

# 用于闭塞性心肌梗死的心电图诊断和风险分层的机器学习

## 摘要

闭塞性心肌梗死（OMI）且心电图无 ST 段抬高的患者越来越多。这些患者的预后较差，需要立即进行再灌注治疗，但目前还没有准确的工具可以在初步分诊时识别这些患者。据我们所知，我们在此报告了第一项为 OMI 的心电图诊断开发机器学习模型的观察性队列研究。通过使用来自多个临床站点的 7313 名连续患者，我们得出了一个智能模型并进行了外部验证，该模型的表现优于临床医生和其他广泛使用的商业解释系统，大大提高了精确度和灵敏度。我们推导出的 OMI 风险评分提高了常规护理相关的入院和出院准确性，当与训练有素的急诊人员的临床判断相结合时，它有助于对三分之一的胸痛患者进行正确的重新分类。驱动我们模型的心电图特征得到了临床专家的验证，为心肌损伤提供了可信的机理联系。

## 概要

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。指南主要以 ST 段抬高（STE）为依据来区分 ST 段抬高型心肌梗死（STEMI）患者和其他形式的 ACS<sup>5-8</sup>。在心电图上没有 STE 的情况下，建议采用生物标志物驱动的方法。这种诊断模式有两个重要的局限性。首先，约 24%-35% 的非 STEMI 患者存在冠状动脉全闭塞，即闭塞性心肌梗死（OMI），需要进行紧急导管检查。虽然文献中经常描述 OMI 的重要心电图特征，但这些特征并不明显，涉及整个 QRST 复合体，且具有空间性（即多个导联的变化被稀释。因此，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。

第二个局限是，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，约三分之一的患者在连续采样后仍处于生物标记物不确定的“观察区”。更重要的是，约 25% 的急性心肌梗死病例初始 hs-cTn 为阴性，这在 STEMI 和 OMI 亚组中均可观察到。因此，25%-30% 的 OMI 患者未得到及时治疗，而在急诊科接受胸痛评估的患者中，约有 63%（四分位间范围为 38%-81%）因初步评估结果不确定而入院治疗。这些诊断上的局限性造成了一种高成本、低效率的临床实践模式，即大多数胸痛患者受到过度监护，而一些 OMI 患者却延误了诊断和治疗，这可能是导致非 STE ACS（NSTE-ACS）组患者死亡风险高出 14-22% 的原因。

在我们之前的工作中，我们设计了人工智能（AI）支持的心电图分析原型算法，并证明了在院前环境中筛查 ACS 的临床可行性。据我们所知，我们在此描述了首个多地点、前瞻性、观察性队列研究，以评估机器学习在首次医疗接触时和无 STEMI 模式时对 OMI 的心电图诊断和风险分层的诊断准确性（扩展数据图 2）。我们的智能模型是在美国多个临床站点的 7313 名胸痛患者身上得出并经过外部验证的。结果表明，在没有 STEMI 模式的情况下，机器学习在检测指示 OMI 的细微缺血性心电图变化方面具有优越性，优于临床医生和其他广泛使用的商业心电图解读软件。我们确定了驱动模型分类的最重要的心电图特征，并确定了与心肌损伤之间似是而非的机理联系。与 HEART 评分相比，我们推导出的 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。这种新临床路径在临床结果方面的益处应在前瞻性试验中进行评估。

## 结果

### 样本特征

在排除心脏骤停、室性心动过速、确诊院前 STEMI 和重复心电图的患者后，我们的衍生队列包括 4026 名连续胸痛患者（年龄为  $59 \pm 16$  岁，47% 为女性，5.2% 为 OMI）。两个外部验证队列共包括 3287 名患者（年龄为  $60 \pm 15$  岁，45% 为女性，6.4% 为 OMI）（图 1 和表 1）。推导组和验证组中的大多数患者均为正常窦性心律（>80%），约 10% 为心房颤动。约 3% 的患者有左束支传导阻滞（BBB），约 10%

的患者有左心室肥厚 (LVH) 的心电图证据。推导队列和验证队列在年龄、性别、基线临床特征和 30 天心血管死亡率方面相似。不过，验证队列中黑人和西班牙裔少数民族裔较多，ACS 和 OMI 的发生率略高。

## 算法推导和测试

心肌梗死溶栓 (TIMI) 血流分级为 0-1 级的冠状动脉是模型训练的阳性级别，由对所有心电图分析保密的独立审稿人根据病历判定。TIMI 血流分级为 2 级、冠状动脉严重狭窄 (>70%) 且第四代肌钙蛋白 (非高灵敏度) 峰值为 5-10 纳克/毫升也是 OMI 的指征。模型训练的阴性类别是无 OMI，包括所有其他非 ACS 病因或非冠状动脉闭塞性 ACS 亚型。

模型训练的输入数据基于院前 12 导联心电图。我们采用数据驱动和领域专业知识混合方法，从 554 个时空指标中挑选出 73 个形态心电图特征<sup>18</sup>。利用这些特征，我们训练了 10 个分类器来学习 ACS 组和非 ACS 组之间的缺血模式，并估计 OMI 的概率。我们选择这些分类器是为了最大限度地找到最适合的方法来学习将复杂心电图数据与潜在生理学相关联的数学表示。

