任务，深度学习模型的表现已达到或超过心脏病专家的水平（Hannun等人，2019年；Jiang等人，2024年；Ribeiro等人，2020年）。重要的是，此类模型已在许多医院投入临床使用，因此由此产生的潜在预测变量对临床医生而言熟悉且易于理解。

需要指出的是，传统的心电图诊断任务与利用心电图检测结构性心脏病有很大不同。前者反映的是明显、可识别且具有临床定义的波形异常，而后者则依赖于一些细微的生理特征，这些特征可能无法通过视觉解读（Siontis 等人，2021年）。尽管如此，许多经临床验证的心电图诊断标签所对应的对数几率，为我们提供了一组丰富的潜变量，能够捕捉多种电生理特征。通过使用函数 $f_j(\cdot)$对其影响进行非参数建模，所提出的框架能够灵活地表示结构性心脏病与这些对数几率之间的非线性关联，同时保留单个预测因子层面的可解释性。第 4 节将通过大规模真实世界数据展示这些优势。

关于作为预测器提取器的网络 $h_j(\cdot)$ 的选择，已有许多心电图建模方法被提出（Behrouz 等人，2024年；Cheng 等人，2023年；Li 等人，2025年；Na 等人，2024年；Shi 等人，2023年；Strodthoff 等人，2021年；Zhang 等人，2023年；Zhou 等人，2025年a）。原则上，任何能够准确、全面地预测心电图诊断结果的模型都可承担这一角色。近年来的心电图基础模型尤其适合，因为它们具有良好的可扩展性、广泛的适应能力，并且可通过广泛可用的开源实现轻松使用。在本研究中，我们采用此类模型，并应用后训练策略来构建所提出的框架。

3 方法

第 2 节介绍了所提出的混合建模框架。本节将详细介绍一种具体的估计策略，该策略将心电图基础模型预测变量提取与非参数统计建模相结合。实施该框架需要明确三个组成部分：

(1) 选择合适的心电图基础模型；

(2) 选择可解释预测变量，以及用于学习 $h_j(\cdot)$、 $j = 1$、...、 J 的相应后训练策略；

(3) 未知函数 $f_j(\cdot)$、 $j = 1$、...、 J 的估计方法。

我们将在下面的各个小节中介绍每个组成部分。

3.1 心电图基础模型

如第2.3节所述，任何能够准确、广泛地预测心电图诊断结果的模型，都可以用作特征提取器 $h_j(\cdot)$。在本研究中，我们采用了一种可扩展、适应性广且易于使用的心电图基础模型，其开源实现可公开获取。

近年来，已有多种心电图基础模型问世，包括 ECG-FM（McKeen 等人，2024年）、HuBERT-ECG（Coppola 等人，2024年）、ECGFounder（Li 等人，2025年）和 ST-MEM（Na 等人，2024年），涵盖基于卷积神经网络和基于Transformer的架构。我们使用 ST-MEM，这是一种在掩码自编码器框架中训练的 Transformer模型变体（He 等人，2022年）；相关变体已在大规模真实世界心电图场景中展现出稳健性（Diao 等人，2025年；Zhou 等人，2025年a）。值得注意的是，ST-MEM 的预训练权重和实现均已公开，可确保研究结果可复现且易于使用。其他基础模型也可能适用于我们的框架，未来可以进一步探索。

3.2 可解释预测变量与后训练

传统心电图诊断标签通常呈现层次结构，其中细粒度的节律或形态学标注可归入更宽泛的诊断类别。例如，SCP-ECG陈述分类体系（Wagner等人，2020年）将左心室肥厚、右心室肥厚及相关异常汇总为更高层级的肥厚类别。为保留诊断粒度并最大化信息含量，我们采用最低层级（即最具体）的诊断类别作为候选预测变量。由于可用的常规心电图诊断标签数量通常远小于训练数据集的规模（见第4.1.1节），我们保留全部最低层级标签，而不再进行额外的特征选择。在对相应对数几率施加Sigmoid变换后，这些输出在所提出的混合框架中充当可解释潜在预测变量。

为获得这些预测因子的可靠估计，我们采用改编自Zhou等人（2025年b）的后训练策略，该策略相较ECG基础模型的标准微调过程已展现出显著的性能提升。该策略包含两个阶段。首先，在初始化阶段，我们应用线性探测，基于公开可用的预训练表示估计最终分类层的权重。其次，在正则化阶段，我们引入随机深度和dropout，以提高稳健性并减轻过拟合。此后训练过程提升了常规心电图诊断任务上的性能，从而改善了本框架加性建模阶段所用潜在预测变量的质量。

3.3 非参数估计

对于每个观测值，潜在诊断预测因子满足 $\sigma\{h_j(\mathbf{X})\} \in (0, 1)$，因此非常适合进行平滑的非参数建模。我们使用阶数为 ζ 的 B样条基来近似每个未知函数 $f_j(\cdot)$：

$$f_j(x) \approx \sum_{k=1}^{K} \alpha_{j,k} b_{j,k}(x), \tag{4}$$

其中 $b_{j,1}(x)$ ……，$b_{j,K}(x)$ 表示样条基函数，而 $\alpha_{j,k}$则是未知系数。由于不同诊断信号之间的预测变量分布可能存在显著差异，因此允许样条基函数的设定随 j、$j = 1$、……、J而变化。将展开式代入（3）可得

样条近似形式：

$$m(\mathbf{X}) \approx \sum_{j=1}^{J} \sum_{k=1}^{K} \alpha_{j,k} b_{j,k}\big[\sigma\{h_j(\mathbf{X})\}\big]. \tag{5}$$

由于样条基满足$\sum_{k=1}^{K} b_{j,\,k}\,(x)\;=1$，因此每个函数都有一个多余的自由度 $f_j\,(\cdot)$。为确保可识别性，我们对每个预测变量移除一个基函数，并加入一个全局截距。忽略近似误差，模型（1）可表示为：

$$g\{\mathbb{E}(y \mid \mathbf{z}, \mathbf{X})\} = \nu + \boldsymbol{\gamma}^{\top}\mathbf{z} + \sum_{j=1}^{J} \sum_{k=1}^{K-1} \alpha_{j,k} b_{j,k}\big[\sigma\{h_j(\mathbf{X})\}\big], \tag{6}$$

