

检测急性心肌梗塞（AMI）的深度学习方法

摘要

背景：在日常实践中，急性心肌梗塞（AMI）的诊断延误或误诊并不少见。由于12导联心电图（ECG）对于检测AMI至关重要，因此加强ECG解读的系统算法可能对改善诊断具有重要意义。

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。

方法：这项回顾性队列研究包括了737/287名经冠状动脉造影（CAG）验证的STEMI/NSTEMI患者的1051/697份ECG和来自76775名非AMI患者的140336份ECG。在这些ECG中，DLM分别在80%和20%的数据上进行了训练和验证。进行了人机比赛。接收器操作特征曲线下面积（AUC）、灵敏度和特异性被用来评估DLM的性能。

结果：在人机比赛中，DLM对STEMI的检测的AUC为0.976，显著优于最佳医生的表现。此外，DLM独立展现了足够的诊断能力用于STEMI的检测（AUC=0.997；灵敏度，98.4%；特异性，96.9%）。关于NSTEMI的检测，结合DLM和传统心脏肌钙蛋白I（cTnI）的AUC提高到0.978，这优于DLM（0.877）或cTnI（0.950）的表现。

结论：DLM可能作为一个及时、客观和精确的诊断决策支持工具，以协助基于紧急医疗系统的网络和一线医生检测AMI，并随后启动再灌注疗法。

引言

急性心肌梗塞（AMI）尽管在诊断和管理上取得了全球性进展，但仍然是一个重大的公共卫生问题。**AMI指的是通过异常的心脏生物标志物检测到的急性心肌损伤**，伴随在临床环境中与心肌缺血一致的坏死。基于12导联心电图（ECG）的表现，**ST段抬高型心肌梗塞（STEMI）和非ST段抬高型急性冠状动脉综合征（NSTE-ACS）**的分类通常被包括在**急性冠状动脉综合征（ACS）**的概念中。有心肌缺血症状的患者和ECG上有ST段抬高需要及时再进行再灌注治疗，以减少心脏发病率和死亡率。同样，被认为处于非常高风险或高风险类别的**非ST段抬高型心肌梗塞（NSTEMI）**患者需要立即/早期的侵入性策略来防止更糟的预后。

然而，及时的管理依赖于快速识别和准确诊断。尽管已经建立了诊断AMI的标准，但对急诊医生来说，快速识别仍然是一个严峻的挑战。以往的研究报告称，**首次医疗接触时AMI的误诊率在2%到30%之间**。

未能识别患有AMI患者中的高风险ECG发现会导致较低质量的护理和更高的不良结果。诊断过程中漏诊的主要原因之一是错误解释了诊断测试。因此，改进ECG解读的系统化过程可能对改善诊断具有重要意义。由于主要的AMI诊断工具是12导联ECG，对ECG进行更详细的分析可能会显著加快这一过程。

当前由深度学习模型（DLM）开始的人工智能革命为我们提供了一个前所未有的机会来改进医疗系统，并且已经被证明在医疗应用中是有效的。

此外，当DLM被大量标注的ECG数据集训练时，它们被证实优于心脏病专家在ECG解读方面。据我们所知，可用的适用于AMI的ECG数据库相对较小。我们的研究旨在通过12导联ECG及时、客观和准确地开发一个DLM来检测AMI。通过DLM学习了超过100,000份与AMI相关的ECG。

借助系统的强大计算能力，训练模型的性能与包括心脏病专家和急诊医生在内的医生进行了比较。还**评估了DLM和传统心脏肌钙蛋白I（cTnI）对STEMI和NSTEMI的诊断能力**。

方法

研究设计

这是一项单中心、病例对照研究。数据由台北三军总医院提供，回顾性设计已经得到了机构审查委员会（IRB编号：2-107-05-168）的伦理批准。构建了一个电子健康系统用于收集ECG和医疗记录。研究期间从2012年1月至2018年12月。

研究人群

征集到急诊部（ED）呈现的AMI患者，他们接受了冠状动脉造影（CAG）来确诊I型AMI并确认STEMI的梗塞相关动脉（IRA）。排除了没有电子ECG、右侧ECG、后侧ECG和起搏器节律的AMI患者。

同一时期呈现到ED的非AMI患者也被招募，同时排除了有AMI史或在ED逗留期间任何升高的cTnI的患者。本研究中AMI、STEMI、NSTEMI、非AMI和非STEMI的定义见补充表1和补充附录1。

按日期将AMI病例分为开发（80%）和验证（20%）队列。开发队列中的ECG从验证队列中排除。这两个队列之间没有患者重叠。

最终诊断的裁定

最终诊断的裁定是由三位未参与人机竞赛并根据临床表现、连续ECG、连续cTnI水平和血管造影发现，回顾性并独立审查AMI案例的董事会认证的介入性心脏病专家进行的，以根据当前指南推荐作出STEMI和NSTEMI的最终诊断。如果对诊断有异议，案例将在共识会议中进行审查和裁决。