随机森林 (RF) 模型在训练和内部测试中实现了最佳偏差-方差权衡。我们将随机森林模型与临床医师的心电图判读进行了比较，并与美国食品药品监督管理局 (FDA) 批准用于 "急性心肌梗死" 诊断的商用心电图判读系统的性能进行了比较。在保留测试集上，RF 模型 (接收器操作特征下面积 (AUROC) 0.91 (95% 置信区间 (CI) 0.87-0.96) ) 的表现优于临床医师 (AUROC 0.79 (95% CI 0.73-0.76) ,  $P < 0.001$ ) 和商用心电图系统 (AUROC 0.78 (95% CI 0.70-0.85) ,  $P < 0.001$ ) (图 2a)。

接下来，我们使用 OMI (+) 和 OMI (-) 等级的概率密度图来表示风险预测的最佳分离边缘。根据指南<sup>6</sup> 的建议，我们定义了一个风险评分来识别低风险 (OMI 评分 <5)、中度风险 (OMI 评分 5-20 分) 和高度风险 (OMI 评分 >20) 患者，这些临界值在不同等级之间产生了极好的分离 (log-rank chi-square, 133.04; 自由度 = 2;  $P < 0.001$ ) (图 2b 左)。我们的 OMI 评分将 74.4% 的患者划分为低风险，4.6% 的患者划分为高风险。在排除策略中使用低风险组的灵敏度为 0.91，阴性预测值 (NPV) 为 0.993，总体漏诊率为 0.5%。使用高危组来制定 "纳入" 策略的特异性为 0.976，阳性预测值 (PPV) 为 0.514，总体误诊率为 2%。最后，我们将 OMI 评分与 HEART 评分进行了比较，后者使用了患者病史、心电图数据、年龄、风险因素和肌钙蛋白值 (图 2b, 右)。我们的 OMI 评分仅基于心电图数据，与 HEART 评分相比，多 66% 的患者被归类为低风险，假阴性率小于 1%，与 HEART 评分相似，而被归类为高风险的患者更少，精确度更高 (51% 对 33%)。OMI 评分也减少了 50% 的中危患者，但在 OMI 检测方面仍有更好的区分度 (11.2% 对 5.6%)。

## 模型的可解释性

我们使用树形 SHAP 算法生成了一个重要性排序，根据对前 25 个特征估计的 SHAP 值来解释 RF 模型的输出 (图 3a)。对分类输出影响最大的特征包括：V1、V2、I 和 aVL 导联的 ST 轻度压低；III 和 V4-V6 导联的 ST 轻度抬高；前导联的凹陷模式消失；II 和 aVF 的 T 波增大以及 I 和 aVL 的 T 波变平或倒置；T 峰-腱间期延长；T 轴偏离；再极化弥散增加；以及激活和恢复模式的方向扭曲。这些心电图模式大多与心肌缺血有生理上的联系，这表明它们作为 OMI 检测特征具有临床价值。

为了更直观地显示我们的模型检测到的这些全局心电图模式，我们创建了 OMI (+) 类患者 (n = 414 张心电图) 的集合人群中位搏动，并将这些中位搏动叠加到正常窦性心律和 OMI (-) 状态患者 (n = 9072 张心电图) 的集合人群中位搏动上 (图 3b)。该图的结果支持上述 SHAP 值所描述的模式。具体而言，OMI 与 V1-V2、I 和 aVL 的 ST 压低和 T 波变平有关；前导联 ST 轻度抬高，凹陷模式消失；下导联 T 波峰值；T 峰 - Tend 延长 (见于许多导联)；全局性的再极化弥散 (表现为某些导联的 T 波峰值和其他导联的平坦)；T 轴偏离 (偏离左心室)；以及激活和恢复模式扭曲 (在水平面上表现为心前区导联的 R 波进展消失，T 波不协调性增加)。由于该队列中普遍存在多血管疾病，因此无论罪魁祸首位于何处，这些 OMI 模式都相对一致。

尽管如此，为了检查局部特征重要性的可解释性，我们在单个病例上使用了力图，以确定在特定心电图上符合射频模型贡献阈值的特征。研究人员也对这些力图进行了检查，以进一步证实模型预测的临床有效性。扩展数据图 3 显示了 12 导联心电图的一个选定示例及其相应的局部特征贡献力图。

