# FedScore：联邦评分系统的隐私保护框架

## 开发

Siqi Li[1], Yilin Ning[1], Marcus Eng Hock Ong[2,3,4], Bibhas Chakraborty[1,2,5,6], Chuan Hong[6], Feng Xie[1,2], Han Yuan[1], Mingxuan Liu[1], Daniel M. Buckland[7], Yong Chen[8], Nan Liu[1,2,9*1] 定量医学中心，杜克-新加坡国立大学医学院，新加坡，新加坡[2] 卫生服务与系统研究项目，杜克-新加坡国立大学医学院，新加坡，新加坡

[3] 卫生服务研究中心，新加坡保健服务，新加坡，新加坡[4] 急诊医学系，新加坡中央医院，新加坡，新加坡[5] 统计与数据科学系，新加坡国立大学，新加坡，新加坡

[6] 生物统计学与生物信息学系，杜克大学，美国北卡罗来纳州达勒姆[7] 急诊医学系，杜克大学医学院，美国北卡罗来纳州达勒姆[8] 生物统计学、流行病学与信息学系，宾夕法尼亚大学，美国宾夕法尼亚州费城

[9] 新加坡国立大学数据科学研究所，新加坡，新加坡

* 通讯作者：Nan Liu，杜克-新加坡国立大学医学院定量医学中心，College Road 8 号，新加坡 169857，新加坡。电话：+65 6601 6503。电子邮箱：liu.nan@duke-nus.edu.sg

关键词：临床决策、分布式算法、电子健康记录、联邦学习、评分系统

字数：3399

## 摘要

### 目标

我们提出 FedScore，这是一种隐私保护联邦学习框架，可在多个站点生成评分系统，促进跨机构合作。

### 材料与方法

FedScore框架包含五个模块：联邦变量排序、联邦变量转换、联邦评分推导、联邦模型选择和联邦模型评估。为说明其用法并评估FedScore的性能，我们使用新加坡一家三级医院划分出的10个模拟站点，构建了一个假设的全球评分系统，用于预测患者就诊急诊科后30天内的死亡率。我们采用现有的评分生成器，在每个站点分别独立构建10个本地评分系统，并使用集中式数据构建了一个评分系统作为对照。

### 结果

我们使用受试者工作特征（ROC）分析，将所获得的FedScore模型的性能与其他评分模型进行比较。FedScore模型在所有研究点的平均曲线下面积（AUC）值为0.763，标准差（SD）为0.020。我们还计算了各本地模型的平均AUC值和标准差。FedScore模型表现出良好的准确性和稳定性，其平均AUC值较高，最接近合并模型的AUC值，且其标准差低于大多数本地模型。

### 结论

本研究表明，FedScore 是一种保护隐私的评分系统生成器，具有潜在的良好泛化能力。

## 1. 引言

近年来，跨机构合作作为加速医学研究和促进质量改进的一种方式，日益受到重视[1]。医疗保健行业广泛推进数字化，使得可利用数据驱动证据构建临床预测模型[2],；理想情况下，这些模型可使用尽可能多来源汇集而成的集中式数据来构建。跨区域合作的一些例子包括：Collaborative European NeruroTrauma Effectiveness Research in Traumatic Brain Injury[3], 、基因型到表型数据库[4],、心血管疾病大数据[5], 、安大略省院前高级生命支持[6],、凯撒医疗研究库[7] 以及泛亚复苏结局研究[8]。然而，此类合作关系需要数据共享，而数据共享通常费时费力，并且有时甚至会因各种隐私法规[9], [10], （例如欧盟《通用数据保护条例》[11]）而无法实现。

联邦学习（FL），有时也称为分布式学习或分布式算法，无需交换患者级数据即可通过协同训练算法来避免数据共享[12], ；通过将模型训练分配给数据所有者并汇总其结果来保护患者隐私[13]。除了打破数据孤岛，联邦学习还有望加速开发亟需的人工智能模型[14]。例如，在COVID-19疫情期间，Dayan等人[14] 利用联邦学习，在20家研究机构之间构建了临床结局预测模型。Luo等人[15] 使用无损单轮联邦学习算法[15]研究了与COVID-19患者住院时长相关的人口统计学和临床因素。联邦学习在医学图像数据方面已有许多应用，其中大多数使用计算机视觉领域的黑箱模型。

相较之下，尽管可解释模型在临床研究中广受欢迎，但其联邦学习应用案例较少。

作为一种可解释风险评分模型[16], ，评分系统几乎已应用于医学的所有诊断领域[17], ，因为无需使用计算机，它们就能快速、简便地评估多种严重医学疾病的风险[16]。一些传统评分系统，例如格拉斯哥昏迷量表[18] ，最早于1974年提出，很大程度上依赖临床医生的领域专业知识。近年来，出现了更多数据驱动的评分系统构建方法，包括能更好地处理稀疏性的超稀疏线性整数模型[19] 、计算效率更高的近似方法[20], [21],，以及支持灵活运用领域专业知识的交互界面，例如区间编码评分[22] 和AutoScore[23]。

无论采用何种开发策略，评分系统通常都是基于单一数据源构建的；如果开发数据的样本量不足或缺乏代表性，就会限制其在其他站点的应用。虽然可以基于汇总数据开发评分系统[24], ，但如前所述，由于隐私限制，汇总数据的过程耗时且难以实现。因此，需要能够以联邦方式构建评分系统的框架，以克服这些困难。为弥补这一空白，我们提出了FedScore，这是首个用于在多个站点间构建联邦评分系统的框架，并通过一项使用真实世界数据的概念验证实验，展示了该框架的有效性和潜在泛化能力。

## 2. 方法

评分系统是线性分类模型，用户只需进行简单的加、减、乘运算即可作出预测[16], ；由于具有可解释性和透明度，这类系统已广泛用于临床决策[25]–[27] 中的风险分层。它们还有助于纠正医生对医疗结局概率的错误估计，而这种情况可能相当常见[28]。临床医生在实施临床模型时，通常会考虑模型的简约性[29], ；这意味着，模型只有同时具备稀疏性（即尽可能少使用变量）和良好的预测准确度[29]，才称得上简约。例如，AutoScore框架[23] 是一种计算工具，可利用机器学习方法便捷地创建此类评分，并受到临床医生的广泛认可[30], [31], ，因为它融合了领域知识与数据驱动证据。然而，无论评分系统的具体生成方式如何，也无论是否考虑模型可解释性，AutoScore及其他类似方法都只能利用一组汇总数据来开发评分系统。为充分利用不断增长的数据源并构建偏倚较小的模型，我们提出 FedScore框架，以便在联邦数据上实现良好的简约性和可解释性，同时满足潜在的隐私限制。

### 2.1 FedScore框架

FedScore框架由五个模块组成：1）联邦变量排序；2）联邦变量转换；3）联邦评分推导；4）联邦模型选择；5）联邦模型评估。FedScore的工作流程如图1所示。

1）联邦变量排序

为兼顾简约性，变量选择是开发评分系统时必不可少的一步。要构建适用于多个站点的全局模型，需要预先确定一组统一的候选变量。我们采用随机森林进行变量重要性测量，该方法已得到广泛应用[32]–[36]。在FedScore中，首先在每个本地站点分别通过随机森林进行变量排序，然后根据变量在所有$K$ 站点中的加权排名重新排列变量，从而生成全局变量排序。具体而言，对于单个变量$X_m$ ，其中$1 \le m \le P$ ，且$P$ 为预测变量总数，设$q_j \in N$ 表示该变量在站点$j$的排名，则将各站点的$\sum_{j=1}^{K} w_j q_j$ 映射至整数集合$[1, P] \subset Z$，即可得到其全局排名。此处，$w_j$ 是站点$j$的归一化权重，满足$\sum_{j=1}^{K} w_j = 1$；默认设置为$1/K$，表示所有站点的权重相等。也可采用基于样本量的权重$w_i = S_j/S_0$ ，其中$S_j$ 为站点$j$ 的样本量，$S_0$ 为总样本量。若希望适应特定的研究目标，用户也可以采用其他自定义权重。

2）联邦变量转换

创建分类变量可以实现非线性效应建模[16], [23],，这一做法已广泛应用于[37]–[43] 临床评分系统的开发中。遵循这一常见做法，FedScore 在建立统一变量排名后，将连续变量转换为分类变量。此类转换的最大类别数预先设定（例如，通常设为 5）；如果类别数超过上限，则合并类别以满足要求。在本研究中，连续变量的分位数设为 0%、 $k_1$%、$k_2$%、 $k_3$%、 $k_4$% 和 100%，其中

$k_1$、 $k_2$、 $k_3$、 $k_4$ 的默认值分别为 5、20、80 和 95。每个连续变量的统一截断值通过对各站点获得的$k$ 值进行加权计算，其方式与确定全局排名相同。

3）联邦评分推导

二元结局在临床决策中很常见，而逻辑回归是用于对此类结局建模的一种重要方法。联邦逻辑回归可通过不同的现有联邦学习框架实现。为进行演示，我们采用了一种称为 ODAL2[44]，的单轮隐私保护分布式算法，该算法通信高效，并已被证明具有较低偏差和较高统计效率[44]。该算法利用主要本地站点的信息（可访问数据），以及远程站点（无法访问数据）的似然函数一阶和二阶梯度，来构建全局似然函数的近似值。随后，可通过优化近似的全局似然函数获得全局逻辑回归系数。令$x_1$、 $x_2$、 ..... $x_{p-1}$ 表示$p-1$ 预测变量，$y$ 表示一个二元结局，则逻辑回归模型可表示为

$$logit\big(Pr(y=1|x)\big) = x^T\beta$$

其中， $x=(1, x_1, x_2, \ldots x_{p-1})^T$, β 是截距和斜率系数的向量，且$logit(t) = log\ t/(1-t)$。假设总计有$N=\sum_{j=1}^{K} n_j$ 个相同且独立分布（i.i.d.）的观测值分布在$K$ 个站点，则汇总所有站点的数据得到的全局逻辑回归的对数似然函数（LLR）为

$$L(\beta) = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{n_j}\left[Y_{ij}x_{ij}^{T}\beta - log\{1 + exp(x_{ij}^{T}\beta)\}\right]$$

合并估计量$\hat{\beta}$ 可以通过优化$L(\beta)$获得。当数据无法共享且无法使用合并似然函数时，仍可对似然函数进行近似。ODAL2 算法采用泰勒展开的思路，提出使用 LLR 的一阶和二阶梯度进行近似[44]:

$$\tilde{L}^{2}(\beta) = L_1(\beta) + \left\{\nabla \mathrm{L}(\bar{\beta}) - \nabla \mathrm{L}_1(\bar{\beta})\right\}^{T}\beta + \frac{1}{2}(\beta - \bar{\beta})^{T}\{\nabla^{2}\mathrm{L}(\bar{\beta}) - \nabla^{2}\mathrm{L}_1(\bar{\beta})\}(\beta - \bar{\beta})$$