其中，$g\,(\cdot)$是固定的连接函数，$\nu \in \mathbb{R}$ 是截距，$\boldsymbol{\gamma} \in \mathbb{R}^p$ 是未知的回归参数。

为清晰起见，本文聚焦于（6）中的逻辑斯蒂连接函数。为确保数值稳定性，施加 ℓ_2 惩罚后，估计问题表述为：

$$\underset{\nu,\,\boldsymbol{\gamma},\,\alpha_{j,k}}{\arg\min} \left[-\sum_{i=1}^{n} \Big\{ y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) \Big\} + \lambda_1 \|\boldsymbol{\gamma}\|_2^2 + \lambda_2 \sum_{j=1}^{J} \sum_{k=1}^{K-1} \alpha_{j,k}^2 \right], \tag{7}$$

其中

$$\hat{y}_i = \sigma\left(\nu + \boldsymbol{\gamma}^{\top}\mathbf{z}_i + \sum_{j=1}^{J} \sum_{k=1}^{K-1} \alpha_{j,k} b_{j,k}\big[\sigma\{h_j(\mathbf{X}_i)\}\big] \right),$$

其中，λ_1 和 λ_2 是调参参数。该表述对应于一个带惩罚项的逻辑斯蒂回归模型，其预测变量经过样条展开。该优化问题可以利用常用软件平台（如 Python 或 R）中提供的标准数值实现高效求解。

所提出的框架引入了多个调参组件，包括样条阶数ζ、基函数数量 K、节点布置策略以及正则化参数 λ_1 和 λ_2。为平衡建模灵活性与计算简便性，我们采用一组成熟且实用的选择。具体而言，按照通用惯例，我们将样条阶数设为 $\zeta = 4$，从而得到三次B样条基（Huang等人，2010年）。基函数数量选择为 $K = \lceil 2n^{1/5} \rfloor$，其中 $\lceil \cdot \rfloor$ 表示四舍五入到最近整数，遵循非参数回归文献中的建议（Fan等人，2014年）。采用数据驱动方式，利用等距经验分位数生成节点，以确保在不同特征分布下的稳健性。对于正则化，我们将$\lambda_1 = \lambda_2$ 设为相同值，以降低超参数搜索维度。执行小规模网格搜索，并根据留出的验证集上最高的受试者工作特征曲线下面积（AUROC）分数选择最终配置。

4 实验

本节使用大规模真实世界心电图数据集评估所提出的方法。我们首先介绍本研究所使用的两个数据集：一个传统心电图数据集，用于

后训练基础模型，以获得强大的预测器提取器；另一个是 ECG–ECHO 数据集，用于评估所提出的方法在结构性心脏病（SHD）检测中的效果。随后，我们研究该模型在 ECG–ECHO 数据集上的不同设置下的性能。

4.1 数据集

4.1.1 TraditionalECG 数据集

我们采用 PTB-XL 作为源数据集，以获得 $h_j(\cdot)\ \ j=1,\ \ldots\ldots,\ J$的最终权重。PTB-XL 是心电图建模中采用最广泛的基准资源之一，这得益于其数据集规模、丰富的标注结构以及标准化评估方案（Strodthoff，2024年；Wagner 等人，2020年）。它包含 21,837 条12导联心电图记录，每条时长 10 秒，采集自 18,885 名不同患者。每条记录由至多两名心脏病学家标注了 71 个传统心电图诊断类别（见表 1）。该数据集包含预定义的训练集、验证集和测试集。在本研究中，PTB-XL 仅用于心电图基础模型的后训练阶段。

表 1： 所提出模型中使用的临床变量和传统心电图诊断类别。7 个预处理后的临床变量，以及与 71 个临床定义的心电图诊断类别相关的估计风险，被纳入式（6）所定义的广义加性模型。诊断类别缩写遵循 SCP-ECG 术语；完整定义见Wagner 等人（2020年）。

7 个临床变量	71 种传统心电图诊断类别
sex，心室率，心房率、PR 间期，QRS 时限、QT 间期、患者年龄（截至心电图时间	正常、低电压、窦性心律、窦性心动过缓、异常QRS波群、下壁心肌梗死、窦性心律失常，心房扑动、心房颤动、非诊断性T波异常、非特异性ST段异常、洋地黄效应、左心室肥厚、左后分支传导阻滞、长QT间期，左前分支传导阻滞、右心房增大、不完全性右束支传导阻滞、右心室肥大、室内传导延迟、侧壁心肌梗死、前间壁心肌梗死，急性心肌梗死、一度房室传导阻滞、前侧壁缺血、窦性心动过速、心肌缺血，起搏心律，前侧壁缺血，SEHYP，下侧壁缺血，ILMI，PVC，完全性右束支传导阻滞，完全性左束支传导阻滞，动脉瘤，ALMI，缺血性心脏病，TAB_，高电压，LOWT，STD_，PAC，EL、NT_、Q波、T波倒置、PR间期延长、左心室肥厚电压标准、左心房负荷过重/左心房扩大，不完全性左束支传导阻滞，下壁缺血，SVTAC，INJAL，INJAS，IPMI，WPW，ISCAN，INJLA，BIGU，TRIGU，IPLMI，3AVB，INJIL，2AVB、PRC(S)、PSVT、PMI、STE_、INJIN、SVARR

4.1.2 心电图-超声心动图数据集

我们使用 EchoNext 数据集（Elias 和 Finer，2025年）评估所提出的方法。该数据集包含来自哥伦比亚大学欧文医学中心 36,286 名不同患者的 100,000 份去标识化心电图–超声心动图配对记录。每份记录均包含一段以 250 赫兹采样、时长为 10 秒的标准 12 导联心电图波形。人口统计信息和由心电图衍生的表格测量数据包括年龄、性别、心室率和心房率、PR 间期、QRS 时限及 QT 间期，如表 1所示。与先前研究（Poterucha 等人，2025年）一致，这 7 个临床变量经过预处理后，与原始心电图信号一同纳入我们的建模流程。