## 外部验证

我们在两个独立的外部临床机构的 3287 名患者身上测试了最终的锁定模型。机器学习工程师对其他研究机构的结果数据视而不见，预填充模型预测结果由临床研究人员独立评估。我们的模型具有良好的通用性，并保持了较高的分类性能（AUROC 0.87 (95% CI 0.85-0.90)），优于商业心电图系统（AUROC 0.75 (95% CI 0.71-0.79)， $P < 0.001$ ）和临床医师（AUROC 0.80 (95% CI 0.77-0.83)， $P < 0.001$ ）（图 4a）。我们的 OMI 风险评分是预测 OMI 的有力指标，不受年龄、性别和其他冠状动脉风险因素的影响（高风险等级的比值比 (OR) 为 10.60 (95% CI 6.78-16.64)，中风险等级的比值比 (OR) 为 2.85 (95% CI 1.91-4.28)）（图 4b）。该风险评分将 69% 的低风险组患者分流，假阴性率为 1.3%，将 5.1% 的患者识别为高风险，可接受的真阳性率大于 50%。OMI 规则入和规则出策略的总体灵敏度、特异性、PPV 和 NPV 分别为 0.86 (95% CI 0.81-0.91)、0.98 (95% CI 0.97-0.99)、0.54 (95% CI 0.46-0.62) 和 0.99 (95% CI 0.98-0.99)。这一诊断准确率在基于年龄、性别、种族、合并症和基线心电图结果的亚组中保持相对相似，表明没有聚集偏差（图 4c）。相比之下，临床医生对心电图重读的敏感性、特异性、PPV 和 NPV 分别为 0.58、0.93、0.36 和 0.97，而商用心电图系统的敏感性、特异性、PPV 和 NPV 分别为 0.79、0.80、0.22 和 0.98。

接下来，我们评估了我们得出的风险评分在对首次就医患者进行重新分类时的增益（图 5）。急诊人员的初步评估基于修改后的 HEAR（病史、心电图、年龄和危险因素）评分，将患者分流为低危、中危和高危组 36。基线时，急救人员将 48% 的患者分流为低风险，NPV 为 99.0%，将 3% 的患者分流为高风险，PPV 为 54.1%。近 50% 的患者仍处于不确定的观察区域。应用我们的 OMI 风险评分将有助于将 45% 的患者分流为低风险，同时将 NPV 保持在 98.8%，并有助于检测出 85% 的 OMI 患者，同时将 PPV 保持在 50.0%。OMI 评分还有助于将处于不确定观察区的患者人数减少一半以上。这些数字换算成净再分类改善 (NRI) 指数为 41% (95% CI 33-50%)。为了验证这一临床效用的增加，我们手动查看了被正确重新分类为 OMI (+) 的心电图（扩展数据图 4）。其中许多心电图显示了微妙或非特异性的变化，根据指南 5，这些变化是非诊断性的，这表明它在增强医疗人员解释“模糊”心电图时的信心方面具有潜在价值。

最后，我们调查了验证数据中假阴性的潜在来源。在漏报 OMI 事件的患者中 ( $n = 28$ , 0.9%)，许多患者的初始心电图存在高频噪声和基线游走 ( $n = 13/28$ , 46%) 或低电压心电图 ( $n = 14/28$ , 50%)，大多数患者 ( $n = 24/28$ , 86%) 的良性心电图没有任何可诊断的 ST-T 变化（扩展数据图 5）。此外，我们还发现假阴性和真阳性患者在人口统计学或临床特征方面没有明显差异，但大多数假阴性患者既往有心肌梗死病史 (93% 对 27%)。鉴于我们的 OMI 模型对已知冠状动脉疾病 (CAD) 患者的特异性略低（图 4c），后一个发现很耐人寻味。这一发现与最近的证据显示胸痛和已知冠状动脉疾病患者的 NPV 值降低相一致。

## 筛查任何 ACS 事件

我们进一步建立了一个模型，用于在首次医疗接触时筛查任何潜在的 ACS 事件。使用同一组心电图特征，我们训练并优化了一个射频分类器，该分类器可表示发生任何 ACS 事件的可能性。该模型在训练过程中表现良好（AUROC 0.88 (95% CI 0.87-0.90)），在内部测试过程中泛化良好（AUROC 0.80 (95% CI 0.76-0.84)）（扩展数据图 6）。在外部验证中，该模型继续保持良好的通用性（AUROC 0.79 (95% CI 0.76-0.8)），优于商业系统（AUROC 0.68 (95% CI 0.65-0.71)， $P < 0.001$ ）和临床医师（AUROC 0.72 (95% CI 0.69-0.74)， $P < 0.001$ ）。我们得出的风险评分为任何 ACS 事件提供了次优的排除分类（灵敏度 68.2% 和 NPV 92.5%），但提供了优越的入选准确性（特异性 98.9% 和 PPV 82.5%）。

## 讨论

在这项研究中，我们开发并验证了一种机器学习算法，用于对从美国多个临床站点招募的连续胸痛患者进行 OMI 的心电图诊断。该模型的表现优于临床医生和其他商业解释系统。与参考标准相比，得出的风险评分提供了更高的 OMI 规则入和规则出准确性，灵敏度提高了约 28 个百分点，精确度提高了约 32 个百分点。结合经验丰富的急诊人员的判断，我们推导出的 OMI 风险评分有助于对三分之一的胸痛患者进行正确的重新分类。据我们所知，这是第一项使用机器学习方法和新型心电图特征来优化急性胸痛患者的 OMI 检测的研究，这些患者的心电图呈阴性 STEMI 模式。