这里，$\bar{\beta}$ 是本地站点执行回归模型后获得的初始值，并存储起来以便广播给远程站点。$L_j(\beta) = \frac{1}{n_j}\sum \quad \left[Y_{ij}x_{ij}^{T}\beta - log\{\right.^{n_j}_{i=1} \qquad 1 + exp\ \ x_{ij}^{T}\beta \quad \left.\right]()\}$是第$j$个站点的对数似然比（LLR）（假定$j = 1$ 为本地站点）。$\nabla L(\bar{\beta}) = \sum \quad n_j\nabla L_j(^{K}_{j=1} \qquad \bar{\beta} \quad N\ )$/是对数似然比$L(\bar{\beta})$在$\bar{\beta}$ 处计算得到的第一个梯度，而$\nabla L_j(\bar{\beta}) = \frac{1}{n_j}\sum\{^{n_j}_{i=1}\ Y_{ij} - p_{ij}(\bar{\beta})\}x_{ij}$ 是站点$j$的对数似然比的第一个梯度，其中$p_{ij}(\bar{\beta}) = 1 + exp(-x_{ij}^{T}\bar{\beta})^{-1}$和$\nabla^{2}L_j(\bar{\beta}) = \frac{1}{n_j}\sum \quad p_{ij}($ $^{n_j}_{i=1} \qquad \bar{\beta} \quad 1 - p_{ij})\{\ (\bar{\beta})\}x_{ij}x_{ij}^{T}$ 是站点$j$的对数似然比的第二个梯度。这两个梯度均在各远程站点计算，并传回本地站点。

最后，通过优化替代似然函数得到 β 的全局 beta 估计量。构建全局模型的这一过程是一次性的[44]，如图1所示；两个共享文件均不包含任何患者级信息，从而保障隐私。联邦评分是通过将全局逻辑回归模型中的系数

四舍五入为整数并映射到区间 $[0, S_{max}]$，其中 $S_{max}$ 是用户预先确定的最高分，例如 100。

4）联邦模型选择

模型选择通过验证数据生成的简约性图来完成。一般的模型选择标准可定义为最大化 $\Psi_{\mathrm{m}} = \sum w_i \phi_i(p_1, p_2, p_3, \ldots p_m)$，其中$w_i$ 是先前所述站点 $i$ 的权重，$\phi_i$ 衡量评分在第$i$个验证集上的表现（例如 AUC值），而 $m$ 是预先指定的要纳入的变量总数，该数量应在所有站点间保持一致。优化任务可以添加不同的约束。例如，变量总数 $m$ 不得超过整数 $D$。变量集合 $\{p_1, p_2, \ldots p_m\}$ 也可以设定为满足用户要求的某些主观标准。例如，用户可能会（根据领域知识）决定一组变量 $\{x_1, x_2, . . ^{x_q}\}$，其中必须纳入 $q \leq m$，无论变量重要性分析的结果如何，都要将其纳入最终评分系统。此外，只要将变量数量从 $d$ 增加到 $m$ 对 $\Psi$ 的变化影响很小，就可以使用少于 $m$个变量的数量 $d$ 来最大化 $\Psi$： $|\Psi_m - \Psi_d| \leq \epsilon$，其中 $\epsilon$ 的大小可由用户根据简约性图直观地确定。

根据选定模型确定最终变量后，通过模块 2) 重新拟合一个新模型，以使最终模型尽可能简约。

5）联邦模型评估

在参与 FedScore框架的每个站点上验证最终模型的性能。根据步骤 4) 中$\Psi_m$ 的定义，联邦评分的整体加权性能为$M_1 = \sum w_j \mu_j(p_1, p_2, p_3, \ldots\ p_m)$，其中 $\mu_j$ 是该评分在第 $j$个测试集上的性能；而$M_2 = (\sum w_j(M_1 - \mu_j)^2)^{1/2}$是衡量站点间加权性能变异的指标。$M_1$ 值越高、$M_2$ 值越低，表示评分的性能和泛化能力越好。

FedScore框架使用 R 4.0.3 实现，代码可在
https://github.com/nliulab/FedScore 获取。

2.2 实验

初始研究队列通过筛选 2016 年和 2017年新加坡中央医院（SGH）的急诊科（ED）就诊记录构建，所用电子健康记录数据提取自新加坡保健服务电子健康智能系统。电子健康记录数据的收集和回顾性分析获准豁免知情同意，且该研究已获得新加坡保健服务的集中式机构审查委员会批准，所有数据均已去标识化。排除 18 岁以下患者及存在缺失值的患者后，将剩余队列随机划分为 10个站点，用于演示，所占比例依次为 4%、5%、7%、9%、10%、11%、12%、13%、14% 和 15%。图2展示了队列构建流程。

本研究的结局是患者在急诊科收治后30天内是否死亡。候选变量根据一项近期研究确定[30], 该研究的队列同样来自SGH急诊科数据。候选预测变量共包括5类29个变量：(1) 人口统计学信息：年龄、性别和种族；(2) 患者病情严重程度分类量表[45] 分诊类别（P1、P2、P3和P4）、轮班时间（上午8点至下午4点、下午4点至午夜、午夜至上午8点）以及星期几

（星期五、星期一、周末、周中）；(3) 生命体征：脉搏（次/分钟）、呼吸（次/分钟）、外周毛细血管血氧饱和度（$SpO_2$；%）、舒张压（mm Hg）和收缩压（mm Hg）；(4) 合并症：心肌梗死、充血性心力衰竭、外周血管疾病、中风、痴呆、慢性肺病、类风湿性疾病、消化性溃疡病、糖尿病、偏瘫或截瘫、肾脏疾病和肝病；(5) 既往医疗服务使用情况：过去一年内的急诊就诊、外科手术、重症监护病房收治和高依赖病房收治。

我们使用 AutoScore 创建基线模型，以便与 FedScore框架进行本地和合并比较。我们进行了三组分析：1）在每个站点上独立训练的 10 个本地评分；2）使用所有站点的数据、但不进行数据共享而训练的一个联邦评分；3）使用集中式数据生成的一个合并评分，这是理想情况，但在大多数现实场景中通常无法实现。所有模型均根据相应的简约性图进行选择，预先设定的标准为：模型中的变量最多不超过 8 个，并持续添加变量，直至曲线下面积不再显著提升。为便于直接比较，评分系统开发过程中使用的截断值和权重均采用第 2.1 节规定的默认值；所有相关流程均由数据驱动，未进行融入临床实践专家知识的调整。

## 3. 结果

共将80,613例急诊入院事件随机划分至10个站点，每个站点的样本量为3,224至12,092例，各站点的训练集、验证集和测试集

分别按70%、10%和20%的比例随机划分，如图2所示。表1总结了总体队列及各站点队列的基线特征。

我们比较了 FedScore 开发的联邦评分、使用全部数据开发的合并评分，以及在每个站点独立开发的 10 个本地评分的性能。图3通过 12 个子图展示了各评分在每个站点测试数据集上的表现。每个子图中，横线表示某个评分模型在 10 个站点上的表现，标注了相应的AUC值及其 95% 置信区间。每个子图中灰色矩形框的垂直边缘表示 10 个 AUC值的均值加减其标准差，因此，灰色矩形框的宽度反映了模型在所有站点间的性能变异程度。详细的 AUC值、置信区间和标准差见补充表1。各模型的评分表及相应的简约性图也见补充材料中的电子表格 2 和图1。

根据图3和电子表格1所呈现的信息，我们总结出以下主要观察结果：1）联邦评分表现良好，所有站点的平均AUC值为0.763，优于本地模型，且接近合并模型；2）联邦评分的曲线下面积方差处于最低之列，尽管站点 2 和站点 5 的本地模型标准差似乎略小，但其平均AUC值较低；3）联邦评分在某些站点上的表现优于该站点本地开发的模型（例如，站点7的0.7804 > 0.7300）。

## 4. 讨论

FedScore是首批旨在多个站点之间实现统一评分泛化、同时保护隐私的框架之一。其可扩展且适应性强的架构为提升孤立临床数据集上的模型泛化能力和稳定性提供了潜在解决方案。

尽管评分系统已广泛应用于临床领域，但现有联邦学习应用中很少有关注评分系统的，这与其普遍应用的现状形成反差，反映出一种现象：现有生物医学联邦学习应用倾向于采用黑箱模型[46]，而非更具可解释性的机器学习模型。为了满足医生对模型简洁性和透明度的期望，与采用计算机科学界已有成熟联邦学习框架的黑箱模型实现相比，可解释模型的联邦学习应用需要进行更多定制和修改。例如，在急诊科进行风险分层时，临床医生可能更倾向于使用简单明了、AUC值较低的评分表，而非AUC值较高的黑箱模型。因此，可解释模型的联邦学习应用需要更加审慎的设计；FedScore通过强调模型简约性并使用户能够灵活地进行过程监控来应对这一问题。如果研究问题更倾向于透明的解决方案，而非仅仅关注模型性能，未来的临床科学联邦学习应用也应考虑类似因素。

生物医学领域的FL研究与计算机科学领域的研究虽然起源相似，但二者存在差异。在许多标准的工程领域的联邦学习场景中，由于单个客户端无法独立创建模型，因此人们一直关注数据划分方案等技术细节以及

各种隐私机制[47]。然而，在临床领域，数据通常以医院或机构为单位形成，因此在这些情况下可以构建本地模型。在此背景下，泛化能力（模型将其性能泛化至新环境的能力[48]）以及全局模型相对于本地模型的稳定性变得更加重要，但许多正在开发的现有联邦学习框架并未充分考虑这些因素。第3节的结果表明，通过联邦学习进行协同训练过程，全局模型预测框架（如FedScore）在保持良好性能的同时，其表现变化可以小于本地开发的模型。对于寻求可靠高风险决策的医学研究而言，联邦学习的这一优势很有前景。

数据限制，例如偏倚数据和小型数据集，被认为是机器学习滥用的一个来源[49], ，然而，调查此类不当行为并不像开发模型那样可行。尽管人们强调[48] 外部验证，但据报道，完成此类验证的临床预测研究不足10%[50]。与其在单一站点上训练模型，随后在其他站点对其进行测试和修改，不如通过隐私保护方法，利用充分且具有代表性的数据构建模型，这可能是更可行的解决方案。FedScore及其未来的扩展有望帮助减少不同队列之间的模型不一致性，从而为医学研究带来更值得信赖的决策。

尽管我们仅使用一个二元结局示例进行说明，但FedScore框架具有可扩展性和多功能性，因为各模块可以进行适当调整，以适应不同的临床研究问题。例如，评分推导模块可以调整为适用于生存结局或有序结局，还可以增加其他保护隐私的