结构性心脏病（SHD）标签源自超声心动图报告，并以二元结局的形式提供。如第 1 节所述，结构性心脏病包括心脏瓣-

膜、心肌和心腔的多种异常，例如心脏瓣膜病和心力衰竭。先前研究表明，模型精确率会受到各结构性心脏病亚型患病率的影响；使用复合终点或可提高预测精确率，而不会增加临床负担，因为预测风险较高的患者无论如何都会被分诊接受超声心动图检查（Poterucha 等人，2025年）。因此，我们采用表示中度或重度结构性心脏病的复合标签作为预测目标。

数据集的使用遵循官方纳入与排除规则（Elias 和 Finer，2025年）。数据被划分为训练集、验证集和测试集，确保同一患者的所有心电图仅出现在一个划分中，以防止数据泄漏。训练集可以包含每位患者的多份心电图，以支持模型学习；而验证集和测试集则仅保留每位患者最近的一份心电图，以便无偏地评估泛化性能。应用官方筛选标准后，最终数据集包含 82,543 份心电图。表 2 总结了各数据集划分的特征。

表 2： ECG–ECHO 数据集中训练集、验证集和测试集的患者特征（Elias 和 Finer，2025年；Poterucha 等人，2025年）。数值以计数和百分比表示。

	训练集	验证集	测试集
患者（n）	26,218	4,626	5,442
心电图（n）	72,475	4,626	5,442
年龄组			
18–59	29,783 (41.1%)	1,787 (38.6%)	2,124 (39.0%)
60–69	18,745 (25.9%)	1,093 (23.6%)	1,318 (24.2%)
70–79	14,898 (20.6%)	975 (21.1%)	1,154 (21.2%)
80+	9,049 (12.5%)	771 (16.7%)	846 (15.5%)
Sex			
女性	33,524 (46.3%)	2,356 (50.9%)	2,731 (50.2%)
Male	38,951 (53.7%)	2,270 (49.1%)	2,711 (49.8%)
种族/族裔			
西班牙裔	22,806 (31.5%)	1,351 (29.2%)	1,649 (30.3%)
白人	21,289 (29.4%)	1,385 (29.9%)	1,569 (28.8%)
黑人	11,559 (15.9%)	728 (15.7%)	846 (15.5%)
亚裔	2,602（3.6%）	134（2.9%）	153（2.8%）
其他	5,272（7.3%）	380（8.2%）	457（8.4%）
未知	8,947（12.3%）	648（14.0%）	768（14.1%）
临床背景			
急诊	22,811 (31.5%)	1,688 (36.5%)	1,971 (36.2%)
住院	34,906 (48.2%)	1,903 (41.1%)	2,203 (40.5%)
门诊	12,423 (17.1%)	858 (18.5%)	1,059（19.5%）
操作相关	2,335（3.2%）	177（3.8%）	209（3.8%）

4.2 实验设置

4.2.1 心电图基础模型的后训练

如第 3.1 节所述，我们采用 ST-MEM 作为预测器提取器的骨干网络$h_j(\cdot)$ $j = 1, \ldots\ldots, J$。我们在包含 71 种传统心电图诊断标签的 PTB-XL 上，采用了 Zhou 等人 （2025年b）提出的后训练策略。具体而言，该后训练策略包含两个阶段：初始化阶段和正则化阶段。在初始化阶段，我们采用与（Na 等人，2024年）相同的线性探测配置，基于公开可用的权重训练 100 个周期。在正则化阶段，我们保留（Na 等人，2024年）中的微调超参数，同时加入 0.1 的随机深度率和 0.01 的 dropout 率，以增强稳健性并防止过拟合。我们使用这 71 个 ECG 诊断标签的平均二元交叉熵作为损失函数。模型在每个周期结束后于验证集上进行评估，并采用验证集宏平均 AUROC 最高的检查点。为清晰起见，我们将最终的预测器提取器表示为

$$\hat{\boldsymbol{h}}(\mathbf{X}) = \big(\sigma\{\hat{h}_1(\mathbf{X})\}, \cdots, \sigma\{\hat{h}_J(\mathbf{X})\}\big), \tag{8}$$

在此设置下， $J = 71$ 由于其辅助作用，我们仅报告后训练方法在基准测试集上的 AUROC，即 0.945；这一结果显著优于先前的小型深度学习模型（Strodthoff 等人，2021年）。

4.2.2 所提出的广义加性模型

基于式（8）中定义的预测器提取器和输入的 ECG 波形信号，我们为每个样本获得$J = 71$ 个预测因子。这些预测因子与第4.1.2节所述的 7 个预处理临床变量一起，被纳入式（7）所构建的优化问题中，以估计最终的结构性心脏病检测模型。如第3.3节所述，我们采用统计学界常用的超参数选择策略。具体而言，正则化参数设为 $\lambda = \lambda_1 = \lambda_2$，候选网格取值为 0.001、0.01、1、10、100、1000。遵循 Poterucha 等人（Poterucha 等人（2025年）的评估方案，使用受试者工作特征曲线下面积（AUROC）、精确率-召回率曲线下面积（AUPRC）和 F1 分数评估模型性能。对于每项指标，均基于 1000 个自助样本计算 95% 置信区间。

所提出的模型在完整训练数据集上进行训练，并在整个测试集上进行评估。为探究所提出方法的特性，我们针对不同群体开展亚组分析，包括不同年龄组、性别、种族和族裔以及临床背景。此外，我们通过改变训练样本量来研究所提出方法的样本效率；详见第 4.3 节。

4.3 结果

4.3.1 性能比较

我们将所提出的广义加性模型与五种方法进行比较：

(i) 使用7个临床变量的惩罚逻辑回归（逻辑回归–1），

(ii) 使用7个临床变量及（8）中定义的71个基础模型预测因子的惩罚逻辑回归（逻辑回归–2），(iii) 使用7个临床变量的支持向量机（支持向量机-1），(iv) 使用7个临床变量及（8）中定义的71个基础模型预测因子的支持向量机（支持向量机–2），以及 (v) 哥伦比亚小型模型（Poterucha 等人，2025年），代表当前最先进的水平。