心肌缺血是一个区域代谢紊乱问题，而冠状动脉闭塞是动脉粥样硬化斑块破裂导致血流减少的问题，将心肌缺血与冠状动脉闭塞相联系是一个复杂的过程<sup>1</sup>。从本质上讲，缺血会不成比例地扭曲不同心肌节段的动作电位，从而产生组织规模的电流，通常称为"损伤"电流。以往的研究将明显的 ST 抬高与冠状动脉全闭塞相关的跨膜损伤电流相联系。在确定哪些患者可从紧急再灌注治疗中获益时，这一直是目前 STEMI 与"其他"(除 STEMI 以外的任何 ACS)二分法的驱动因素。然而，近 65% 的 ACS 患者在基线心电图上没有 ST 段抬高，而在后者中，24-35% 的患者冠状动脉完全闭塞，需要进行紧急导管治疗。因此，确定哪些患者可从再灌注治疗中获益仍是一项裁定性诊断。

从概念上讲，缺血心肌细胞产生的损伤电流具有总结性质，这也解释了 ST 波幅的变化如何在表面心电图上减弱(扩展资料图 7)。然而，这些损伤电流会扭曲兴奋和恢复途径的传播，从而彻底改变 QRS 波群和 ST-T 波形的构型。因此，更全面的心电图缺血检测方法应侧重于：(1) 评估整个波形片段的时间特征，而非特定时间点的电压(例如，J 点 + 80 毫秒)；(2) 评估波形形态中导联与导联之间的空间特征，而非孤立心电图导联的绝对变化。

这项研究发现了几种表明急性冠状动脉闭塞的心电图模式，超出了临床指南推荐的标准。耐人寻味的是，这些心电图模式与文献中描述的心电图模式重叠。2012 年的一份共识报告指出，在急性疼痛发作时，有几种心电图模式应被视为 STEMI：V1-V3 的 ST 压低；V1-V3 的小倒置 T 波；心前导联的深负 T 波；广泛的 ST 压低；突出的正 T 波。最近也出现了类似的心电图模式：V1-V4 (相对于 V5-V6) 的 ST 波压低；往复 ST 波压低，最大 ST 波压低向量朝向心尖(导联 II 和 V5, aVR 中的往复 STE)；细微的 ST 波抬高；急性病理性 Q 波；超急性 T 波；末端 S 波消失。这些由专家主导的模式中，许多都依赖于评估再极化波幅的比例或 QRS 波幅下的面积。它们在很大程度上还依赖于对波形形态的直观评估，并可能在心电图判读员之间引入高度的主观性和变异性。我们证明，机器学习模型不仅在识别 OMI 方面优于临床医师，而且还提供了一种客观、不依赖于观察者的方法来量化与 OMI 相关的微妙心电图模式。

我们的机器学习模型识别出的许多数据驱动特征都很微妙，临床专家不易察觉。T 波特征指数是这些最重要的特征之一，包括 T 峰-T 峰间期延长、T 波变平和 T 峰前拐点处的 T 波特征(图 3a)。从机制上讲，缺血性损伤电流会干扰信号传播，导致激活时间延长。这些晚期激活电位导致末端 S 波消失和恢复时间延长，两者都表现为 T 波变平、T 峰移动和初始 T 波凹陷消失(图 3b)。这些 STEMI 等效模式以前在文献中被描述为小的或负的 T 波，伴有广泛的 ST 凹陷或微妙的 ST 抬高。我们的模型发现的另一个重要的微妙特征是心室复极化弥散增加，该特征通过 ST-T 波形的主成分(即主成分分析(PCA)指标)之间的比率、T 轴方向以及激活和恢复途径之间的角度(例如，总余弦 R-T)来测量。损伤电流会不成比例地影响不同心肌节段的复极化持续时间和速度，从而导致 ST-T 波形形态的导联间差异。这些高风险心电图模式以前曾被描述为深负性 T 波和突出/超急性 T 波或互变 T 波的混合物<sup>20,21</sup>。我们的机器学习模型提供了一种更全面的定量方法来评估这种细微的再极化形态的导联间变异。

机器学习非常适合解决 12 导联心电图解读中的许多难题。心肌缺血会扭曲 Q 波、R 峰、R'、QRS 波群、ST 段和 T 波的持续时间和振幅，以及这些波形的形态和构型(如上斜、下斜、凹陷、对称和切迹)。这些畸变具有导联特异性，但又与导联间的动态相关性有关。因此，心电图解读涉及许多复杂的方面和参数，是一个高维度的决策空间问题<sup>1</sup>。很少有经验丰富的临床医生擅长这种模式识别，这也是为什么很多 OMI 患者没有得到及时再灌注的原因；这也是为什么使用简单回归模型的基于规则的商业系统在 OMI 检测中并不理想的原因。机器学习算法可以提供强大的工具来解决 12 导联心电图数据中的高维、非线性数学表征问题。

尽管有关冠心病心电图诊断的机器学习文献比比皆是，但也存在许多严重的局限性。首先，许多研究侧重于检测已知的 STEMI 组，而不是没有 ST 段抬高的临界 OMI 组。其次，之前的大多数研究都使用了开放源代码的心电图数据集，如 PTB 和 PTB-XL，它们都是经过严格筛选的数据集，侧重于心电图判断诊断。我们独特的队列包括未经选择的连续患者，其临床特征和疾病流行率与现实世界中的情况类似。第三，许多研究使用了基于心电图数据和临床数据元素(如患者病史、体格检查异常、实验室值和诊断测试)的各种输入特征，这限制了对真实世界环境的适用性。第四，据我们所知，大多数研究使用单一的衍生队列进行训练和测试，而没有使用独立的验证队列。最后，以前的研究很少关注模型的可解释性，对缺血的新标志物和新途径的揭示少于已有的研究。如果没有具有临床意义的解释辅助工具，用于心电图解读的机器学习模型的临床实用性将非常有限。

