arXiv:2301.11146v2 [stat.AP] 6 Mar 2024

重症监护获得性感染的两步可解释建模

G. Lancia[1]**,**乌得勒支大学数学研究所,
Budapestlaan 6，3584 CD，荷兰乌得勒支,
g.lancia@uu.nl

M. Varkila,Julius健康科学与初级保健中心，乌得勒支大学医学中心，荷兰；重症医学科，乌得勒支大学医学中心，荷兰

O. Cremer,Julius健康科学与初级保健中心，乌得勒支大学医学中心，荷兰；重症医学科，乌得勒支大学医学中心，荷兰

C. Spitoni,乌得勒支大学数学研究所，荷兰,
C.Spitoni@uu.nl

摘要

我们提出一种新方法，将高分辨率 纵向数据与生存模型的动态预测能力相结合。其目标有两个：在提升预测能力的同时，保持模型的可解释性。为超越人工神经网络的黑箱 范式，我们提出一种简约且稳健的半参数方法（即标志点竞争风险模型），将常规收集的低分辨率 数据与从卷积神经网络中提取的预测特征相结合；该网络是在高分辨率时变信息上训练的。随后，我们使用显著性图 分析并解释该模型额外的预测能力。为说明这一方法，我们聚焦于入住重症监护病房的患者所发生的医疗保健相关感染。

关键词 标志点方法 · 卷积神经网络 · 动态预测 · 重症监护病房获得性感染 ·显著性图。

[1]现任地址：热那亚大学数学系，Via Dodecaneso 35，16461,
意大利热那亚

1 引言

与更传统的生存模型相比，人工神经网络（ANN）是非常准确的预测工具（Topol，2019年；Zeng 等，2022年；Ivanov 等，2022年）。然而，它们常被视为黑箱，因为通常无法用闭式表达人工神经网络的预测与输入数据之间的联系。

人工神经网络模型确实很难解释，也难以确定哪些预测因子最为重要（May 等，2011）。相比之下，半参数风险率生存模型（Andersen 等，1993）是可解释模型的典型示例，其风险可以直接或间接衡量每个协变量对感兴趣结局的影响。

为了恰当地对生存过程的时间演变进行建模，将纵向信息（例如生物标志物、健康状况、临床测量）作为时变协变量纳入模型通常很有帮助。这些协变量通常是内部 协变量，要准确预测生存函数，还需进行额外建模（Cortese 和 Andersen，2010）。联合建模（JM）试图对纵向协变量和事件时间进行联合建模，因此可能是自然的选择（Proust-Lima 和 Taylor，2009；Rizopoulos，2011，2012）。尽管模型设定正确时，联合模型能够有效估计潜在参数，但它们容易受到纵向轨迹设定错误的影响（Ferrer 等，2019），且估计过程复杂。

基于这些原因，我们考虑采用标志时间法（LM）对感兴趣的结局进行动态预测（例如，重症监护病房获得性感染）。LM 是一种务实的方法，无需为纵向协变量指定模型，并且在纵向过程模型设定错误时仍具有稳健性（Van Houwelingen, 2007; van Houwelingen and Putter, 2011）。LM 的核心思想是选定一个时间点 s，称为标志时点。通过选取在 s 时仍处于风险中的研究对象（即在时间 s 左截断），并在时间$s+w$（时限时间）进行行政性右删失，即可构建标志点数据集。因此，对于时间依赖协变量 $Z(t)$，仅考虑其在 s 时的取值 $Z(s)$，这样便可采用标准方法分析所得的 LM 数据集： $Z(s)$实际上被视为时间恒定协变量。对于存在竞争事件的情况，LM 方法可推广至竞争风险模型（LM-CR），参见 Nicolaie 等人（2013）。

该手稿的创新之处在于，LM-CR 模型纳入了来自高分辨率电子健康记录（EHR）数据的时间依赖信息：重症监护病房（ICU）监护仪记录并每分钟采样的生命体征信号（即心率、平均动脉压、脉压、动脉血氧饱和度和呼吸频率）。我们使用一种用于预测标志时间 s之前信号中存在的模式的深度神经网络——卷积神经网络（CNN）——作为特征提取器，并将其纳入主要的 LM-CR 模型中。事实上，我们假设这些模式代表了低分辨率协变量中未包含的额外信息。

尽管 LM-CR 本身是一个可解释模型，但我们希望从患者医学状况的角度解释 CNN 评分所带来的额外预测能力。因此，我们研究了 CNN 执行的模式识别，并通过显著图顺序等价（SMOE）尺度（Mundhenk 等人，2019）使其具有可解释性；该算法描述了网络隐藏层中激活特征图的统计特征。借助 SMOE 尺度，我们能够将输入数据中对预测具有最高显著性的区域可视化。随后，我们提取具有最高累积显著性的信号子集，以对更有可能在即将到来的预测窗口内发生结局的患者进行数据驱动聚类。该方法为未来应用我们的方法提供了概念验证。