为确保比较公平，所有惩罚回归模型均在相同的l_2 惩罚值网格上进行调参。两种支持向量机模型均采用径向基函数核，核宽度参数固定为默认的自适应设置（`gamma="scale"`），该设置位于`scikit-learn`中；正则化参数 C 则在与逻辑回归模型相同的网格上进行调参。Columbia mini模型采用公开发布的训练权重进行评估（Poterucha 等人，2025年）。为简洁起见，本节后文将所提出的方法称为加性模型。

表3 汇总了比较结果。所提出的加性模型在全部三项评估指标上均表现最佳，AUROC为 82.8，精确率-召回率曲线下面积为 79.7，F1分数为 71.8。与 Columbia mini模型相比，这些提升分别为 0.98%、1.01% 和 1.41%。值得注意的是，逻辑回归–2 和支持向量机-2 的性能均较逻辑回归–1 和支持向量机-1 有显著提升，分别凸显了源自基础模型的预测因子的预测效用。

为进一步评估模型在一系列决策阈值下的性能，图1展示了受试者工作特征（ROC）曲线和精确率-召回率（PR）曲线。在这两项指标上，对于大多数阈值，加性模型的曲线始终高于所有竞争方法的曲线，这表明其优势稳定，并非仅限于某个特定工作点的性能提升。

我们还通过使用逐步增大比例的可用训练集训练加性模型，考察了其样本效率。结果也列于表3，显示随着训练数据量从 10% 增至 100%，AUROC、精确率-召回率曲线下面积和F1分数均稳步提升。值得注意的是，即使仅使用 10% 的训练数据，加性模型也能达到与完全训练的 Columbia mini模型相当的 F1分数。仅使用 30% 的数据进行训练时，加性模型的表现便可媲美甚至超过使用完整数据集训练的 Columbia mini模型。这些发现凸显了所提方法出色的样本效率和稳健性。

4.3.2 亚组分析

为评估模型在不同患者特征和临床环境中的稳健性与公平性，我们开展了亚组分析，将加性模型与Columbia mini模型进行比较。亚组按年龄、性别、种族和族裔以及临床背景划分。所有性能指标均使用AUROC和精确率-召回率曲线下面积报告，并分别给出相应的95% 置信区间（表4）。

总体而言，加性模型在大多数亚组中表现出相当或更优的性能。这与近期研究中采用的评估设计一致（Poterucha 等人，2025年），

表 3： 所提出的加性模型在不同训练数据规模下的性能指标，并与基线方法并列呈现。所有数值均以括号内的95%自助法置信区间报告。标有“(our)”的模型表示使用本研究采用的心电图基础模型所衍生预测变量的模型。各指标的最佳性能以粗体突出显示。

模型	训练数据	AUROC	精确率-召回率曲线下面积	F1分数
逻辑回归-1	100%	74.8 (73.6–76.0)	67.7（65.6-69.8）	66.3（65.0-67.7）
支持向量机-1	100%	75.8 (74.5-77.0)	67.8（65.6-69.8）	66.3（65.1-67.8）
逻辑回归-2（我们的方法）	100%	80.8（79.6–81.9）	76.9（74.9–78.6）	70.9（69.7-72.4）
支持向量机-2（我们的）	100%	80.8 (79.7-82.0)	76.7（74.7-78.4）	71.0（69.7-72.4）
哥伦比亚 mini模型	100%	82.0（80.9–83.1）	78.9 (77.2–80.4)	70.8 (69.5–72.2)
加性模型（我们的）	10%	81.5 (80.4–82.6)	77.6 (75.8–79.4)	70.6（69.2–72.0）
	20%	81.9（80.8–83.0）	78.5（76.7–80.1）	71.1（69.8–72.5）
	30%	82.1（81.0–83.1）	78.9（77.2–80.5）	71.2（69.9–72.5）
	40%	82.2（81.1–83.3）	79.0（77.4–80.6）	71.3（69.9–72.7）
	50%	82.3（81.2–83.4）	79.2（77.6–80.8）	71.2（69.9–72.7）
	60%	82.4 (81.3–83.5)	79.4（77.7–81.0）	71.0 (69.6–72.3)
	70%	82.5（81.4–83.5）	79.6（78.0–81.1）	71.3 (70.0–72.6)
	80%	82.7 (81.6–83.7)	**79.7** (78.0–81.2)	**71.8** (70.5–73.2)
	90%	82.7 (81.6–83.7)	**79.7** (78.0–81.1)	71.6（70.3–73.1）
	100%	**82.8** (81.7–83.8)	**79.7** (78.1–81.2)	**71.8** (70.4–73.2)

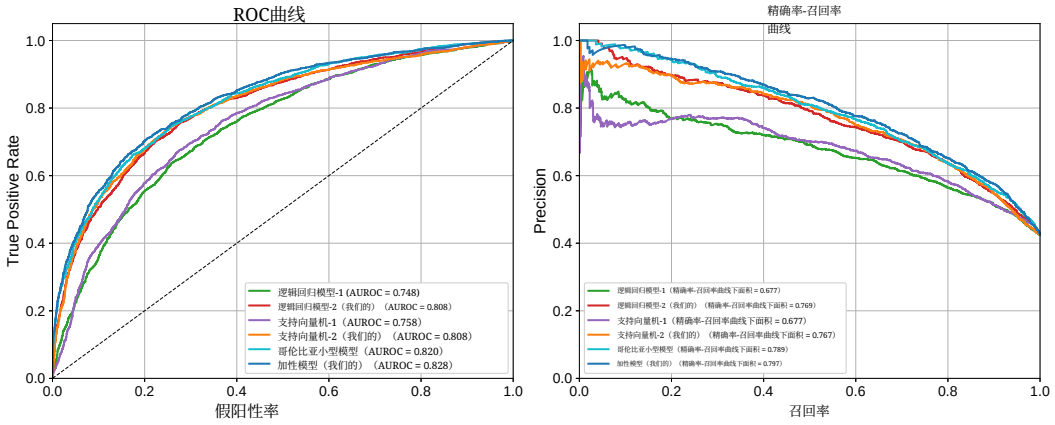


图 1： 四种不同模型的 ROC 曲线和 PR 曲线比较。左侧面板显示带有 AUROC 值的 ROC 曲线，右侧面板显示带有精确率-召回率曲线下面积值的 PR 曲线。标注“（我们的）”的模型表示使用了本研究所采用的心电图基础模型所派生预测因子的模型。