这项研究具有重要的临床意义。我们的模型可集成到医疗系统中进行实时部署，临床医生可在采集心电图时随时获得风险评分分配。尽管没有 STEMI 模式，但增强的决策支持可帮助急救人员多识别出 85% 的危重冠状动脉闭塞患者，而且不会降低精确度。我们的模型还能帮助为 50% 以上初步评估不确定的患者提供护理信息，使 OMI 低风险组的患者增加 45%，而 NPV 不会有任何损失。这种规则输入和规则输出准确性的递增有助于将关键的急诊资源重新分配给最需要的人，同时优化临床工作流程。这将影响首次医疗接触时的许多决策，包括有针对性的院前干预、导管室启动、抗缺血疗法的应用、住院目的地决策、会诊需求、快速诊断检测（如心电图和成像扫描）转诊以及早期出院决策。此外，到目前为止，临床医生还没有敏感或高度特异的工具，可以在没有 STEMI 模式的情况下超早识别 OMI。通过加强诊断，可以设计和实施前瞻性干预试验，以评估针对这一弱势群体的干预措施的治疗效果（例如，早期使用上游 P2Y12 抑制剂、紧急再灌注治疗与延迟再灌注治疗以及葡萄糖-胰岛素-钾输注）。

有几个限制因素值得考虑。首先，我们用于建立模型的特征是基于特定制造商的软件。众所周知，不同制造商在心电图预处理方面存在差异，这意味着在使用不同软件进行信号处理时，我们的模型需要重新训练。另外，深度神经网络也可用于分析原始心电信号，而无需明确的特征工程。然而，这些技术需要大量的训练样本（例如，大于 10,000 个），与基于特征工程的机器学习相比，在传统的基于 12 导联心电图的诊断中可能不会产生有意义的改进。其次，我们发现推导组群和验证组群在疾病患病率和临床医生心电图解读准确性方面存在细微差别。这些队列来自美国两个不同的地区，而紧急医疗系统（EMS）遵循各州的特定协议。急救医疗系统协议和院内实践的差异可能导致接受院前 12 导联心电图检查的患者类型和比例以及相应的结果判定略有不同。不过，令人欣慰的是，我们的模型在研究地点之间仍然具有良好的通用性。第三，值得注意的是，我们的“任何 ACS 事件”模型只提高了规则入组的性能。这意味着，低风险判定表明特定患者不太可能发生 OMI，但他们可能有不太明显的 NSTEMI-ACS 表型，不需要再灌注治疗。连续心电图检测可能会提高对漏诊事件的检测率，因为患者可能在随后几小时内转为高风险类别，但这一点仍有待证实。冠状动脉闭塞是一个动态过程，会随着时间的推移而发生变化，因此我们的模型得出的初始低风险等级不应导致降低主动监测水平。最后，虽然本研究使用的是前瞻性患者，但所有分析都是离线完成的。需要进行前瞻性验证，实时提供 OMI 概率和决策支持。

总之，我们在美国多个地点的 7313 名胸痛患者中开发了心电图诊断 OMI 的模型，并进行了外部验证。结果表明，机器学习在以一种独立于观察者的方法检测指示 OMI 的细微缺血性心电图变化方面具有优越性。这些模型的表现优于临床医生和商业心电图解读软件，显著提高了精确度和召回率。对驱动我们模型的心电图特征进行了评估，提供了与心肌损伤的合理机理联系。与 HEAR 评分相比，我们推导出的 OMI 风险评分提高了判入和判出的准确性，当与训练有素的急救人员的临床判断相结合时，该评分有助于对三分之一的胸痛患者进行正确的重新分类。这种新临床路径在临床结果方面的益处应在前瞻性试验中进行评估。未来的工作还应该侧重于实时提供 OMI 概率和决策支持的前瞻性部署。

## 方法

### 道德声明

衍生队列包括来自匹兹堡市紧急医疗服务局的院前数据和来自匹兹堡大学医疗中心（UPMC）医疗系统三家三级医院的院内数据：这三家医院分别是：UPMC Presbyterian 医院、UPMC Shadyside 医院和 UPMC Mercy 医院（宾夕法尼亚州匹兹堡市）。所有符合条件的连续患者都是在放弃知情同意的情况下招募的。该观察性试验已获得匹兹堡大学机构审查委员会的批准，并在 <https://www.clinicaltrials.gov/> 上注册（标识符为 NCT04237688）。本文中描述的分析是由美国国立卫生研究院资助的试验方案预先规定的。第一个外部验证队列包括来自橙县紧急医疗服务机构（北卡罗来纳州教堂山）的数据。这项研究得到了符合条件的患者的积极同意，并获得了北卡罗来纳大学教堂山分校机构审查委员会的批准。第二个外部验证队列包括来自梅克伦堡县急救医疗服务机构和 Atrium Health（北卡罗来纳州夏洛特市）的数据。数据是通过医疗保健登记处收集的，所有连续符合条件的患者都是在放弃知情同意的情况下登记的。本研究还获得了北卡罗来纳大学教堂山分校机构审查委员会的批准。这两个外部数据集由同一地方投资机构收集。