联邦学习框架和拓扑结构，以提供更多选择。我们预计，FedScore及其未来扩展版本能够共同为创建更可信的临床评分系统奠定基础，同时采用保护数据隐私的方法。

局限性与未来工作

结果来自单一来源拆分出的同质数据，未体现不同站点之间真实世界的异质性。由于模块3中当前使用的 ODAL2 算法要求不同站点的数据同质，这与目前使用的大多数 FL 和分布式方法类似，因此 FedScore 在处理异质医疗数据时可能会遇到问题[12], [51]。不过，所提出的框架具有可扩展性，可以通过能够更好地解决这一问题的前沿方案持续更新和增强。未来的工作将通过国际合作，利用更多异质数据集开发 FedScore 流程。我们计划引入两种能够处理站点间异质性的先进 FL 算法，以扩展 FedScore。第一种策略是采用受新型成对条件逻辑回归启发的 dCLR算法[52], ，先估计共同的回归系数，再由各站点在本地估计站点特异性截距。第二种策略是采用无损的小样本 dPQL算法[53], ；该算法已用于对不同医院的表现进行排名，同时考虑各站点的病例组合情况（即，不同医院收治的患者各不相同）。

## 5. 结论

我们提出了 FedScore，这是一种隐私保护评分系统，并以 30天死亡率预测任务进行了概念验证。我们展示了它在构建有效的

联邦临床评分方面的潜力：这类评分具有更强的泛化能力，且在不同站点间的性能差异更小。FedScore 是首创的基于分布式算法构建评分系统的框架，填补了当前医学研究中的一项空白。尽管目前已展示其在二元结局中的应用，但随着联邦学习和临床预测方法的进一步发展，FedScore 还可扩展至其他类型临床结局，以及站点间异质性更大的情形，从而应用于广泛的不同医疗情境。

表1 研究队列描述[a] （N = 80613）

| | 所有研究点 | 站点 1 | 站点 2 | 站点3 | 站点4 | 研究点5 | 站点6 | 站点7 | 站点 8 | 研究点9 | 研究点10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # 事件 | 80,613 | 3,224 | 4,031 | 5,643 | 7,255 | 8,061 | 8,867 | 9,674 | 10,480 | 11,286 | 12,092 |
| 年龄，均值 (SD) | 63.51 (17.74) | 64.09 (17.35) | 63.54 (17.77) | 63.32 (17.78) | 63.80 (17.70) | 63.38 (17.70) | 63.24 (18.08) | 63.52 (17.72) | 63.59 (17.76) | 63.29 (17.67) | 63.65 (17.68) | |
| 性别 | | | | | | | | | | | |
| 女性 | 40492（50.2） | 1599（49.6） | 2015 (50.0) | 2830（50.2） | 3604（49.7） | 4034（50.0） | 4528（51.1） | 4899（50.6） | 5258（50.2） | 5705 (50.5) | 6020 (49.8) |
| Male | 40121（49.8） | 1625（50.4） | 2016 (50.0) | 2813 (49.8) | 3651 (50.3) | 4027 (50.0) | 4339 (48.9) | 4775 (49.4) | 5222 (49.8) | 5581 (49.5) | 6072 (50.2) |
| 种族/族裔 | | | | | | | | | | | |
| 华人 | 56966 (70.7) | 2289 (71.0) | 2847 (70.6) | 3956 (70.1) | 5125 (70.6) | 5763 (71.5) | 6273 (70.7) | 6769 (70.0) | 7358 (70.2) | 7941 (70.4) | 8645 (71.5) |
| 印度人 | 8888 (11.0) | 320 (9.9) | 452 (11.2) | 610 (10.8) | 811 (11.2) | 819 (10.2) | 1026 (11.6) | 1120 (11.6) | 1171 (11.2) | 1266 (11.2) | 1293 (10.7) |
| 马来语 | 9793 (12.1) | 386 (12.0) | 478 (11.9) | 711 (12.6) | 865 (11.9) | 1006 (12.5) | 1026 (11.6) | 1203 (12.4) | 1308 (12.5) | 1368 (12.1) | 1442 (11.9) |
| 其他 | 4966 (6.2) | 229 (7.1) | 254 (6.3) | 366 (6.5) | 454 (6.3) | 473（5.9） | 542（6.1） | 582（6.0） | 643（6.1） | 711（6.3） | 712（5.9） |
| PACS 分诊类别 | | | | | | | | | | | |
| P1 | 19169 (23.8) | 778 (24.1) | 950 (23.6) | 1279 (22.7) | 1752 (24.1) | 1865 (23.1) | 2095 (23.6) | 2311 (23.9) | 2579 (24.6) | 2665 (23.6) | 2895 (23.9) |
| P2 | 44572（55.3） | 1800（55.8） | 2210 (54.8) | 3166 (56.1) | 3989 (55.0) | 4547 (56.4) | 4852 (54.7) | 5275 (54.5) | 5762 (55.0) | 6286 (55.7) | 6685 (55.3) |
| P3 和 P4 | 16872 (20.9) | 646 (20.0) | 871 (21.6) | 1198 (21.2) | 1514 (20.9) | 1649 (20.5) | 1920 (21.7) | 2088 (21.6) | 2139 (20.4) | 2335 (20.7) | 2512 (20.8) |
| 轮班时间 | | | | | | | | | | | |
| 上午 8 点至下午 4 点 | 42594 (52.8) | 1707 (52.9) | 2133 (52.9) | 3012（53.4） | 3851（53.1） | 4275（53.0） | 4649（52.4） | 5158（53.3） | 5609（53.5） | 5893 (52.2) | 6307 (52.2) |
| 下午4点至午夜 | 28141（34.9） | 1115（34.6） | 1419 (35.2) | 1937 (34.3) | 2537 (35.0) | 2829 (35.1) | 3112 (35.1) | 3378 (34.9) | 3584 (34.2) | 3961 (35.1) | 4269 (35.3) |
| 午夜至8点 AM | 9878 (12.3) | 402 (12.5) | 479 (11.9) | 694 (12.3) | 867 (12.0) | 957 (11.9) | 1106 (12.5) | 1138 (11.8) | 1287 (12.3) | 1432 (12.7) | 1516 (12.5) |
| 星期 | | | | | | | | | | | |
| 星期五 | 11060（13.7） | 445（13.8） | 555 (13.8) | 763 (13.5) | 983 (13.5) | 1088 (13.5) | 1234 (13.9) | 1313 (13.6) | 1388 (13.2) | 1620 (14.4) | 1671 (13.8) |
| 周中 | 35464 (44.0) | 1435 (44.5) | 1781 (44.2) | 2590 (45.9) | 3178 (43.8) | 3614 (44.8) | 3880 (43.8) | 4272 (44.2) | 4588 (43.8) | 4820 (42.7) | 5306 (43.9) |
| 星期一 | 13311 (16.5) | 528 (16.4) | 640 (15.9) | 905 (16.0) | 1200 (16.5) | 1306 (16.2) | 1486 (16.8) | 1610 (16.6) | 1756 (16.8) | 1897 (16.8) | 1983 (16.4) |
| 周末 | 20778（25.8） | 816（25.3） | 1055 (26.2) | 1385 (24.5) | 1894 (26.1) | 2053 (25.5) | 2267 (25.6) | 2479 (25.6) | 2748 (26.2) | 2949 (26.1) | 3132 (25.9) |
| 生命体征，均值（标准差） | | | | | | | | | | | |
| 脉搏，次/分钟 | 86.38（18.41） | 86.14（18.75） | 86.68（18.26） | 85.89（18.03） | 86.39（18.58） | 86.26（18.34） | 86.51（18.59） | 85.49（18.56） | 86.65（18.38） | 86.36（18.21） | 86.27（18.42） | |
| 呼吸，/min | 18.27 (2.16) | 18.26 (2.16) | 18.31 (2.18) | 18.22 (2.05) | 18.28 (2.21) | 18.22 (2.11) | 18.28 (2.19) | 18.25 (2.14) | 18.30 (2.14) | 18.27 (2.18) | 18.27 (2.22) |
| $SpO_2$, % | 97.37 (4.11) | 97.40 (3.66) | 97.37 (4.42) | 97.25 (4.93) | 97.28 (4.50) | 97.34 (4.03) | 97.43 (3.83) | 97.40 (3.95) | 97.40 (3.85) | 97.42 (3.65) | 97.32 (4.45) |
| 血压，毫米 Hg | | | | | | | | | | | |
| 舒张期 | 72.58 (14.06) | 72.62 (14.07) | 72.32 (13.78) | 72.70 (13.71) | 72.78 (14.09) | 72.67 (14.14) | 72.63 (14.20) | 72.54 (14.25) | 72.62 (14.18) | 72.54 (14.07) | 72.44 (13.87) | |
| 收缩压 | 137.46 (27.98) | 137.95 (28.34) | 136.91 (27.23) | 137.56 (27.78) | 137.64 (27.98) | 137.73 (28.14) | 137.12 (27.72) | 137.45 (28.36) | 137.79 (28.36) | 137.48 (27.93) | 137.13 (27.70) | |
| 合并症 | | | | | | | | | | | |
| 心肌梗死 | 5095 (6.3) | 207 (6.4) | 262（6.5） | 335（5.9） | 462（6.4） | 512（6.4） | 540（6.1） | 621（6.4） | 666 (6.4) | 699 (6.2) | 791 (6.5) |
| 充血性心力衰竭 | 8841 (11.0) | 363 (11.3) | 453 (11.2) | 600 (10.6) | 767 (10.6) | 901 (11.2) | 956 (10.8) | 1044 (10.8) | 1212 (11.6) | 1214 (10.8) | 1331 (11.0) |
| 外周血管疾病 | 4649 (5.8) | 181 (5.6) | 219 (5.4) | 298 (5.3) | 429 (5.9) | 479 (5.9) | 522 (5.9) | 528 (5.5) | 654 (6.2) | 631 (5.6) | 708 (5.9) |
| 中风 | 9527（11.8） | 400（12.4） | 498（12.4） | 653（11.6） | 838（11.6） | 969（12.0） | 1055（11.9） | 1141（11.8） | 1260（12.0） | 1307（11.6） | 1406（11.6） |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 痴呆 | 2854 (3.5) | 133 (4.1) | 137 (3.4) | 195 (3.5) | 253 (3.5) | 278 (3.4) | 334 (3.8) | 322 (3.3) | 369 (3.5) | 386 (3.4) | 447 (3.7) |
| 慢性<br>肺<br>病 | 7089 (8.8) | 290 (9.0) | 378（9.4） | 491（8.7） | 634（8.7） | 678（8.4） | 804（9.1） | 841（8.7） | 887 (8.5) | 994 (8.8) | 1092 (9.0) |
| 类风湿性<br>疾病 | 1165 (1.4) | 53（1.6） | 54（1.3） | 85（1.5） | 97（1.3） | 125（1.6） | 126（1.4） | 136 (1.4) | 155 (1.5) | 164 (1.5) | 170 (1.4) |
| 消化性溃疡<br>疾病 | 2455 (3.0) | 106 (3.3) | 121 (3.0) | 167 (3.0) | 222 (3.1) | 263 (3.3) | 224 (2.5) | 300 (3.1) | 307 (2.9) | 359 (3.2) | 386 (3.2) |
| 糖尿病 | | | | | | | | | | | |
| None | 50165（62.2） | 2004（62.1） | 2526（62.7） | 3601（63.9） | 4491（61.9） | 5007（62.1） | 5529（62.3） | 6018（62.2） | 6448（61.5） | 7020（62.2） | 7521（62.2） |
| 糖尿病<br>无<br>慢性<br>并发症 | 3640（4.5） | 147（4.6） | 189（4.7） | 273（4.8） | 334（4.6） | 339（4.2） | 414（4.7） | 447（4.6） | 502（4.8） | 459（4.1） | 536 (4.4) |
| 伴有<br>复杂性 | 26808 (33.3) | 1073 (33.3) | 1316 (32.6) | 1769 (31.3) | 2430（33.5） | 2715（33.7） | 2924（33.0） | 3209（33.2） | 3530（33.7） | 3807（33.7） | 4035 (33.4) |
| 偏瘫或<br>截瘫 | 3594 (4.5) | 156 (4.8) | 188 (4.7) | 269 (4.8) | 345 (4.8) | 373 (4.6) | 396 (4.5) | 411 (4.2) | 456 (4.4) | 445 (3.9) | 555 (4.6) |
| 肾脏<br>疾病 | 19142（23.7） | 747（23.2） | 947 (23.5) | 1312（23.3） | 1753（24.2） | 1923（23.9） | 2080（23.5） | 2264（23.4） | 2534（24.2） | 2700（23.9） | 2882（23.8） |
| 肝病 | | | | | | | | | | | |
| None | 75111（93.2） | 3006（93.2） | 3764（93.4） | 5251 (93) | 6779 (93.5) | 7510 (93.1) | 8288 (93.4) | 8968 (92.7) | 9745 (93) | 10520 (93.2) | 11280 (93.3) |
| 轻度肝脏<br>疾病 | 3937 (4.9) | 151 (4.7) | 202 (5.0) | 276 (4.9) | 342 (4.7) | 392 (4.9) | 423 (4.8) | 490 (5.1) | 534 (5.1) | 538 (4.8) | 589 (4.9) |
| 严重肝脏<br>疾病 | 1565 (1.9) | 67 (2.1) | 65（1.6） | 116（2.1） | 134（1.8） | 159（2.0） | 156（1.8） | 216（2.2） | 201 (1.9) | 228 (2.0) | 223 (1.8) |
| 医疗保健<br>使用，均值<br>(SD) | | | | | | | | | | | |
| 急诊<br>入院<br>过去一年 | 1.07 (2.40) | 1.03 (2.34) | 1.06 (2.38) | 1.04 (2.38) | 1.06 (2.26) | 1.07 (2.44) | 1.05 (2.35) | 1.09 (2.52) | 1.12 (2.51) | 1.04 (2.25) | 1.06 (2.46) |
| 运营<br>过去一年 | 0.29 (0.98) | 0.27 (0.88) | 0.28 (0.97) | 0.28 (1.04) | 0.29 (0.97) | 0.30 (1.00) | 0.29 (0.99) | 0.29 (0.98) | 0.30 (1.02) | 0.29 (0.96) | 0.28 (0.92) |
| ICU<br>入院次数<br>过去一年 | 0.03 (0.28) | 0.03 (0.31) | 0.02（0.22） | 0.03（0.37） | 0.03（0.33） | 0.03（0.32） | 0.02（0.25） | 0.03（0.29） | 0.02 (0.25) | 0.03 (0.26) | 0.03 (0.28) |
| HD<br>过去一年的入院<br>过去一年 | 0.08 (0.44) | 0.08 (0.46) | 0.08 (0.45) | 0.08 (0.47) | 0.08 (0.44) | 0.07 (0.42) | 0.08 (0.46) | 0.08 (0.44) | 0.08 (0.43) | 0.08 (0.45) | 0.08 (0.45) |
| 结局 | | | | | | | | | | | |
| 30 天<br>死亡率 | 4249 (5.3) | 158 (4.9) | 207 (5.1) | 294 (5.2) | 400 (5.5) | 405 (5.0) | 450 (5.1) | 478 (4.9) | 512 (4.9) | 640 (5.7) | 705 (5.8) |