年龄和性别既作为模型输入，也作为亚组标识符。在 16 项亚组比较中的 14 项中，加性模型的 AUROC 和精确率-召回率曲线下面积均较高，其余比较则未显示实质性差异。

不同种族和族裔群体中的性能表现总体稳定。加性模型在每个亚组中的 AUROC 和精确率-召回率曲线下面积均为最高，包括西班牙裔、白人、黑人、亚裔以及其他/未知类别。亚裔等样本量较小的亚组置信区间较宽，反映的是样本量减少，而非区分能力减弱的证据。不同临床护理场景中的结果同样一致。在急诊、住院、门诊和操作环境中，加性模型均保持性能优势，其中门诊和操作相关就诊中的优势幅度最大。

综合来看，这些发现表明，加性模型能够很好地推广到异质的患者群体和护理环境中，体现了其在真实世界部署场景中的潜在应用价值。

表 4： 哥伦比亚 mini模型与我们的加性模型在人口统计学和临床亚组中的性能比较。所有数值均以括号内的 95% 置信区间表示。标注“（我们的）”的模型表示使用了由本研究所采用的心电图基础模型生成的预测因子。各亚组中每项指标的最佳表现以粗体标出。

		哥伦比亚 mini模型		加性模型（我们的）	
亚组	n	AUROC	精确率-召回率曲线下面积	AUROC	精确率-召回率曲线下面积
年龄组					
18–59	2124	83.0 (81.1–84.8)	73.7（70.7–76.8）	**83.5** (81.7–85.4)	**74.7** (71.8–77.7)
60–69	1318	80.6 (78.2–83.0)	75.2（71.6–78.7）	**81.8** (79.3–84.2)	**77.5** (73.8–81.0)
70–79	1154	77.9 (75.1–80.5)	80.4（77.3–83.2）	**79.0** (76.3–81.5)	**81.2** (78.1–84.1)
80+	846	75.9（72.7–79.0）	**85.5** (82.7–88.3)	**76.0** (72.9–79.1)	85.1（82.3–88.0）
Sex					
女性	2731	**81.9** (80.3–83.6)	**74.9** (72.5–77.6)	81.8（80.1–83.5）	74.7（72.0–77.3）
Male	2711	81.3（79.5–82.9）	81.7（79.6–83.6）	**82.9** (81.2–84.3)	**83.1** (81.0–85.1)
种族/族裔					
西班牙裔	1649	82.7（80.7–84.8）	76.8（73.6–79.9）	**83.3** (81.1–85.2)	**77.1** (74.0–80.4)
白人	1569	80.4（78.3–82.5）	76.1（72.9–79.3）	**81.5** (79.3–83.4)	**77.7** (74.7–80.5)
黑人	846	81.7 (78.8–84.5)	81.8（78.1–85.0）	**82.4** (79.5–84.9)	**82.6** （79.1–85.8）
亚裔	153	81.3（74.3–87.9）	81.8（73.5–89.3）	**84.1** （77.9–89.9）	**83.8** (76.2–90.2)
其他	457	81.5（77.8–85.2）	78.9 (74.0–83.8)	**82.4** (78.6–86.3)	**80.9** （76.1–85.4）
未知	768	83.3（80.4–86.1）	83.4（79.9–87.0）	**83.9** (80.9–86.6)	**83.5** (79.8–87.1)
临床背景					
急诊	1971	82.2（80.3–84.1）	77.6（74.8–80.2）	**83.5** (81.8–85.3)	**78.8** (76.1–81.4)
住院	2203	79.7 (77.8–81.5)	82.5（80.4–84.5）	**80.0** (78.2–81.7)	**82.8** (80.7–84.7)
门诊	1059	79.5（76.2–82.6）	64.5（59.0–70.2）	**80.5** （77.2–83.7）	**67.1** （61.5–72.6）
操作相关	209	79.5（73.2–85.3）	76.6（67.1–85.3）	**82.2** (75.9–87.4)	**80.7** (72.4–88.2)

4.3.3 估计的逐项函数

Let

$$\left\{\hat{\nu},\ \hat{\boldsymbol{\gamma}},\ (\hat{\alpha}_{j,k})_{j=1,k=1}^{j=J,k=K-1}\right\}$$

是优化问题（7）的解。为确保可识别性，我们重构了每个预测变量对应的逐项估计函数 j 作为

$$\hat{f}_j(x) := \sum_{k=1}^{K-1} \hat{\alpha}_{j,k} b_k(x) - \sum_{k=1}^{K-1} \int_0^1 \hat{\alpha}_{j,k} b_{j,k}(x) dx, \qquad j = 1, \ldots, J. \tag{9}$$

该中心化约束确保 $\int_0^1 \hat{f}_j(x)dx = 0$，从而防止各分量之间出现任意常数偏移，并使非线性效应能够得到一致的解释。

回顾一下，加性模型中使用的预测因子来源于心电图基础模型，其中每个预测变量表示模型对相应传统ECG诊断的估计概率。为说明这些诊断的部分效应，图2展示了四个估计函数示例 f_j（x），分别对应于下壁心肌梗死（IMI）、非诊断性T波异常（NDT）、心房颤动（AFIB）和左心室肥厚（LVH）。