数据收集与DLM实施：

数据收集和DLM实施如补充附录1和补充图1所示。ECG记录是使用飞利浦12导联ECG机器（PH080A）收集的，DLM基于之前开发的ECG12Net。DLM的输出是STEMI、NSTEMI和非AMI的概率。

人机比赛：

我们通过一个包含450个ECG的比赛集来评估参与医生的性能，其中包括174个STEMI、138个NSTEMI和138个非AMI ECG。根据IRA，STEMI ECG进一步分类为左主干冠状动脉（LMCA）、左前降动脉（LAD）、左回旋动脉（LCx）或右冠状动脉（RCA）。五名心脏病学家和一名急诊主治医师参加了比赛。此外，还包括飞利浦12导联算法在比赛中检测AMI。医生们无法获取任何病人信息，也不知道数据内容。他们的回应被输入到一个在线标准化数据输入程序中。我们计算灵敏度、特异性和卡帕值，以将他们的结果与DLM的结果进行比较。

统计分析：

研究队列被划分为训练、验证和比赛集。我们适当地以平均值和标准差、病人数量或百分比展示他们的特征。使用学生t检验或卡方检验进行比较。统计分析使用R软件版本3.4.4（R Foundation for Statistical Computing, Vienna, Austria）执行。

所有分析都基于ECG而非病人。在整个分析中使用了 $p < 0.05$ 的显著性水平。主要分析是评估人机比赛中DLM、医生和飞利浦算法对STEMI检测的性能。接收器操作特征（ROC）曲线和ROC曲线下的面积（AUC）被用于评估比赛结果。我们还使用精确-召回ROC（PRROC）来评估假设现实世界情况下的模型性能。由于比赛集中STEMI、NSTEMI和非AMI的比例失真，我们根据现实世界中的发生率（分别为STEMI、NSTEMI和非AMI案例的0.1%、0.2%和99.7%）重新加权样本。在整个验证队列上进行了次要分析。我们包括更多临床信息，如患者特征和实验室测试，以提高模型性能。使用多变量逻辑回归模型将DLM和临床信息整合。一系列逻辑回归模型识别了不同临床信息对STEMI和NSTEMI检测性能的影响。AUC被用于评估模型性能的变化。本研究的研究兴趣、模型比较和统计方法在补充表2中详细总结。

结果

队列的基线特征：

本研究包括了737名STEMI患者的1051份CAG前ECG、287名NSTEMI患者的697份CAG前ECG以及76775名非AMI患者的140336份ECG。开发队列和验证队列分别包含来自58056和19743名患者的记录。特征和实验室数据展示在补充表3中，详细描述显示在补充附录2中。

STEMI、NSTEMI和非AMI的预测：

人机比赛的结果总结在图1中。DLM在人机比赛中对STEMI检测的AUC为0.976，相应的灵敏度和特异性分别为89.7%和94.6%。相比之下，医生和飞利浦算法对STEMI检测的灵敏度和特异性范围分别为60.5-92.6%和76.0-97.5%，低于DLM的表现。PRROC分析展示了自动ECG筛选系统的可行性，显示了DLM对STEMI检测的AUC在假设的真实世界中为0.586。DLM在适当的截断点达到了63.2%的精确度和50.3%的召回率。这些值显著优于所有医生和飞利浦算法的结果。

在人机比赛中对STEMI检测进行了DLM、医生和飞利浦算法的性能排名和一致性分析（图2）。DLM在全球性能上表现最好（ $\kappa^{(3)}=0.645$ ）（图2A），而医生在STEMI检测上相对较好，但对NSTEMI和非AMI的区分较差。DLM、医生和飞利浦算法之间AMI检测的一致性分析展示在热图中（图2B）。

STEMI的IRA分析：

对于STEMI的IRA（梗死相关动脉）检测，DLM实现了最佳的全球性能（ $\kappa=0.629$ ）（补充图2）。如补充图3所示，在排除LMCA（左主干冠状动脉）和LCx（左回旋动脉）后，DLM对前壁STEMI检测的AUC为0.975，相应的灵敏度为92.6%，超过了所有参与的医生。此外，DLM在下壁STEMI检测中的AUC为0.974，相应的灵敏度为84.8%，优于除一位最佳医生外的所有人。在前壁和下壁STEMI的联合检测中，DLM的性能优于所有医生（AUC，0.975；灵敏度，89.4%）。