缩写：HD，高依赖；ICU，重症监护病房；PACS，患者病情严重程度分类量表；$SpO_2$, 脉搏血氧测定测得的血氧饱和度。[a] 除非另有说明，数据以患者人数（百分比）表示。

图1. FedScore框架的流程图。

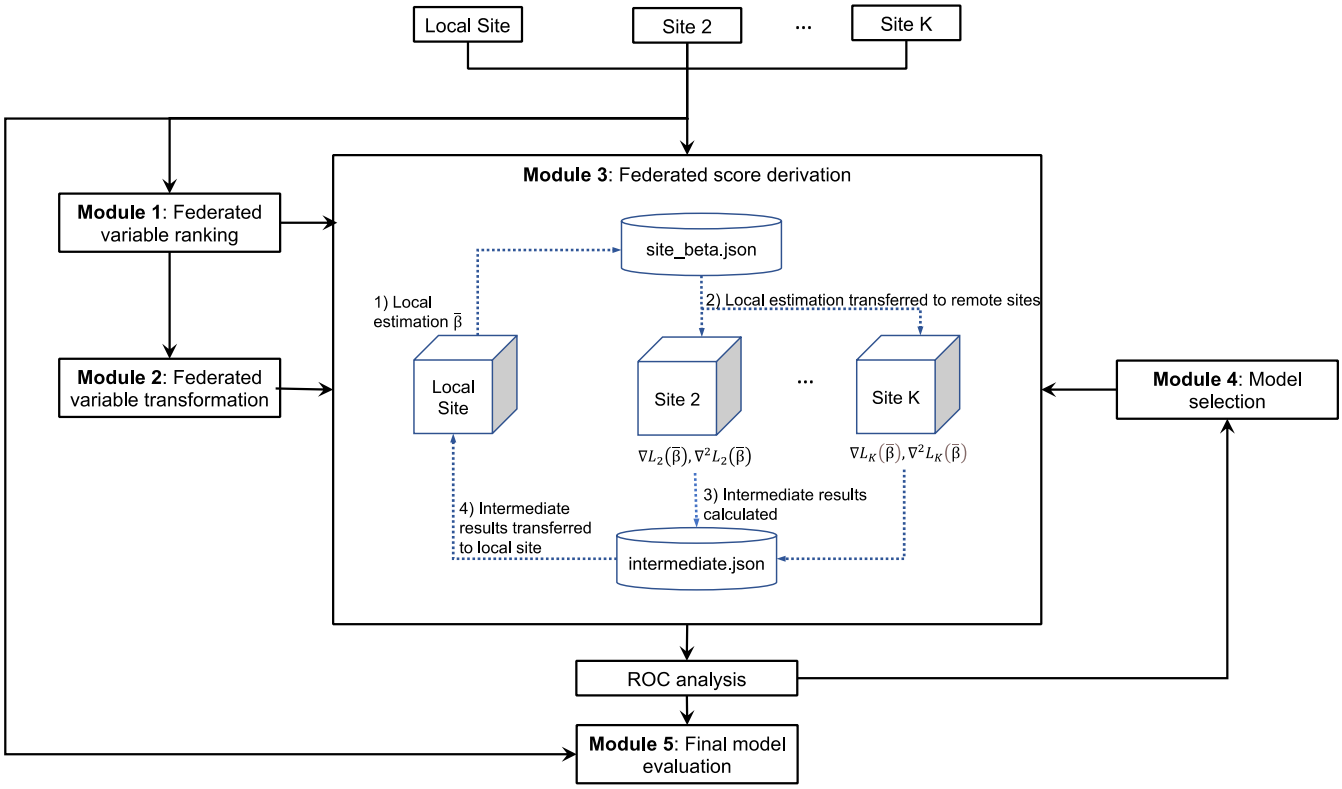



图2. 研究队列形成的流程图。SGH：新加坡中央医院

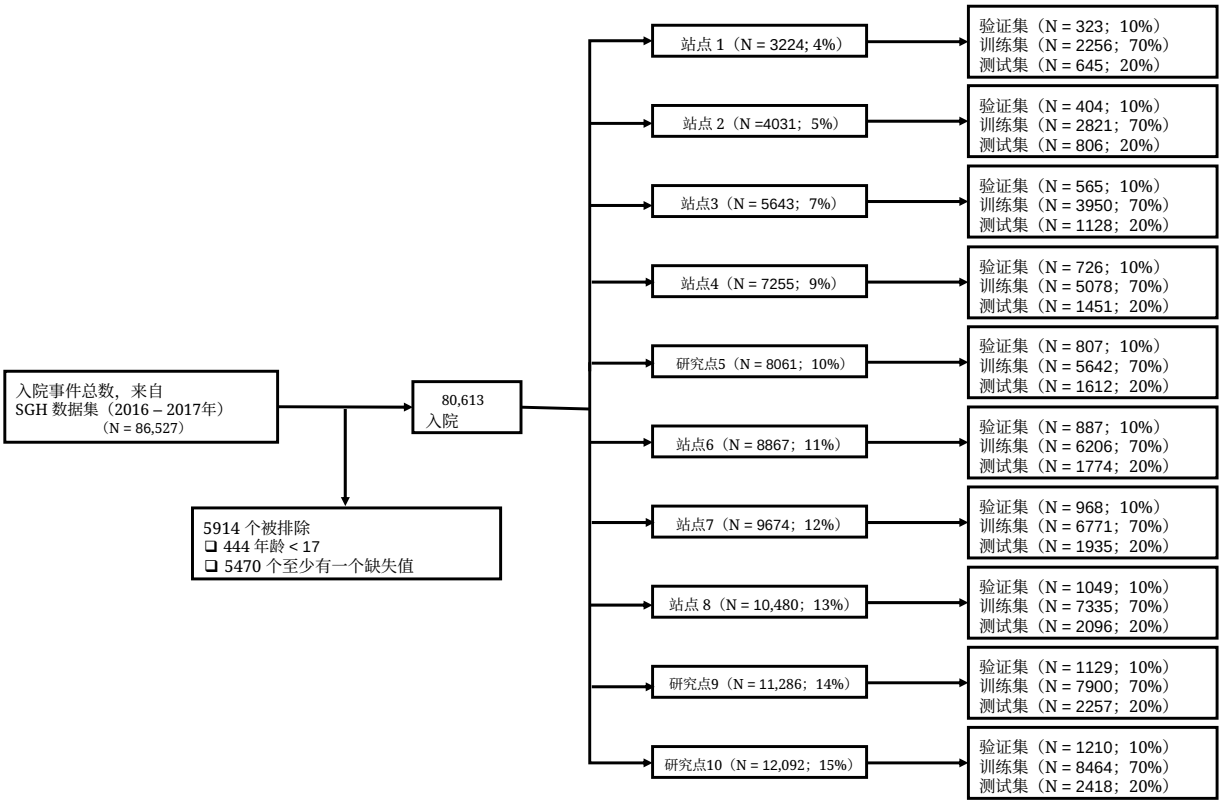
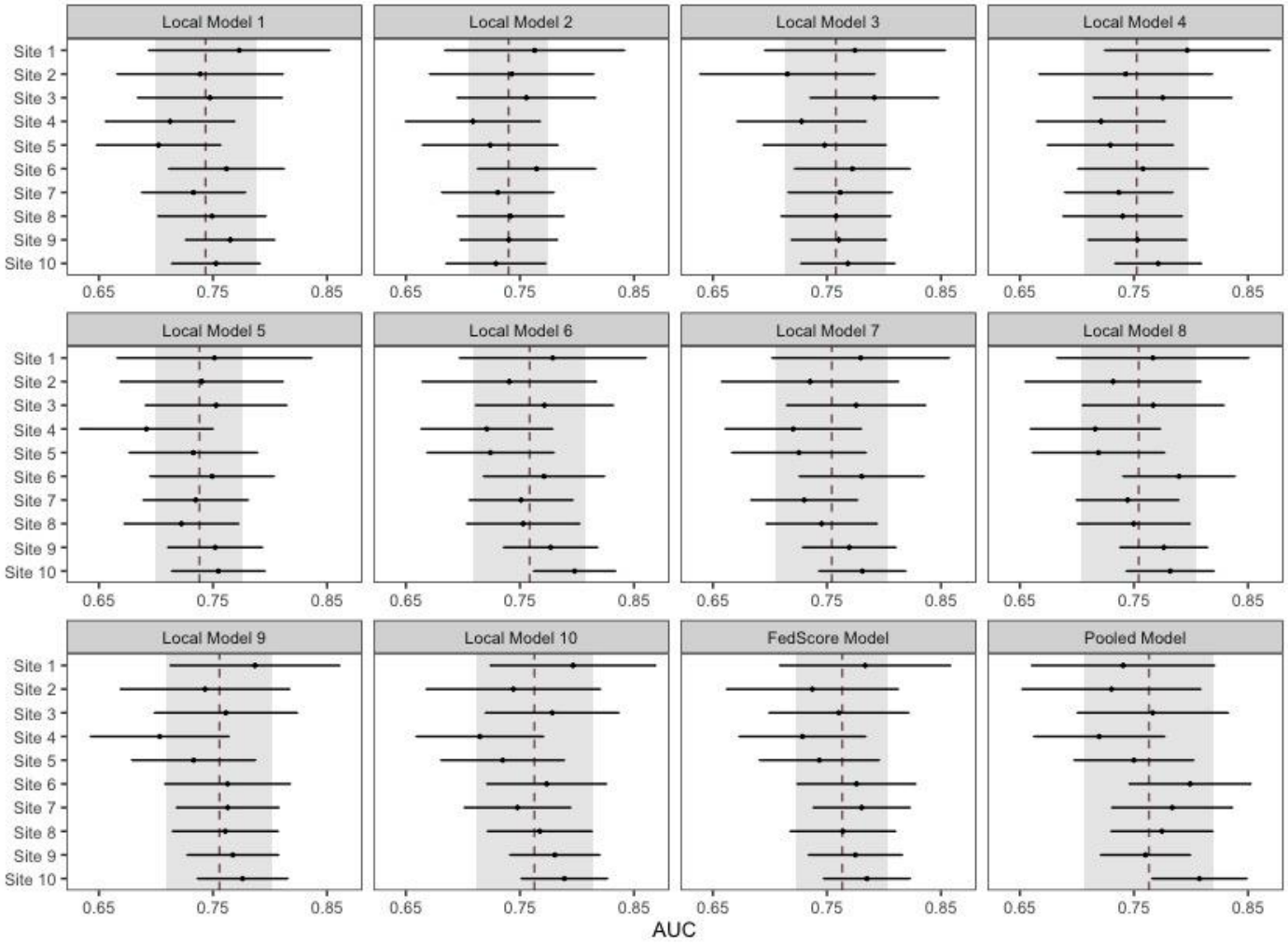


图3. FedScore 与本地评分的性能比较。



T. Kuo 和 A. Pham，“检测去中心化医疗联邦学习中的模型违规行为”，《国际医学信息学杂志》，第 158 卷，文章编号 104658，2022年2月，doi: 10.1016/j.ijmedinf.2021.104658。[2] D. Gotz 和 D. Borland，“数据驱动医疗：交互式可视化的挑战与机遇”，《IEEE计算机图形与应用》，第 36 卷，第 3 期，第 90–96 页，2016年5月，doi: 10.1109/MCG.2016.59。[3] A. I. R. Maas 等，“欧洲创伤性脑损伤神经创伤疗效协作研究（CENTER-TBI）：一项前瞻性纵向观察性研究”，《神经外科》，第 76 卷，第 1 期，第 67–80 页，2015年1月，doi: 10.1227/NEU.0000000000000575。[4] A. J. Webb、G. A. Thorisson、A. J. Brookes，以及代表 G. Consortium，“支持现代基因型到表型研究的信息学项目与在线‘知识中心’”，《人类突变》，第 32 卷，第 5 期，第 543–550 页，2011年，doi: 10.1002/humu.21469。[5] S. Anker、F. W. Asselbergs、G. Brobert、P. Vardas、D. E. Grobbee 和 M. Cronin，“心血管疾病大数据”，《欧洲心脏杂志》，第 38 卷，第 24 期，第 1863–1865 页，2017年6月，doi: 10.1093/eurheartj/ehx283。[6] I. G. Stiell 等，“多中心基础生命支持/除颤系统中与心脏骤停生存率改善相关的可调节因素：OPALS研究第一阶段结果”，《急诊医学年鉴》，第 33 卷，第 1 期，第 44–50 页，1999年1月，doi: 10.1016/S0196-0644(99)70415-4。[7] “凯撒医疗研究库——凯撒医疗”，凯撒医疗研究库。https://researchbank.kaiserpermanente.org/（访问日期：2022年8月10日）。[8] M. E. H. Ong 等，“泛亚复苏结局研究（PAROS）：研究依据、方法学与实施”，《学术急诊医学》学会官方期刊，第 18 卷，第 8 期，第 890–897 页，2011年8月，doi: 10.1111/j.1553-2712.2011.01132.x。