## 研究设计和数据收集

这是一项前瞻性、观察性队列研究。每个研究队列的方法已在其他地方详细描述。所有研究队列均招募了成年患者，这些患者出现了非创伤性胸痛或类似心绞痛的症状（如手臂、肩膀或下巴疼痛、气促、出汗或晕厥），并拨打了急救电话。符合条件的患者由救护车运送，并至少有一次记录的院前12导联心电图。根据性别、种族、合并症或疾病严重程度等进行选择性排除的标准均不存在。对于这项预先指定的分析，我们只包括来自唯一患者接触的非重复心电图，并删除了院前心电图显示室速或室颤的患者（即这些患者由ACLS算法管理）。我们还删除了确认的院前STEMI患者，其中包括机器生成的急性心肌梗死警告、急救队员记录的STEMI和可能需要导管室激活的医疗咨询。

独立审阅员从医院系统中提取符合资格标准的所有患者的数据要素。如果院前心电图没有患者标识符，我们采用概率匹配方法将每次接触与正确的医院记录关联起来。这一先前验证的数据链接协议基于心电图标记的出生日期、性别和日期/时间日志，以及基于急救调度日志和接收医院记录。所有概率匹配均由研究专家手动审核以确保准确性。匹配成功率范围从98.6%到99.8%。

## 临床结果

每个本地站点的独立审阅员在索引接触后的30天内审阅了所有可用的医疗记录，并作出了裁决。审阅员对所有心电图分析和模型预测都是盲审的。OMI被定义为在至少三条主要冠状动脉（左前降支（LAD）、左回旋支（LCX）和右冠状动脉（RCA））或其主要分支中至少一条具有急性致病性病变的冠状动脉造影证据，并且TIMI流动等级为0-1。TIMI流动等级为2且冠状动脉严重狭窄 $> 70\%$ 和峰值肌钙蛋白为 $5-10.0 \text{ ng ml}^{-1}$ 也被认为是OMI的指示<sup>17,21</sup>。这些裁决是由两位独立的审阅员进行的。两位审阅员之间的kappa系数统计为0.771（即，有实质性的一致性）。所有分歧均由第三位审阅员解决。

根据第四版心肌梗死通用定义，ACS被定义为出现缺血症状（即，胸部、上肢、下颌或上腹部持续20分钟以上的弥漫性不适感）并且满足以下至少一项标准：（1）住院期间出现不稳定的缺血性心电图变化（例如，ST段变化和T波倒置）；（2）住院期间心肌肌钙蛋白升高（即， $>$ 第99百分位数）并在连续检测中升高和/或降低；（3）冠状动脉造影显示大于70%的狭窄，无论是否治疗；和/或（4）心功能评估（压力测试）显示心电图、超声心动图或放射核素证据表明局部心肌缺血。对于有心肌梗死的第2型患者和有既往亚急性冠状动脉闭塞的患者，被标记为ACS和OMI为阴性。这包括大约10%的患者，其肌钙蛋白阳性但在连续检测中浓度未升高和/或降低（即，慢性泄漏）或肌钙蛋白泄漏归因于非冠状动脉闭塞性疾病，如心包炎。对于随机选取的少量患者（ $n = 1,209$ ），ACS裁决的kappa系数统计范围从0.846到0.916（即，实质性至完美的一致性）。

## 心电图方法

院前心电图是由急救人员在现场作为常规护理的一部分获取的。心电图是使用Heart Start MRX（飞利浦医疗）或LIFEPAK-15（Physio-Control）监护除颤仪设备获取的。所有数字12导联心电图的采样率为每秒500个样本（0.05-150 Hz），并传输到相应的急救服务机构和接收医院。数字心电图文件以.xml格式导出并存储在每个本地站点的辅助服务器上。心电图图像被去标识化，并由独立审阅员或研究专家手动注释；质量差或缺少导联的心电图被从研究中删除。然后，数字.xml文件被传输到飞利浦高级算法研究中心（马萨诸塞州剑桥市）进行离线分析。

心电图特征提取在其他地方有详细描述。简言之，使用制造商特定软件（飞利浦DXL诊断12/16导联心电图分析程序）进行了心电图信号预处理和特征提取。首先对心电图信号进行预处理以消除噪声、伪迹和基线游走。异位搏动被去除，并为每个导联计算了代表性的中位搏动。中位搏动是指经过R峰的时间对齐后，在给定心电图导联中顺序搏动的代表性平均值（或中位数）。接下来，我们使用均方根（RMS）信号识别全局波形基准点，包括P波、QRS波群和T波的起始、终止和峰值。然后识别了导联特定的基准点，以进一步将单个波形分割为Q、R、R'、S、S'和J点。