在这些示例中，部分效应往往随预测值增加，这与临床直觉大体一致，即传统心电图异常的预测概率越高，可能与结构性心脏病的可能性越大相关。非线性关系的存在也有助于解释为何某些与结构性心脏病相关的细微心电图模式难以通过肉眼识别。尽管这些观察仍属初步结果，但它们表明，将某些非线性决策成分纳入未来基于心电图的临床指南，可能有助于支持更早发现结构性心脏病。

5 讨论

本研究提出了一种简单而有效的框架，将可解释的加性建模策略与源自心电图基础模型的预测因子相结合，用于通过标准12导联心电图检测结构性心脏病（结构性心脏病）。基于一个包含8万多份心电图的基准数据集，所提出的方法表现出优于 Columbia mini模型的一致提升（Poterucha 等人，2025年），AUROC、精确率-召回率曲线下面积和F1分数分别提高了0.98%、1.01%和1.41%。此外，不同训练样本量下的实验表明，该方法的数据效率有所提高：仅使用30%的训练数据，所提出的方法就取得了略优于 Columbia mini模型的性能。在不同患者亚群和护理环境中开展的亚组分析进一步证明了该方法的稳健性。重要的是，加性模型中的每个预测因子都对应一个具有临床意义的潜在预测因子。估计得到的逐项函数揭示了这些预测因子的风险与结构性心脏病之间的非线性关联。综上，这些特性使该方法成为一种兼具可解释性和实际应用价值的工具，能够支持临床工作流程中的透明决策。

本研究提出了一种通过心电图数据检测结构性心脏病的新范式。传统上，人们认为结构性心脏病无法通过心电图诊断，因为相关的心电图模式可能十分细微，难以被人类察觉（Siontis 等人，2021年）。近期

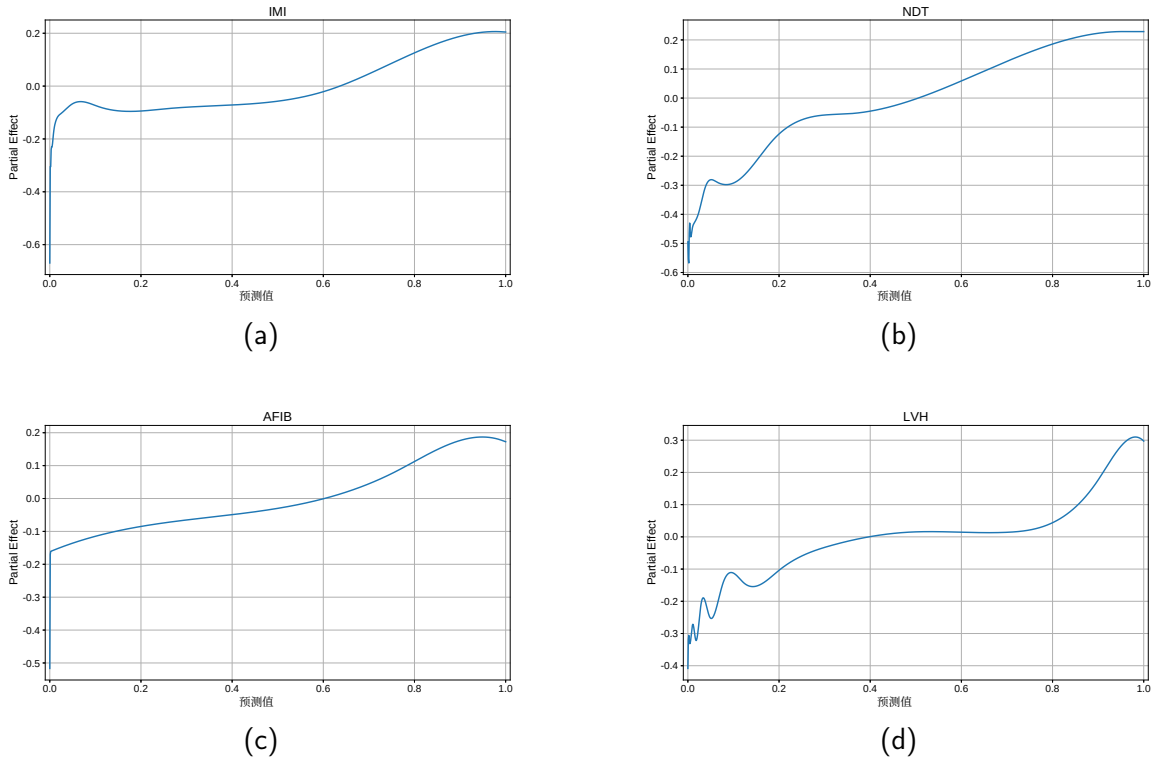


图2：所提出的加性模型中估计的逐项函数示例。“部分效应”ˆ ˆ指 $f_j(x)$，其定义见（9），而“预测值”指相应的 $\sigma\{h_j(\mathbf{X})\}$，其中ˆ$h_j(\mathbf{X})$ 定义见（8）。标题 IMI、NDT、AFIB 和 LVH 分别表示下壁心肌梗死、非诊断性T波异常、心房颤动和左心室肥厚。

研究进展表明，通过在大型配对心电图–超声心动图数据集上训练深度学习模型，可以发现此类模式（Attia 等人，2019年；Cohen-Shelly 等人，2021年；Diao 等人，2025年；Poterucha 等人，2025年）；然而，这些方法以端到端黑箱系统的形式运行，限制了透明度和临床可解释性。所提出的框架代表了一种概念上的转变：不再仅将深度学习用于端到端预测，而是将可接受的深度学习方法的输出转换为可解释预测变量，并采用加性结构进行建模。这种表述保留了现代人工智能（AI）模型所捕获的预测信号，同时能够对每个预测变量的贡献进行明确的可视化和解释，为基于心电图的结构性心脏病检测提供一种更具可解释性且更便于临床应用的方法。