DLM与医生对STEMI ECG的解读：

在人机比赛中选出的STEMI ECG示例显示在图3中。一个典型的LAD梗死相关的STEMI ECG（图3A）被DLM和医生一致检测到。一个RCA（右冠状动脉）梗死相关的STEMI ECG（图3B）被DLM误诊但被最佳心脏病医生正确识别。一个RCA梗死相关的STEMI ECG（图3C）被DLM和最佳心脏病医生都误诊。DLM正确检测到ECG（图3D）为LAD梗死相关的STEMI，这被最佳心脏病医生误诊。

在138个NSTEMI ECG中，58个ECG被DLM检测为非AMI，准确率为58.0%，低于最佳心脏病医生（75.4%）。这一差异是由于DLM更为保守的AMI诊断策略。相反，在138个非AMI ECG中，DLM的特异性为96.4%，远高于两位最佳心脏病医生（82.6%和64.5%）。在调整特异性后，DLM的NSTEMI误诊明显少于最佳心脏病医生（表1）。尽管如此，在最佳心脏病医生的标准化下，DLM在AMI检测中提供了最佳性能。针对STEMI检测及相应IRA的ECG导联特定分析展示在补充图4中，详细描述显示在补充附录2中。

STEMI和NSTEMI的逻辑回归分析：

在开发队列中进行的单变量和多变量逻辑回归分析显示，男性、既往冠状动脉疾病（CAD）、心肌肌钙蛋白I（cTnI）、血红蛋白、总胆固醇和低密度脂蛋白（LDL）水平是STEMI和NSTEMI检测的独立危险因素（补充图5）。

诊断价值分析：

我们评估了在调整了显著的患者特征、疾病历史和实验室数据后的DLM性能，以确保在验证队列中的广泛假定的混杂变量上保持一致性。DLM在检测STEMI方面的性能显著优于cTnI，具有0.997的AUC，相应的灵敏度和特异性分别为98.4%和96.9%（图4A）。然而，在检测NSTEMI方面，cTnI的性能显著优于DLM。通过DLM和首次记录的cTnI的组合检测NSTEMI的AUC提高到0.978，相应的灵敏度和特异性分别为91.6%和96.7%，这优于DLM（0.877）或cTnI（0.950）单独使用的情况（图4B）。独立使用DLM足以检测STEMI，并且增加患者特征并未显著提高其性能。然而，发现cTnI比任何额外的特征更能提高NSTEMI的诊断准确性（补充图6）。

讨论：

在这项研究中，我们建立了一个深度学习模型（DLM）通过ECG分析精确检测STEMI，该模型应用了深度卷积网络从超过100,000个ECG中提取显著特征。所有AMI患者都通过CAG进行了验证，且识别了STEMI的相应IRA。最重要的是，我们的DLM在STEMI检测方面的表现超过了医生，具有89.7%的高灵敏度和94.6%的高特异性。

近年来，在心血管领域应用深度学习技术来诊断心律失常、电解质异常和瓣膜心脏病日益受到欢迎。然而，还没有为AMI检测设计大规模研究。之前针对12导联ECG的AMI检测的DLM主要使用了Physikalisch-Technische Bundesanstalt (PTB) 诊断ECG数据库。

这些研究可能受到限制，因为它们没有进一步验证。此外，缺少DLM与医生之间的比较。与以往研究相比，我们招募了最大数量的临床验证ECG进行开发和验证。此外，我们进一步确认了cTnI在协助NSTEMI检测方面的作用。所有这些结果都凸显了当前研究的优势。

DLM在STEMI检测方面的灵敏度和特异性都优于医生。ECG是所有客观检测AMI方法中最及时的工具。然而，医生间对ECG的解读存在灵敏度低和意见不一致的问题。利用12导联ECG进行AMI检测的主观解读的灵敏度只有61%到74%，特异性介于72%到89%之间。相比之下，以往的院前计算机算法解读对STEMI的灵敏度约为69%。我们的DLM提供了杰出的性能，可以支持临床实践中的决策系统。

DLM可以客观地基于分析和学习大量ECG来识别STEMI。此外，医生容易错过的STEMI最早期阶段的细微ECG变化可以被DLM正确识别。

尽管如此，以往的心肌梗死或心肌病可能会因基线ST-T改变而误导DLM。因此，可能需要以前可用的ECG信息或心血管疾病历史来加强DLM对STEMI检测的能力。

我们进一步评估了DLM在检测STEMI等效物和STEMI模仿物方面的性能。STEMI等效物包括de Winter征象、Wellens'综合症、超急性T波、在aVR导联ST抬高伴广泛ST降低、存在束支阻滞时的ST抬高和后壁AMI，代表冠状动脉闭塞而不符合传统ST抬高标准，对及时识别至关重要。