[9] R. S. Antunes、C. André da Costa、A. Küderle、I. A. Yari 和 B. Eskofier，“医疗联邦学习：系统综述与架构提案”，《ACM智能系统与技术汇刊》，第 13 卷，第 4 期，文章编号 54:1-54:23，2022年5月，doi: 10.1145/3501813。[10] D. C. Nguyen 等，“智慧医疗联邦学习：综述”，《ACM计算调查》，第 55 卷，第 3 期，文章编号 60:1-60:37，2022年2月，doi: 10.1145/3501296。[11] C. J. Hoofnagle、B. van der Sloot 和 F. Z. Borgesius，“欧盟《通用数据保护条例》：内容及其意义”，《信息通信技术法》，第 28 卷，第 1 期，第 65–98 页，2019年1月，doi: 10.1080/13600834.2019.1573501。[12] N. Rieke 等，“联邦学习助力数字健康的未来”，《npj数字医学》，第 3 卷，第 1 期，文章编号 1，2020年9月，doi: 10.1038/s41746-020-00323-1。[13] M. J. Sheller 等，“医学领域的联邦学习：促进多机构协作且无需共享患者数据”，《科学报告》，第 10 卷，第 1 期，文章编号 1，2020年7月，doi: 10.1038/s41598-020- 69250-1。[14] I. Dayan 等，“用于预测新冠肺炎患者临床结局的联邦学习”，《自然医学》，第 27 卷，第 10 期，文章编号 10，2021年10月，doi: 10.1038/s41591-021-01506-3。[15] C. Luo 等，“DLMM：用于协作式多站点分布式线性混合模型的无损单步算法”，《自然通讯》，第 13 卷，第 1 期，文章编号 1，2022年3月，doi: 10.1038/s41467-022-29160-4。[16] C. Rudin、C. Chen、Z. Chen、H. Huang、L. Semenova 和 C. Zhong，“可解释机器学习：基本原则与十大重大挑战”，《统计调查》，第 16 卷，期号 none，第 1–85 页，2022年1月，doi: 10.1214/21-SS133。[17] V. Fleig、F. Brenck、M. Wolff 和 M. A. Weigand，“[重症医学中的评分系统：原则、模型、应用与局限],”，《麻醉学家》，第 60 卷，第 10 期，第 963–974 页，2011年10月，doi: 10.1007/s00101-011-1942-8。[18] “昏迷与意识障碍评估——《柳叶刀》。”https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(74)91639-0/fulltext（访问日期：2022年6月6日）。[19] B. Ustun 和 C. Rudin，“用于优化医学评分系统的超稀疏线性整数模型”，《机器学习》，第 102 卷，第 3 期，第 349–391 页，2016年3月，doi: 10.1007/s10994-015-5528-6。