在这些发现的基础上，本研究还强调了统计学与人工智能的互补作用，正如Redman 和 Hoerl （2024年）最近所强调的那样。尽管现代人工智能模型已展现出强大的预测能力，但它们往往以牺牲透明度和可解释性为代价。可解释人工智能旨在通过设计不仅能够做出准确预测、还能提供可解释依据的模型来弥合这一差距（Broniatowski 和 Broniatowski, 2021年）。然而，这种透明度可能需要在预测准确性或计算效率方面作出权衡（Min 等人，2024年）。当前研究正致力于解决这些缺陷（Liu 等人，2025年；Longo 等人，2024年）。所提出的框架通过整合临床认可且已在实践中得到应用的预测变量与经典统计建模方法，为这一方向作出了贡献。本文结果表明，在保留可解释性的同时，可以保留现代人工智能的大部分预测优势，这为人工智能时代未来应用统计方法学的发展提供了潜在路径。

从临床角度来看，估计得到的平滑函数揭示了传统心电图诊断风险与结构性心脏病之间的非线性关联。尽管与结构性心脏病相关的电学特征通常难以通过目视察觉，基础模型预测因子仍反映了临床工作流程中常用且可解释的心电图特征。观察到的非线性结构提示，既有的心电图标准可能包含与结构性心脏病相关、但尚未系统纳入临床指南的潜在诊断信息。这些发现推动未来研究将估计效应转化为具有临床可操作性的决策阈值、风险方程或筛查指南，以实现基于心电图的结构性心脏病早期检测。

最后，本研究也存在若干局限。首先，结构性心脏病包含多种表型，但本分析聚焦于复合结局。尽管这遵循了既往研究中确立的做法（Poterucha 等人，2025年；Ulloa-Cerna 等人，2022年），但所提出的框架能否在各个结构性心脏病亚型标签中保持一致的泛化能力，目前仍不得而知。其次，本研究仅使用单个心电图基础模型进行评估。采用更大规模或更多样化数据集训练的较新模型（Diao 等人，2025年）可能提供更优的潜在预测变量，并进一步提升性能。第三，实证分析基于单中心数据，缺乏外部或多中心验证，因此对模型泛化能力的评估受到限制。这些局限也构成了未来研究的积极方向。

6 结论

总之，本研究提出了一种新颖的框架，将可解释的基础模型衍生心电图预测因子与加性建模结构相结合，以通过标准 12 导联心电图检测结构性心脏病。实验表明，该方法在保持预测准确性和稳健性的同时，还能清晰呈现各个预测因子对风险估计的贡献，展现了一种兼具可解释性与实际应用价值的方法。这些发现也表明，统计学与现代人工智能能够互为补充。未来的工作将评估该方法在各个结构性心脏病亚型、其他基础模型架构以及外部多中心数据集上的表现，进一步考察其泛化能力和临床效用。

参考文献

Attia, Z. I. 等（2019）利用人工智能辅助心电图筛查心脏收缩功能障碍。《自然医学》，**25**，70–74。Behrouz, A.、Santacatterina, M. 和 Zabih, R.（2024）Chimera：利用二维状态空间模型有效建模多变量时间序列。载于 第三十八届神经信息处理系统年会。网址：https://openreview.net/forum?id=ncYGjx2vnE。Broniatowski, D. A. 和 Broniatowski, D. A.（2021）人工智能中可解释性与可阐释性的心理学基础，第4卷。美国商务部、美国国家标准与技术研究院。Cheng, R.、Zhuang, Z.、Zhuang, S.、Xie, L. 和 Guo, J.（2023）Msw-transformer：用于12导联心电图分类的多尺度移位窗口Transformer网络。*arXiv*预印本 *arXiv:2306.12098*。Cohen-Shelly, M.、Attia, Z. I.、Friedman, P. A.、Ito, S.、Essayagh, B. A.、Ko, W.-Y.、Murphree, D. H.、Michelena, H. I.、Enriquez-Sarano, M.、Carter, R. E. 等（2021）利用人工智能通过心电图筛查主动脉瓣狭窄。《欧洲心脏杂志》，**42**，2885–2896。Coppola, E.、Savardi, M.、Massussi, M.、Adamo, M.、Metra, M. 和 Signoroni, A.（2024）HuBERT-ECG：一种用于广泛、可扩展心脏应用的自监督基础模型。*medRxiv*，2024年11月。d'Arcy, J. L.、Coffey, S.、Loudon, M. A.、Kennedy, A.、Pearson-Stuttard, J.、Birks, J.、Frangou, E.、Farmer, A. J.、Mant, D.、Wilson, J. 等（2016）大规模社区超声心动图筛查揭示了老年人中大量未确诊的心脏瓣膜病：OxVALVE人群队列研究。《欧洲心脏杂志》，**37**，3515–3522。Diao, X.、Xu, W.、Cheng, H.、Zhou, Y.、Liu, Y.、Huo, Y.、Lu, J.、Huang, J.、He, J.、Liu, F. 等（2025）Speed-tr：一种自蒸馏预训练Transformer模型，用于增强心电图对三尖瓣反流的检测。《*npj*数字医学》，**8**，650。