然后，我们基于以下内容计算了总共554个心电图特征：（1）全局和导联特定波形的幅度、持续时间、面积、斜率和/或凹凸度；（2）前瞻、水平、空间、x-y、x-z和y-z平面中的QRS和T轴和角度，包括峰值、拐点和初始/末端环的方向；（3）正交心电图导联（I、II和V1-V6）的主成分的特征值，包括各个心电图波形段的PCA比率；和（4）T环形态描述符。删除了分布为零的特征以防止表征偏差。

接下来，我们之前已经确定了与心肌缺血有机联系的最重要的心电图特征的最佳简约列表，具体描述在其他地方<sup>18</sup>。简而言之，为了防止省略特征的偏差，我们使用了一种混合方法，结合了领域知识和数据驱动策略。首先，临床科学家确定了24个已知与心肌缺血相关的经典特征（即导联特异性ST和T波振幅）。接下来，从一个包含554个候选特征的全面列表开始，我们使用数据驱动算法（例如，递归特征消除和LASSO）来识别可能与缺血相关的198个补充特征。LASSO在L1范数正则化后选择具有非零系数的特征，而递归特征消除使用重复的回归迭代来识别对模型预测有显著影响的特征。然后，我们检查了这个扩展列表中的222个特征对，并删除了具有非常高共线性分数的特征，其中包含了冗余信息（例如，如果模型同时选择了QT和QTc，则保留QTc）。最后，我们使用特征重要性排序来确定最简约的特征子集，这些特征互补，并且可以提升分类性能。这种混合方法最终产生了一个包含73个特征子集，可以作为缺血的合理标志物。

## 机器学习方法

我们遵循了“ROBUST-ML”和“ECG-AI压力测试”清单推荐的最佳实践，来设计和评估我们的机器学习算法。为了防止测量偏差，对心电图特征进行了手动审查，以识别错误的计算。生理上合理的异常值被替换为 $\pm 3$ 个标准差。平均每个特征的缺失率为0.34%（范围为0.1-1.6%）。因此，在与临床专家协商后，我们使用该特征的均值、中位数或众数对缺失值进行了插补。然后对心电图指标进行了z-score标准化，并将其作为机器学习模型的输入特征。推导和验证数据集都独立清理，以防止数据泄漏。两个队列在相同的时间窗口内招募，表明没有时间偏差。为了防止与预期使用的潜在不匹配，模型开发的输入特征仅包括心电图数据加上机器生成的年龄。没有其他临床数据用于模型构建。

我们将推导队列随机分成80%的训练集和20%的内部测试集。在训练集上，我们拟合了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。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。训练好的模型使用AUROC曲线进行比较，并使用Wilcoxon符号秩检验进行成对比较。使用Youden指数选择ROC优化的截断点，并使用McNemar检验比较混淆矩阵上的分类。

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。我们生成了这些概率分数的密度图，用于正负类，并基于预先指定的NPV > 0.99和真阳性率 > 0.50选择低风险、中等风险和高风险组的分类阈值。最后，我们使用锁定随机森林分类器在完全未见的外部验证队列上生成概率分数和风险类别。生成概率分数的代码包含在本文的附加材料中。

## 参考标准

为减少评估偏差的风险，我们将我们的机器学习模型与临床实践中常用的多个参考标准进行了基准测试。首先，我们使用商业化的、FDA批准的心电图解释软件（飞利浦DXL诊断算法）来表示缺血性心肌损伤的可能性。该可能性（是/否）基于以下内容的综合：（1）用于“>>急性心肌梗死<<”的诊断代码，包括表示“急性”、“近期”、“年龄不确定”、“可能”或“可能”的描述性语句；以及（2）用于“>>急性缺血<<”的诊断代码，包括表示“可能”、“可能”或“考虑”的描述性语句。将表示“旧的”[梗死]、“非特异性”的[ST段压低]或“由于”的[左心室肥大或心率过快]的诊断语句排除在此综合参考标准之外。

我们还使用临床医生对心电图的复查来表示在不存在STEMI模式的情况下对特定心电图的缺血性心肌损伤的可能性（是/否），这与急诊科医生在临床实践中评估这些患者的方式一致。独立的医生评审员根据第四个心肌梗死的普遍定义标准对每个12导联心电图图像进行标注，包括两个连续导联的ST段抬高（ $\geq 0.2\text{mV}$ ，男性 $\geq 40$ 岁的V2-V3导联， $\geq 2.5\text{mm}$ ，男性 $< 40$ 岁的V2-V3导联；女性 $\geq 0.15\text{mV}$ ，或其他导联 $\geq 0.1\text{mV}$ ）或ST段压低（新的水平或向下倾斜的压低 $\geq 0.05\text{mV}$ ），伴随或不伴随T波倒置（在具有突出的R波或R/S比值 $> 1$ 的导联中 $> 0.1\text{mV}$ ）。评审员还被提示使用他们的临床判断来识别高度可疑的缺血性改变（例如，互补性改变和急性T波改变），并考虑潜在的混杂因素（例如，束支传导阻滞和早期复极化）。在推导队列的随机选定子集（ $n=1,646$ ）中，对解释心电图的两名急诊医生之间的kappa系数统计是0.568（即，中等一致性）。第三位评审员被用于裁定此随机选定子集上的差异。同样，在外部验证队列的随机选定子集（ $n=375$ ）中，对解释心电图的两名董事会认证的心脏病学家之间的kappa系数统计是0.690（即，显著一致性）。