[20] N. Sokolovska、Y. Chevaleyre、K. Clément 和 J.-D. Zucker，“用于学习可解释医疗评分系统的融合套索惩罚”，载于《2017年国际神经网络联合会议（IJCNN）》，2017年5月，第 4504–4511 页。doi: 10.1109/IJCNN.2017.7966427。[21] N. Sokolovska、Y. Chevaleyre 和 J.-D. Zucker，“用于学习可解释评分系统的可证明算法”，载于《第二十一届国际人工智能与统计学会议论文集》，2018年3月，第 566–574 页。访问日期：2022年8月8日。[在线]。可访问：https://proceedings.mlr.press/v84/sokolovska18a.html[22] L. Billiet、S. V. Huffel 和 V. V. Belle，“区间编码评分：可解释评分系统的工具箱”，《PeerJ计算机科学》，第 4 卷，文章编号 e150，2018年4月，doi: 10.7717/peerj-cs.150。[23] F. Xie、B. Chakraborty、M. E. H. Ong、B. A. Goldstein 和 N. Liu，“AutoScore：一种基于机器学习的自动临床评分生成器及其在利用电子健康记录进行死亡率预测中的应用”，《JMIR医学信息学》，第 8 卷，第 10 期，文章编号 e21798，2020年10月，doi: 10.2196/21798。[24] N. Liu 等，“利用机器学习开发并验证用于院外心脏骤停患者院前自主循环恢复（P-ROSC）评分，以进行回顾性研究”，《eClinicalMedicine》，第 48 卷，文章编号 101422，2022年6月，doi: 10.1016/j.eclinm.2022.101422。[25] M. M. Churpek、T. C. Yuen、S. Y. Park、D. O. Meltzer、J. B. Hall 和 D. P. Edelson，“利用病房生命体征推导心脏骤停预测模型*”，《重症监护医学》，第 40 卷，第 7 期，第 2102–2108 页，2012年7月，doi: 10.1097/CCM.0b013e318250aa5a。[26] G. B. Smith、D. R. Prytherch、P. Meredith、P. E. Schmidt 和 P. I. Featherstone，“国家早期预警评分（NEWS）区分早期心脏骤停、意外入住重症监护病房及死亡风险患者的能力”，《复苏》，第 84 卷，第 4 期，第 465–470 页，2013年4月，doi: 10.1016/j.resuscitation.2012.12.016。[27] W. Brady 和 K. de Souza，“HEART评分：其在急诊科中的应用指南”，《土耳其急诊医学杂志》，第 18 卷，第 2 期，第 47–51 页，2018年6月，doi: 10.1016/j.tjem.2018.04.004。[28] H. R. Arkes、S. K. Aberegg 和 K. A. Arpin，“基于事件序列分析医生对医学结局的概率估计”，《JAMA网络开放》，第 5 卷，第 6 期，文章编号 e2218804，2022年6月，doi: 10.1001/jamanetworkopen.2022.18804。[29] L. N. Sanchez-Pinto、L. R. Venable、J. Fahrenbach 和 M. M. Churpek，“临床预测建模中变量选择方法的比较”，《国际医学信息学杂志》，第 116 卷，第 10–17 页，2018年8月，doi: 10.1016/j.ijmedinf.2018.05.006。[30] F. Xie 等，“用于估计急诊入院后死亡率的可解释机器学习分诊工具的开发与评估”，《JAMA网络开放》，第 4 卷，第 8 期，文章编号 e2118467，2021年8月，doi: 10.1001/jamanetworkopen.2021.18467。[31] Y. Ang 等，“在急诊科早期识别急性肾损伤的可解释临床评分的开发与验证”，《科学报告》，第 12 卷，第 1 期，文章编号 1，2022年5月，doi: 10.1038/s41598-022-11129-4。[32] B. Gregorutti、B. Michel 和 P. Saint-Pierre，“随机森林中的相关性与变量重要性”，《统计计算》，第 27 卷，第 3 期，第 659–678 页，2017年5月，doi: 10.1007/s11222-016-9646-1。[33] E. V. A. Sylvester 等，“随机森林特征选择在精细尺度遗传群体归属中的应用”，《进化应用》，第 11 卷，第 2 期，第 153–165 页，2018年，doi: 10.1111/eva.12524。[34] J. K. Jaiswal 和 R. Samikannu，“随机森林算法在特征子集选择以及分类与回归中的应用”，载于《2017年世界计算与通信技术大会（WCCCT）》，2017年2月，第 65–68 页。doi: 10.1109/WCCCT.2016.25。[35] R. Genuer、J.-M. Poggi 和 C. Tuleau-Malot，“使用随机森林进行变量选择”，《模式识别快报》，第 31 卷，第 14 期，第 2225–2236 页，2010年10月，doi: 10.1016/j.patrec.2010.03.014。[36] A. P. Marques Ramos 等，“利用无人机植被光谱指数通过随机森林排序方法预测玉米产量”，《计算机与电子农业》，第 178 卷，文章编号 105791，2020年11月，doi: 10.1016/j.compag.2020.105791。[37] L. G. Forni 等，“识别急性肾损伤风险患者：急性内科患者发生急性肾损伤的预测评分系统”，《Nephron临床实践》，第 123 卷，第 3–4 期，第 143–150 页，2013年，doi: 10.1159/000351509。

[38] M. E. Charlson、P. Pompei、K. L. Ales 和 C. R. MacKenzie，《纵向研究中预后合并症分类的新方法：开发与验证》，《慢性病杂志》，第40卷，第5期，第373–383页，1987年，doi: 10.1016/0021-9681(87)90171-8。[39] J. L. Vincent 等人，《SOFA（脓毒症相关器官衰竭评估）评分：用于描述器官功能障碍/衰竭。代表欧洲重症医学会脓毒症相关问题工作组》，《重症医学》，第22卷，第7期，第707–710页，1996年7月，doi: 10.1007/BF01709751。[40] A. J. Six、B. E. Backus 和 J. C. Kelder，《急诊室胸痛：HEART评分的价值》，《荷兰心脏杂志》，第16卷，第6期，第191–196页，2008年6月。[41] M. Jones，《NEWSDIG：国家早期预警评分开发与实施小组》，《临床医学》，第12卷，第6期，第501–503页，2012年12月，doi: 10.7861/clinmedicine.12-6-501。[42] E. Seth、M. E, BentleyEllison、T. M 和 M. J，《SPOTS系统：针对现代临床前药物开发和毒理学优化的眼部评分系统》，《眼科药理学与治疗杂志》，2017年12月，doi: 10.1089/jop.2017.0108。[43] E. Baldi 等人，《基于Utstein模式的评分：预测患者存活至住院的UB-ROSC评分》，《国际心脏病学杂志》，第308卷，第84–89页，2020年6月，doi: 10.1016/j.ijcard.2020.01.032。[44] R. Duan 等人，《从多个站点的电子健康记录中学习：一种通信高效且保护隐私的分布式算法》，《美国医学信息学会杂志》，第27卷，第3期，第376–385页，2019年12月，doi: 10.1093/jamia/ocz199。[45] R. Y. Fong、W. S. S. Glen、A. K. Mohamed Jamil、W. W. S. Tam 和 Y. Kowitlawakul，《急诊环境下急诊严重程度指数与患者病情严重程度分类量表的比较》，《国际急诊护理》，第41卷，第13–18页，2018年11月，doi: 10.1016/j.ienj.2018.05.001。[46] M. G. Crowson 等人，《生物医学数据联邦学习应用的系统综述》，《PLOS数字健康》，第1卷，第5期，文章编号e0000033，2022年5月，doi: 10.1371/journal.pdig.0000033。[47] C. Zhang、Y. Xie、H. Bai、B. Yu、W. Li 和 Y. Gao，《联邦学习综述》，《知识型系统》，第216卷，文章编号106775，2021年3月，doi: 10.1016/j.knocys.2021.106775。[48] A. A. H. de Hond 等人，《医疗领域基于人工智能的预测模型指南与质量标准：范围综述》，《npj数字医学》，第5卷，第1期，文章编号1，2022年1月，doi: 10.1038/s41746-021-00549-7。[49] V. Volovici、N. L. Syn、A. Ercole、J. J. Zhao 和 N. Liu，《避免在临床研究中滥用和误用机器学习的步骤》，《自然医学》，第1–4页，2022年9月，doi: 10.1038/s41591-022-01961-6。[50] C. Yang 等人，《临床预测模型开发与验证的实施及报告趋势：系统综述》，《美国医学信息学会杂志》，第29卷，第5期，第983–989页，2022年5月，doi: 10.1093/jamia/ocac002。[51] 《联邦学习：挑战、方法与未来方向》。http://ieeexplore.ieee.org/document/9084352（访问日期：2022年6月23日）。[52] J. Tong 等人，《面向异质性临床数据的分布式学习及其在整合230个站点新冠肺炎数据中的应用》，《npj数字医学》，第5卷，第1期，文章编号1，2022年6月，doi: 10.1038/s41746- 022-00615-8。[53] C. Luo 等人，《dPQL：一种用于广义线性混合模型的无损分布式算法及其在隐私保护医院概况分析中的应用》，《美国医学信息学会杂志》，第29卷，第8期，第1366–1371页，2022年8月，doi: 10.1093/jamia/ocac067。

补充材料

**eT**表1：FedScore模型与基线模型的性能比较。

电子表格 2：评分表 (a) - (j)：通过 AutoScore 在站点 1 至站点 10 分别生成的本地模型的评分表。(k)：通过 FedScore 生成的联邦模型的评分表。(l)：通过 AutoScore 生成的合并模型的评分表

图1：简约性图 (a) - (j)：通过 AutoScore 在站点 1 至站点 10 分别生成的本地模型的简约性图。

(k)：通过 FedScore 生成的联邦模型的简约性图。(l)：通过 AutoScore 生成的合并模型的简约性图

电子表格1 FedScore模型与基线模型的性能比较.