Elias, P. 和 Finer, J.（2025）EchoNext：一个用于从心电图检测超声心动图确诊的结构性心脏病的数据集。Fan, J.、Ma, Y. 和 Dai, W.（2014）稀疏超高维变系数模型中的非参数独立性筛选。《美国统计学会杂志》，109，1270–1284。Hannun, A. Y. 等（2019）利用深度神经网络对动态心电图进行心脏病专家级心律失常检测与分类。《自然医学》，25，65–69。Hastie, T. 和 Tibshirani, R.（1990）《广义加性模型》，第43卷。CRC出版社。Hata, E.、Seo, C.、Nakayama, M.、Iwasaki, K.、Ohkawauchi, T. 和 Ohya, J.（2020）利用深度学习对心电图进行主动脉瓣狭窄分类，并使用Grad-CAM进行分析。载于2020年第42届IEEE医学与生物学工程学会国际年会（EMBC），1548–1551页。IEEE。He, K.、Chen, X.、Xie, S.、Li, Y.、Dollár, P. 和 Girshick, R.（2022）掩码自编码器是可扩展的视觉学习器。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，2022年，第16000–16009页。Huang, J.、Horowitz, J. L. 和 Wei, F.（2010）非参数加性模型中的变量选择。《统计年鉴》，38，2282。Jiang, A.、Huang, C.、Cao, Q.、Xu, Y.、Zeng, Z.、Chen, K.、Zhang, Y. 和 Wang, Y.（2024）自监督异常检测预训练提升长尾心电图诊断能力。arXiv预印本 arXiv:2408.17154。Kwon, J.-M.、Lee, S. Y.、Jeon, K.-H.、Lee, Y.、Kim, K.-H.、Park, J.、Oh, B.-H. 和 Lee, M.-M.（2020）基于深度学习的心电图主动脉瓣狭窄检测算法。《美国心脏协会杂志》，9，e014717。Li, J.、Aguirre, A. D.、Junior, V. M.、Jin, J.、Liu, C.、Zhong, L.、Sun, C.、Clifford, G.、Brandon Westover, M. 和 Hong, S.（2025）基于超过1000万条记录构建的心电图基础模型。NEJM AI，2，AIoa2401033。Liu, Z.、Wang, Y.、Vaidya, S.、Ruehle, F.、Halverson, J.、Soljacic, M.、Hou, T. Y. 和 Tegmark, M.（2025）KAN：Kolmogorov–Arnold网络。载于第十三届国际学习表征会议。网址：https://openreview.net/forum?id=Ozo7qJ5vZi。Longo, L.、Brcic, M.、Cabitza, F.、Choi, J.、Confalonieri, R.、Del Ser, J.、Guidotti, R.、Hayashi, Y.、Herrera, F.、Holzinger, A. 等（2024）可解释人工智能（XAI）2.0：开放挑战与跨学科研究方向宣言。《信息融合》，106，102301。McKeen, K.、Masood, S.、Toma, A.、Rubin, B. 和 Wang, B.（2024）ECG-FM：开放式心电图基础模型。arXiv预印本 arXiv:2408.05178。Mensah, G. A.、Fuster, V.、Murray, C. J.、Roth, G. A.、of Cardiovascular Diseases, G. B. 和 Collaborators, R.（2023）1990年至2022年心血管疾病及其危险因素的全球负担。《美国心脏病学会杂志》，82，2350–2473。

Min, J., Song, X., Zheng, S., King, C. B., Deng, X. 和 Hong, Y.（2024年）人工智能时代的应用统计学：综述与展望。*arXiv* 预印本 *arXiv:2412.10331*。

Na, Y., Park, M., Tae, Y. 和 Joo, S.（2024年）引导掩码表征学习，以捕捉心电图的时空关系。载于第十二届国际学习表征会议。网址：https://openreview.net/forum?id=WcOohbsF4H。

Poterucha, T. J., Jing, L., Ricart, R. P., Adjei-Mosi, M., Finer, J., Hartzel, D., Kelsey, C., Long, A., Rocha, D., Ruhl, J. A. 等（2025年）利用人工智能从心电图中检测结构性心脏病。《自然》，**644**，221–230。Redman, T. C. 和 Hoerl, R. W.（2024年）人工智能与统计学：天作之合。《麻省理工斯隆管理评论》*(*在线版*)*，1–4。Ribeiro, A. H. 等（2020年）利用深度神经网络自动诊断 12 导联心电图。《*Nature Commun.*》，**11**，1760。Shi, J., Wang, K. A. 和 Fox, E.（2023年）基于多分辨率卷积记忆的序列建模。载于国际机器学习会议，31312–31327。PMLR。Siontis, K. C., Noseworthy, P. A., Attia, Z. I. 和 Friedman, P. A.（2021年）用于心血管疾病管理的人工智能增强型心电图检查。《*Nature Rev. Cardiol.*》，**18**，465–478。Somani, S. 等（2021年）深度学习与心电图：当前最先进水平综述。《欧洲心律》，**23**，1179–1191。Stone, C. J.（1985年）加性回归及其他非参数模型。《统计年鉴》，689–705。Strodthoff, N.（2024年）以开放科学推动自动心电图分析的发展：现状与未来方向。Strodthoff, N., Wagner, P., Schaeffter, T. 和 Samek, W.（2021年）用于心电图分析的深度学习：来自 PTB-XL 的基准与见解。《*IEEE J. Biomed. Health Inform.*》，**25**，1519–1528。Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Anderson, C. A., Arora, P., Avery, C. L., Baker- Smith, C. M., Beaton, A. Z., Boehme, A. K., Buxton, A. E. 等（2023年）心脏病与卒中统计——2023年更新：美国心脏协会报告。《循环》，**147**，e93–e621。Ulloa-Cerna, A. E., Jing, L., Pfeifer, J. M., Raghunath, S., Ruhl, J. A., Rocha, D. B., Leader, J. B., Zimmerman, N., Lee, G., Steinhubl, S. R. 等（2022年）Rechommend：一种基于心电图的机器学习方法，用于识别未确诊结构性心脏病风险升高且可通过超声心动图检测的患者。《循环》，**146**，36–47。Wagner, P. 等（2020年）PTB-XL：一个大型公开可用的心电图数据集。《*Sci. Data*》，**7**，1–15。

Wang, Z., Ma, L., Liu, M., Fan, J., Hu, S., 心血管健康报告编写委员会（W. C.）和中国疾病（D.）（2023年）《2022年中国心血管健康与疾病报告》摘要。《中华医学杂志》，**136**，2899–2908。Zhang, M., Saab, K. K., Poli, M., Dao, T., Goel, K. 和 Re, C.（2023年）利用简洁的离散状态空间有效地建模时间序列。载于《国际学习表征会议论文集》。Zhou, Y., Diao, X., Huo, Y., Liu, Y., Sun, Z., Fan, X. 和 Zhao, W.（2025年a）增强心电图信号的自动多标签诊断：一种掩码Transformer方法。《生物学与医学中的计算机》，**196**，110674。Zhou, Y., Yang, Y., Fan, X. 和 Zhao, W.（2025年b）弥合基础模型的性能差距：面向 ECGFounder 的后训练策略。*arXiv* 预印本 *arXiv:2509.12991*。