最后，考虑到在缺乏STEMI时临床医生在初步分诊时很大程度上依赖风险评分，这将极大地影响OMI患者在临床实践中的诊断和治疗方式，我们将我们衍生的OMI风险评分与HEART评分进行了比较。这个评分在美国的医院中很常用，并且已经被充分验证用于在急诊科分诊患者60。HEART评分基于患者初次就诊时的历史、心电图解读、年龄、危险因素和初次肌钙蛋白值（范围为0-10）。这个评分将患者分为低风险（0-3）、中等风险（4-6）和高风险（7-10）组。由于通常在第一次医疗接触时通常不会得到肌钙蛋白的结果，我们使用了一个修改后的HEAR评分，删除了肌钙蛋白值，这也是之前验证过的，用于在抵达医院之前由急救人员使用的评分36。这里对HEART评分的比较主要关注于确定使用衍生的OMI评分是否能在初步分诊时提供比常规护理更大的增量。我们比较了由我们衍生的OMI评分分配的新风险类别与HEART评分分配的风险类别之间的一致性 or 差异，这可以提供关于超过常规护理的潜在增益的信息。

## 统计分析

描述性统计数据以平均值 $\pm$ 标准差或  $n$  (%) 报告。缺失数据的随机性进行了评估，并在心电图特征选择过程中进行了处理（见上述“机器学习方法”小节）。在必要时进行了假设检验前分布的正态性评估。心电图特征被  $z$  分数标准化，作为机器学习模型标准输入结构的一部分。使用卡方检验（用于离散变量）和独立样本  $t$  检验或曼-惠特尼  $U$  检验（用于连续变量）进行了队列之间的比较。显著性水平在适用的情况下设置为 $\alpha=0.05$ ，用于双侧假设检验。

所有诊断准确度值均按照诊断准确性研究报告标准（STARD）的建议报告。我们使用 AUROC 曲线、灵敏度（召回率）、特异度、PPV（精确度）和 NPV 报告了分类性能，同时在适用的情况下报告了 95% 的置信区间。对于 10 折交叉验证，我们使用 Wilcoxon 符号秩检验（用于 AUROC 曲线）和 McNemar 检验（用于混淆矩阵）比较多个分类器。我们使用核密度图估计在类别之间导出了最终分类器的低风险、中风险和高风险类别。这些风险类别的适当性是通过在索引入院期间对临床重要结果的累积风险的对数秩卡方进行评估的。

为了评估分类性能的增量增益，我们使用 DeLong 检验将最终模型的 AUROC 与参考标准进行比较。为了便于比较，参考标准（商业系统和临床医生）的 AUROC 的置信区间边界是使用 1,000 个自举样本生成的。为了将增量增益值放置在更广泛的临床工作流程背景下，我们还计算了我们的模型相对于在第一次医疗接触时的初步评估期间使用 HEAR 评分的 NRI 指数。风险评分是对不符合 STEMI 标准的怀疑 ACS 患者的临床工作流程的组成部分。根据 STARD 建议，NRI 指数评估了在正确重新分类旧测试（HEART 评分）的风险类别分配时，使用新测试（衍生的 OMI 评分）之间的净增益。

我们使用 logistic 回归来确定 OMI 风险类别的独立预测价值。我们使用单变量分析中显著的变量，然后使用步进向后选择方法建立多变量模型，使用 Wald 卡方标准进行选择。我们报告了所有显著预测因子的 OR 和 95% 的置信区间。所有分析均使用 Python 版本 3.8.5 和 SPSS 版本 24 完成。



## 数据可用性

---

ECG-SMART 试验利用提取的心电图特征来训练和评估一个 RF 分类器，以表示 OMI 的概率。在衍生和外部验证数据集中使用的心电图特征，以及相关的临床结果，可通过 GitHub (<https://github.com/zeineb-bouزيد/sharing-github-nature-medicine.git>) 公开获取。希望获取源二进制文件以计算自己的特征的研究人员应联系对应作者安排适当的批准和机构数据使用协议。来自非商业实体的有兴趣的研究人员可以通过电子邮件联系对应作者 [ssa33@pitt.edu](mailto:ssa33@pitt.edu) 提交请求。请求将在 2 周的时间内处理。

## 代码可用性

---

评估这些模型的 Python 代码以及推导和外部验证数据集可通过 GitHub 获取 (<https://github.com/zeineb-bouزيد/sharing-github-nature-medicine.git>)。

## 附加信息

---

本文的扩展数据可在 <https://doi.org/10.1038/s41591-023-02396-3> 获取。

补充信息

在线版本包含可在 <https://doi.org/10.1038/s41591-023-02396-3> 获取的补充材料。

信函和材料请求应寄至 Salah S. Al-Zaiti。

同行评审信息

《自然医学》感谢 Chengyu Liu、Antonio Luiz Ribeiro 和 Giulio Guagliumi 对本文的同行评审的贡献。

初审编辑：Lorenzo Righetto，与《自然医学》团队合作。

转载和权限信息可在 [www.nature.com/reprints](http://www.nature.com/reprints) 获取。