| 模型 | 变量数量 | 测试数据 | | | | | | | | | | | | | | | | | | 均值 曲线下面积 每个模型在全部10个站点上 | 标准差 曲线下面积 each 模型在全部10个站点 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SGH1 | | SGH2 | | SGH3 | | SGH4 | | SGH5 | | SGH6 | | SGH7 | | SGH8 | | SGH9 | | SGH10 | | | |
| | | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | 曲线下面积 | 置信区间 | | |
| 模型 1[a] | 7 | 0.7730 | 0.6936-0.8523 | 0.7387 | 0.6659-0.8114 | 0.7473 | 0.6839-0.8106 | 0.7124 | 0.6557-0.7690 | 0.7023 | 0.6479-0.7567 | 0.7618 | 0.7113-0.8123 | 0.7329 | 0.6874-0.7784 | 0.7491 | 0.7018-0.7964 | 0.7651 | 0.7261-0.8040 | 0.7526 | 0.7139-0.7913 | 0.7435 | 0.0226 |
| 模型 2[b] | 5 | 0.7628 | 0.6841-0.8416 | 0.7427 | 0.6708-0.8147 | 0.7556 | 0.6950-0.8162 | 0.7087 | 0.6497-0.7677 | 0.7239 | 0.6645-0.7833 | 0.7645 | 0.7128-0.8163 | 0.7305 | 0.6814-0.7797 | 0.7418 | 0.6952-0.7885 | 0.7402 | 0.6976-0.7828 | 0.7289 | 0.6853-0.7726 | 0.7400 | 0.0177 |
| 模型 3[c] | 7 | 0.7745 | 0.6954-0.8537 | 0.7153 | 0.6386-0.7920 | 0.7915 | 0.7353-0.8477 | 0.7277 | 0.6713-0.7842 | 0.7479 | 0.6941-0.8017 | 0.7723 | 0.7217-0.8229 | 0.7616 | 0.7160-0.8071 | 0.7579 | 0.7098-0.8060 | 0.7604 | 0.7189-0.8020 | 0.7683 | 0.7271-0.8095 | 0.7577 | 0.0225 |
| 模型 4[d] | 7 | 0.7967 | 0.7243-0.8690 | 0.7427 | 0.6667-0.8187 | 0.7752 | 0.7145-0.8358 | 0.7211 | 0.6647-0.7774 | 0.7293 | 0.6742-0.7844 | 0.7579 | 0.7008-0.8149 | 0.7367 | 0.6893-0.7841 | 0.7400 | 0.6878-0.7923 | 0.7530 | 0.7099-0.7961 | 0.7712 | 0.7334-0.8091 | 0.7524 | 0.0233 |
| 模型 5[e] | 4 | 0.7512 | 0.6659-0.8366 | 0.7401 | 0.6688-0.8114 | 0.7527 | 0.6908-0.8146 | 0.6917 | 0.6334-0.7500 | 0.7328 | 0.6765-0.7891 | 0.7492 | 0.6948-0.8035 | 0.7348 | 0.6889-0.7808 | 0.7223 | 0.6721-0.7724 | 0.7519 | 0.7105-0.7933 | 0.7547 | 0.7139-0.7956 | 0.7381 | 0.0195 |
| 模型 6[f] | 7 | 0.7787 | 0.6969-0.8605 | 0.7406 | 0.6643-0.8170 | 0.7714 | 0.7111-0.8317 | 0.7209 | 0.6633-0.7786 | 0.7241 | 0.6685-0.7797 | 0.7710 | 0.7181-0.8240 | 0.7509 | 0.7055-0.7963 | 0.7528 | 0.7033-0.8023 | 0.7769 | 0.7357-0.8180 | 0.7979 | 0.762-0.8338 | 0.7585 | 0.02508 |
| 模型 7[g] | 8 | 0.7795 | 0.7019-0.8571 | 0.7352 | 0.6576-0.8127 | 0.7755 | 0.7147-0.8364 | 0.7204 | 0.6608-0.7801 | 0.7254 | 0.6666-0.7842 | 0.7804 | 0.7258-0.8350 | 0.7300 | 0.6833-0.7767 | 0.7453 | 0.6967-0.7939 | 0.7696 | 0.7288-0.8104 | 0.7810 | 0.7429-0.8190 | 0.7542 | 0.0252 |
| 模型 8[h] | 8 | 0.7666 | 0.6826-0.8507 | 0.7317 | 0.6546-0.8087 | 0.7669 | 0.7049-0.8288 | 0.7161 | 0.6592-0.7731 | 0.7189 | 0.6610-0.7768 | 0.7895 | 0.7404-0.8386 | 0.7444 | 0.6995-0.7892 | 0.7498 | 0.7005-0.7992 | 0.7762 | 0.7378-0.8145 | 0.7818 | 0.7434-0.8202 | 0.7542 | 0.0261 |
| 模型 9[i] | 6 | 0.7868 | 0.7126-0.8610 | 0.7430 | 0.6689-0.8171 | 0.7613 | 0.6987-0.8239 | 0.7033 | 0.6426-0.7640 | 0.7330 | 0.6788-0.7872 | 0.7628 | 0.7078-0.8177 | 0.7629 | 0.7181-0.8077 | 0.7608 | 0.7145-0.8071 | 0.7674 | 0.7273-0.8074 | 0.7759 | 0.7364-0.8155 | 0.7557 | 0.0238 |
| 模型 10[j] | 7 | 0.7965 | 0.7240-0.8689 | 0.7442 | 0.6679-0.8204 | 0.7783 | 0.7198-0.8367 | 0.7148 | 0.6591-0.7706 | 0.7348 | 0.6807-0.7888 | 0.7734 | 0.7210-0.8258 | 0.7478 | 0.7013-0.7944 | 0.7674 | 0.7215-0.8132 | 0.7806 | 0.7412-0.8199 | 0.7889 | 0.7512-0.8267 | 0.7627 | 0.0262 |
| 联邦模型[k] | 8 | 0.7834 | 0.7085-0.8583 | 0.7372 | 0.6621-0.8124 | 0.7604 | 0.6992-0.8217 | 0.7285 | 0.6730-0.7839 | 0.7433 | 0.6908-0.7958 | 0.7758 | 0.7239-0.8278 | 0.7804 | 0.7381-0.8228 | 0.7640 | 0.7178-0.8101 | 0.7749 | 0.7338-0.8160 | 0.7850 | 0.7471-0.8229 | 0.7633 | 0.0204 |
| 模型汇总[l] | 8 | 0.7405 | 0.6603-0.8206 | 0.7303 | 0.6520-0.8085 | 0.7665 | 0.7004-0.8326 | 0.7195 | 0.6621-0.7768 | 0.7499 | 0.6975-0.8023 | 0.7992 | 0.7459-0.8525 | 0.7835 | 0.7307-0.8363 | 0.7745 | 0.7300-0.8191 | 0.7601 | 0.7208-0.7994 | 0.8074 | 0.7658-0.8490 | 0.7631 | 0.0289 |
| 所有 10 个本地模型的平均曲线下面积在各站点上的 | | 0.77663 | | 0.73742 | | 0.76757 | | 0.71371 | | 0.72724 | | 0.76828 | | 0.74325 | | 0.74872 | | 0.76413 | | 0.77012 | | | |

[a-j]通过 AutoScore 在站点 1 至 10 上独立获得的本地模型[k] 通过 FedScore 获得的联邦模型 [l] 通过 AutoScore 获得的合并模型

电子表格 2：评分表 (a) 通过 AutoScore 在站点 1 上生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <70 | 0 |
| | [70,100) | 1 |
| | [100,120) | 3 |
| | >=120 | 6 |
| 收缩压（毫米汞柱） | <98 | 6 |
| | [98,114) | 8 |
| | [114,160) | 4 |
| | [160,188) | 0 |
| | >=188 | 3 |
| 年龄（年） | <30 | 0 |
| | [30,50) | 61 |
| | [50,79) | 69 |
| | [79,88) | 71 |
| | >=88 | 72 |
| 舒张压（毫米汞柱） | <52 | 3 |
| | [52,60) | 1 |
| | [60,83) | 0 |
| | [83,96) | 3 |
| | >=96 | 1 |
| 血氧饱和度（%） | <94 | 4 |
| | [94,96) | 1 |
| | >=96 | 0 |
| 呼吸频率（每分钟） | <16 | 3 |
| | [16,17) | 0 |
| | [17,19) | 2 |
| | [19,22) | 4 |
| | >=22 | 2 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,5) | 2 |
| | >=5 | 4 |

(b) 通过 AutoScore 在站点 2 上生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 1 |
| | [60,71) | 0 |
| | [71,100) | 1 |
| | >=100 | 5 |
| 收缩压（毫米汞柱） | <99 | 12 |
| | [99,114) | 9 |
| | [114,158) | 8 |
| | >=158 | 0 |
| 年龄（年） | <29 | 0 |

| | | |
|---|---|---|
| | [29,49) | 63 |
| | [49,79) | 71 |
| | [79,88) | 73 |
| | >=88 | 75 |
| 舒张压（毫米汞柱） | <53 | 2 |
| | [53,60) | 1 |
| | [60,83) | 0 |
| | [83,96) | 2 |
| | >=96 | 1 |
| 血氧饱和度（%） | <94 | 5 |
| | [94,96) | 1 |
| | [96,99) | 0 |
| | >=99 | 1 |

(c) 通过 AutoScore 在站点3生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 1 |
| | [60,71) | 0 |
| | [71,99) | 1 |
| | [99,117) | 8 |
| | >=117 | 15 |
| 收缩压（毫米汞柱） | <98 | 19 |
| | [98,114) | 13 |
| | [114,159) | 9 |
| | [159,188) | 0 |
| | >=188 | 4 |
| 年龄（年） | <29 | 0 |
| | [29,48) | 10 |
| | [48,79) | 19 |
| | [79,88) | 25 |
| | >=88 | 30 |
| 舒张压（毫米汞柱） | <53 | 11 |
| | [53,61) | 10 |
| | [61,96) | 6 |
| | >=96 | 0 |
| 血氧饱和度（%） | <93 | 10 |
| | [93,99) | 0 |
| | >=99 | 1 |
| 呼吸频率（每分钟） | <19 | 0 |
| | [19,22) | 3 |
| | >=22 | 7 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,4) | 8 |
| | >=4 | 4 |

(d) 通过 AutoScore 在站点4生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 1 |
| | [60,71) | 0 |
| | [71,100) | 1 |
| | [100,119) | 4 |
| | >=119 | 5 |
| 收缩压（毫米汞柱） | <98 | 13 |
| | [98,115) | 10 |
| | [115,160) | 9 |
| | [160,187) | 5 |
| | >=187 | 0 |
| 年龄（年） | <28 | 0 |
| | [28,50) | 60 |
| | [50,80) | 64 |
| | [80,89) | 66 |
| | >=89 | 68 |
| 舒张压（毫米汞柱） | <52 | 3 |
| | [52,61) | 1 |
| | [61,84) | 0 |
| | [84,97) | 1 |
| | >=97 | 3 |
| 血氧饱和度（%） | <94 | 4 |
| | [94,96) | 0 |
| | [96,99) | 1 |
| | >=99 | 2 |
| 呼吸频率（每分钟） | <16 | 4 |
| | [16,17) | 0 |
| | [17,19) | 2 |
| | [19,22) | 3 |
| | >=22 | 5 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,5) | 3 |
| | >=5 | 2 |

（e）通过 AutoScore 在研究点5上生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 1 |
| | [60,71) | 0 |
| | [71,100) | 5 |
| | [100,119) | 17 |
| | >=119 | 24 |
| 收缩压（毫米汞柱） | <98 | 24 |
| | [98,114) | 18 |
| | [114,160) | 6 |
| | [160,188) | 0 |

| | | |
|---|---|---|
| | >=188 | 3 |
| 年龄（年） | <29 | 3 |
| | [29,49) | 0 |
| | [49,79) | 30 |
| | [79,88) | 41 |
| | >=88 | 43 |
| 舒张压（毫米汞柱） | <52 | 9 |
| | [52,61) | 2 |
| | [61,84) | 1 |
| | >=84 | 0 |

(f) 站点6通过 AutoScore 生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <71 | 0 |
| | [71,100) | 2 |
| | [700,120) | 9 |
| | >=120 | 14 |
| 收缩压（毫米汞柱） | <98 | 16 |
| | [98,114) | 15 |
| | [114,159) | 7 |
| | [159,186) | 6 |
| | >=186 | 0 |
| 年龄（年） | <28 | 0 |
| | [28,48) | 13 |
| | [48,80) | 28 |
| | [80,88) | 30 |
| | >=88 | 32 |
| 舒张压（毫米汞柱） | <52 | 11 |
| | [52,84) | 4 |
| | [84,97) | 0 |
| | >=97 | 6 |
| 血氧饱和度（%） | <94 | 9 |
| | [94,99) | 4 |
| | >=99 | 0 |
| 呼吸频率（每分钟） | <16 | 9 |
| | [16,17) | 0 |
| | [17,19) | 2 |
| | [19,22) | 5 |
| | >=22 | 11 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,4) | 6 |
| | >=4 | 3 |