此外，如高钾血症、良性早复极、左室肥大和布鲁加达综合症等模仿STEMI的高起点T表现，通常被误诊，导致错误启动初级PCI。我们的研究证明了DLM在检测STEMI等效物（除了第一型Wellens'综合症外）和高起点T方面表现出卓越的诊断力（

补充图7，补充图8）。

未来需要进一步的前瞻性和大型ECG验证数据集来确认DLM的区分能力。

我们的DLM有几个潜在的临床应用。首先，DLM可以提供决策支持和AMI的高风险警报系统，帮助减少由于紧张的时间压力、繁重的工作负担和匆忙的工作人员而导致的ED医疗错误。其次，DLM可以集成到救护车中的ECG机器中，以便于远程医疗并缩短启动再灌注疗法前的决策时间。第三，我们的DLM可以应用于缺乏专家的农村和偏远地区，以促进ECG解读和提高诊断准确性，从而及时进行管理并改善STEMI患者的预后。最后，DLM可以集成到可穿戴设备中用于AMI检测，特别是对于极高风险的动脉粥样硬化性心血管疾病患者。因此，我们的DLM显示了诊断益处，并可能在不久的将来改善医疗保健质量（中央插图）。

局限性

这项研究的一些局限应当被指出。首先，人机比赛基于一个精心设计的回顾性研究。应进行实际的前瞻性研究来验证DLM的临床影响。其次，只有六位主治医师参加了与DLM的比赛。

虽然他们在AMI检测中的表现与以前的研究相比较一致，但应与更多医师比较以确认DLM的优越性。

第三，研究的患者仅从一个学术医学中心招募，尽管AMI的诊断和管理是基于指南的。需要多中心验证来确认这项研究的价值和应用。第四，NSTEMI病例比STEMI病例少，这可能限制了我们的DLM检测NSTEMI的能力。第五，在研究期间，用于AMI诊断的是cTnI而不是hsTnI。

第六，我们的DLM系统中没有先前ECG的信息来提高诊断性能。第七，没有分析冠状动脉侧支流对AMI期间ST-T改变的影响以及DLM在检测STEMI中的表现。最后，只有急诊室的患者被招募，这可能导致选择偏差并限制了结果的普遍性。

结论

我们建立了一个最优的DLM来基于12导联ECG检测STEMI，其准确性优于医师。

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。需要进一步的前瞻性验证，包括院前和院内ECG测试，以确认我们DLM的性能。

对日常实践的影响

现在可以使用这种心脏病专家级别的算法识别STEMI，实现实时STEMI诊断和早期警报。

已建立了一个包括住院、院前和可穿戴设备的全面生态系统，提高了AMI护理的质量。

资助

本项工作由台湾科学技术部（MOST 108-2314-B-016-001给C. Lin，MOST 109-2314-B-016-026给C. Lin），台湾国家科技发展基金管理协会（MOST 108-3111-Y-016-009 和 MOST 109-3111Y-016-002给C. Lin），以及台湾成功总医院（CHNDMC-109-19给C. Lin）支持。

利益冲突声明

作者声明没有利益冲突。

补充数据

补充数据包含了补充性材料，为读者提供了关于研究方法、结果和深度学习模型（DLM）的额外信息。这些补充数据通常用于支持主文的发现，提供更深入的理解，并展示额外的分析结果。

补充附录1：包括AMI、STEMI、NSTEMI、非AMI和非STEMI的定义；数据收集；到达-球囊/冠状动脉造影（CAG）、到达-ECG和ECG-球囊/CAG的时间线；深度学习模型（DLM）的实施；DLM的实施细节；训练细节；数据增强；模型可视化以及研究兴趣、模型比较和统计方法的总结。

补充附录2：包括队列的基线特征、ECG导联特定分析和讨论。

补充图1：DLM的架构。

补充图2：在人机比赛中DLM、医生和飞利浦算法的STEMI梗死相关动脉检测的性能排名。

补充图3：在人机比赛中前壁（LAD）、下壁（RCA）以及联合前壁和下壁（LAD+RCA）STEMI检测的性能比较。

补充图4：对STEMI、STEMI-LAD和STEMI-RCA的ECG导联特定分析。

- 补充图5：** 在开发队列中STEMI和NSTEMI的单变量和多变量逻辑回归分析。
- 补充图6：** 在验证队列中额外的人口统计变量、cTnI和DLM的诊断价值比较。
- 补充图7：** DLM检测STEMI等效物的测试示例。
- 补充图8：** DLM检测高起点T ECG的测试示例。
- 补充表1：** AMI、STEMI、NSTEMI、非AMI和非STEMI的定义。
- 补充表2：** 研究兴趣、模型比较和统计方法。
- 补充表3：** 在开发和验证队列中STEMI、NSTEMI和非AMI ECGs的相应患者特征和实验室结果。