(g) 站点7通过 AutoScore 生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|

| | | |
|---|---|---|
| 脉搏率（每分钟） | <60 | 5 |
| | [60,71) | 0 |
| | [71,100) | 2 |
| | >=100 | 9 |
| 收缩压（毫米汞柱） | <98 | 31 |
| | [98,114) | 24 |
| | [114,160) | 16 |
| | [160,189) | 7 |
| | >=189 | 0 |
| 年龄（年） | <30 | 0 |
| | [30,49) | 7 |
| | [49,79) | 18 |
| | [79,88) | 22 |
| | >=88 | 25 |
| 舒张压（毫米汞柱） | <52 | 4 |
| | [52,61) | 3 |
| | [61,84) | 0 |
| | [84,97) | 3 |
| | >=97 | 7 |
| 血氧饱和度（%） | <94 | 7 |
| | [94,96) | 3 |
| | [94,99) | 0 |
| | >=99 | 3 |
| 呼吸频率（每分钟） | <16 | 12 |
| | [16,17) | 0 |
| | [17,19) | 5 |
| | [19,22) | 8 |
| | >=22 | 15 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,5) | 5 |
| | >=5 | 4 |
| 星期 | 星期五 | 1 |
| | 周中 | 0 |
| | 其他 | 2 |

（h）通过 AutoScore 在站点 8 上生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <61 | 0 |
| | [61,71) | 2 |
| | [71,100) | 0 |
| | [100,118) | 5 |
| | >=118 | 10 |
| 收缩压（毫米汞柱） | <98 | 15 |
| | [98,114) | 11 |
| | [114,160) | 3 |
| | [160,189) | 2 |

| | | |
|---|---|---|
| | >=189 | 0 |
| 年龄（年） | <28 | 0 |
| | [28,49) | 18 |
| | [48,79) | 31 |
| | [79,88) | 35 |
| | >=88 | 39 |
| 舒张压（毫米汞柱） | <52 | 5 |
| | [52,61) | 3 |
| | [61,84) | 1 |
| | [84,97) | 0 |
| | >=97 | 4 |
| 血氧饱和度（%） | <96 | 7 |
| | [96,99) | 0 |
| | >=99 | 2 |
| 呼吸频率（每分钟） | <16 | 8 |
| | [16,17) | 0 |
| | [17,19) | 3 |
| | [19,22) | 5 |
| | >=22 | 13 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,5) | 9 |
| | >=5 | 2 |
| 星期 | 周中 | 0 |
| | 星期一 | 3 |
| | 其他 | 2 |

(i) 通过 AutoScore 在研究点9上生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 1 |
| | [60,71) | 0 |
| | [71,100) | 4 |
| | [100,119) | 10 |
| | >=119 | 16 |
| 收缩压（毫米汞柱） | <98 | 21 |
| | [98,115) | 13 |
| | [115,159) | 8 |
| | [159,187) | 5 |
| | >=187 | 0 |
| 年龄（年） | <29 | 0 |
| | [29,49) | 14 |
| | [49,79) | 31 |
| | [79,88) | 36 |
| | >=88 | 39 |
| 舒张压（毫米汞柱） | <52 | 5 |
| | [52,61) | 2 |
| | [61,84) | 0 |

| | | |
|---|---|---|
| | [84,96) | 2 |
| | >=96 | 3 |
| 血氧饱和度（%） | <94 | 11 |
| | [94,96) | 1 |
| | [96,99) | 0 |
| | >=99 | 1 |
| 呼吸频率（每分钟） | <16 | 0 |
| | [16,17) | 3 |
| | [17,19) | 2 |
| | [19,22) | 5 |
| | >=22 | 8 |

(j) 通过 AutoScore 在研究点10生成的本地模型评分表。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 3 |
| | [60,71) | 0 |
| | [71,100) | 2 |
| | [100,119) | 8 |
| | >=119 | 14 |
| 收缩压（毫米汞柱） | <99 | 22 |
| | [99,114) | 15 |
| | [114,158) | 10 |
| | [158,186) | 6 |
| | >=186 | 0 |
| 年龄（年） | <29 | 0 |
| | [29,49) | 13 |
| | [49,79) | 25 |
| | [79,88) | 29 |
| | >=88 | 30 |
| 舒张压（毫米汞柱） | <53 | 5 |
| | [53,83) | 3 |
| | [83,96) | 0 |
| | >=96 | 3 |
| 血氧饱和度（%） | <94 | 9 |
| | [94,96) | 1 |
| | >=96 | 0 |
| 呼吸频率（每分钟） | <16 | 3 |
| | [16,17) | 0 |
| | [17,19) | 4 |
| | [19,22) | 8 |
| | >=22 | 13 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,4) | 7 |
| | >=4 | 4 |

(k) 通过 FedScore 生成的联邦模型评分表。

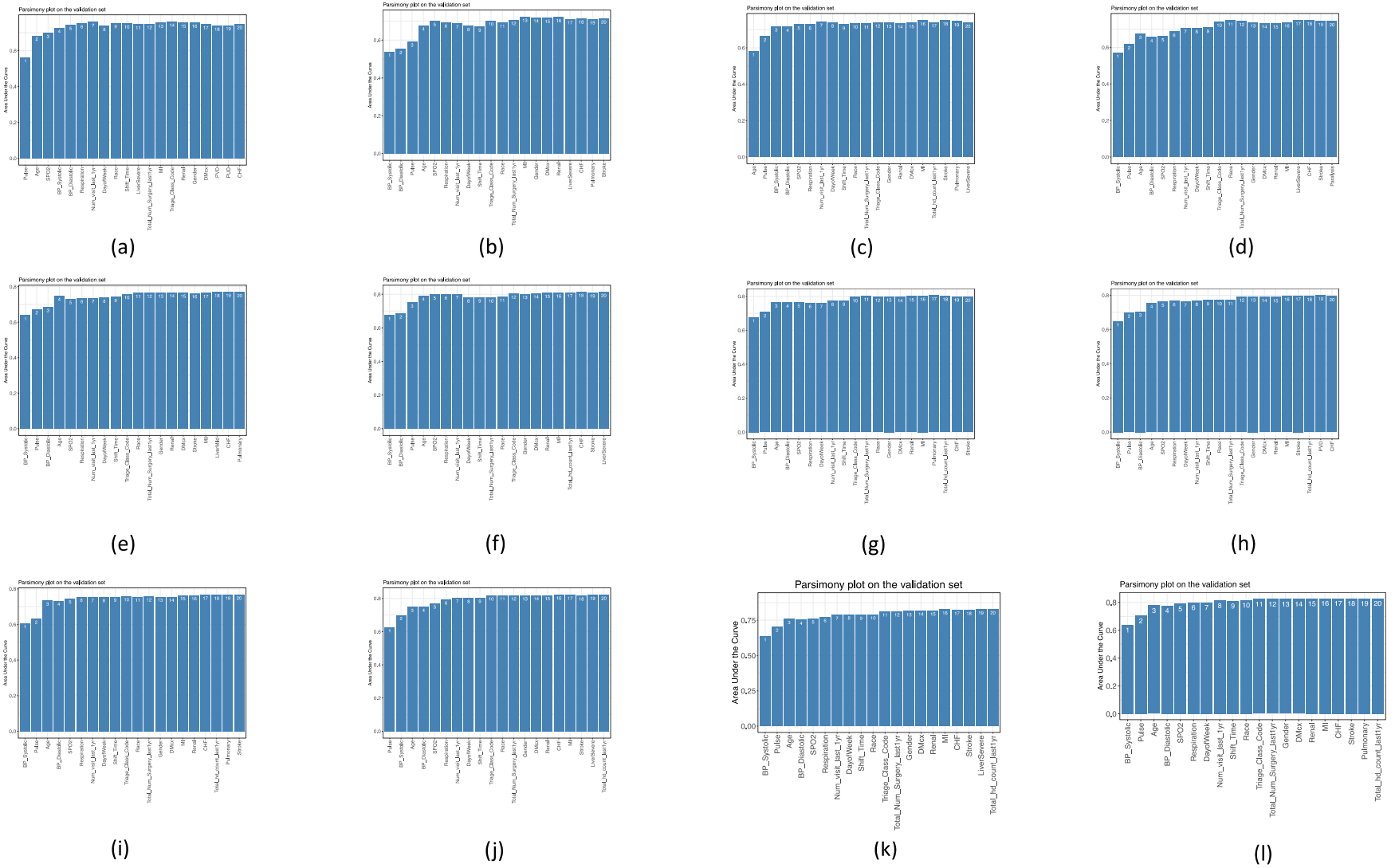
| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <70 | 13 |
| | [70,100) | 15 |
| | [100,118) | 19 |
| | >=118 | 21 |
| 收缩压（毫米汞柱） | <98 | 13 |
| | [98,114) | 8 |
| | [114,159) | 5 |
| | [159,187) | 2 |
| | >=187 | 0 |
| 年龄（年） | <28 | 13 |
| | [28,48) | 22 |
| | [48,79) | 31 |
| | [79,88) | 34 |
| | >=88 | 37 |
| 舒张压（毫米汞柱） | <52 | 13 |
| | [52,60) | 11 |
| | [60,83) | 10 |
| | >=83 | 12 |
| 血氧饱和度（%） | <93 | 13 |
| | [93,96) | 8 |
| | >=96 | 6 |
| 呼吸频率（每分钟） | <16 | 13 |
| | [16,17) | 16 |
| | [17,19) | 14 |
| | [19,22) | 16 |
| | >=22 | 18 |
| 过去一年急诊入院次数 | <1 | 13 |
| | [1,4) | 16 |
| | >=4 | 17 |
| 星期 | 星期五 | 13 |
| | 其他 | 12 |

(l) 通过 AutoScore 获得的合并模型。

| 变量 | 区间 | 点 |
|---|---|---|
| 脉搏率（每分钟） | <60 | 2 |
| | [60,71) | 0 |
| | [71,100) | 2 |
| | [100,119) | 9 |
| | >=119 | 13 |
| 收缩压（毫米汞柱） | <98 | 19 |
| | [98,114) | 14 |
| | [114,159) | 7 |
| | [159,188) | 1 |
| | >=188 | 0 |

| | | |
|---|---|---|
| 年龄（年） | <29 | 0 |
| | [29,49) | 16 |
| | [49,79) | 30 |
| | [79,88) | 34 |
| | >=88 | 37 |
| 舒张压（毫米汞柱） | <52 | 4 |
| | [52,61) | 1 |
| | [61,84) | 0 |
| | [84,97) | 1 |
| | >=97 | 2 |
| 血氧饱和度（%） | <94 | 9 |
| | [94,96) | 2 |
| | >=96 | 0 |
| 呼吸频率（每分钟） | <16 | 8 |
| | [16,17) | 0 |
| | [17,19) | 2 |
| | [19,22) | 5 |
| | >=22 | 11 |
| 过去一年急诊入院次数 | <2 | 0 |
| | [2,5) | 6 |
| | >=5 | 4 |
| 星期 | 星期五 | 0 |
| | 其他 | 1 |

图1




缩写：曲线下面积，曲线下面积；置信区间，置信区间；标准差，标准差；收缩压，收缩压；舒张压，舒张压；$SpO_2$, 通过脉搏血氧测定测得的血氧饱和度；患者病情严重程度分类量表，患者病情严重程度分类量表；急诊科，急诊科