为阐释该方法，我们聚焦于入住重症监护病房（ICU）的患者发生的医疗保健相关感染；这些感染是发病和死亡的主要原因（Vincent 等人，2009；Maki 等人，2008）。因此，及早识别感染事件有助于医生预防和管理重症监护病房中的感染性并发症（Dantes 和 Epstein，2018）。此外，医院感染的动态预测是一项具有挑战性的建模任务。感染是否存在并不容易确定，感染发生的确切时间也无法直接观测。因此，能够预测即将发生感染的方法，或许能为从业者争取宝贵的提前时间，以便采取干预措施。

本文结构如下。第 2 节介绍数据并定义我们要预测的结局；第 3 节介绍两步建模方法；第 4 节说明卷积神经网络的设计、训练，以及风险评分的提取。第 5 节定义并拟合 LM-CR 模型，并纳入由卷积神经网络提取的风险评分 。最后，第 6 节基于对电子健康记录实例的 SMOE 量表分析开展数据驱动聚类。补充材料 文件提供了有关数据、卷积神经网络设计方案选择的更多信息，并对本文所用的 SMOE 量表作了更详细的说明。

2 数据

我们分析了脓毒症分子诊断与风险分层（MARS）队列的数据（Klouwenberg 等，2013）。我们选取了年龄为 >18 岁、住院时长为 >48 小时的患者；这些患者于 2011 年至 2018年间入住参与研究的某个研究中心的重症监护病房。此外，我们还使用了生命体征监护仪采集的高分辨率数据流；这些数据以 1 分钟的分辨率记录在医院信息系统中。

作为主要建模尝试的结局参数，我们采用了从预测时刻起24小时的时间窗口内首次发生疑似重症监护病房获得性感染（ICU-AI）的时间。感染发生时间根据以下两者中较早者确定：开始新的经验性抗菌治疗，或采集血液进行培养（之后也开始抗生素治疗）。因此，该数据集包含5075次重症监护病房入院，其中发生了871例首次ICU-AI。值得注意的是，在重症监护病房住院期间，ICU-AI的发病率相对稳定：入住重症监护病房后的前10天内，平均每天发生0.04（SE 0.01）例。入院后发生时间中位数为5.25天（IQR 3.80–9.45）。

我们依据文献综述、对临床重要性的先验共识以及研究人群中的患病率，从多个变量中筛选候选预测因子。这些协变量既包括反映基线感染风险的时间固定变量，也包括表示患者临床演变动态的时间依赖数据，例如实验室指标、生理反应和器官功能参数；参见补充材料第1节中的表1和表2。

3 两步建模策略

本节简要介绍我们提出的方法。为了充分利用所有纵向临床数据，并纳入时间分辨率不同的观测值，我们采用两步建模方法来设计模型。具体如下：

步骤 1：我们使用 CNN 研究电子健康记录数据的纵向演变。电子健康记录数据是重症监护病房监护仪记录的高频生命体征，采样频率为每分钟一次。

我们对 CNN 进行训练，以获得感染风险评分（或简称风险评分或 CNN 评分），并将其作为预测因子之一纳入步骤 2。感染风险评分旨在预测治疗期间任意时点发生感染事件的可能性。风险评分越高，近期发生感染事件的临床风险就越高。为便于使用，风险评分的取值范围为 0 到 1。需要特别强调的是，风险评分仅由处理电子健康记录中的信息得出。尽管风险评分的取值范围为 0 到 1，但它并不代表感染概率。从理论上讲，CNN 输出并非概率。有关此步骤的更多详情将在第 4 节中讨论

步骤 2：拟合 LM-CR 模型，并纳入所有解释变量，即基线协变量 （例如性别、年龄、ICU 入院类型和入院合并症）、低频预测因子 （例如意识评分、实验室测量结果和细菌定植），以及由卷积神经网络获得的风险评分 。该模型结合了两种模型：Landmark 方法和竞争风险模型。Landmark 模型使我们能够基于 ICU 住院期间某一先前时点的数据，在治疗过程中的任意时刻预测疑似感染事件的发生。竞争风险模型通过实现一个 Cox 比例风险模型来构建，该模型有两种失效原因：院内感染的发生，以及某一竞争事件的发生，例如患者死亡或从 ICU 出院。有关此步骤的更多数学细节和深入说明，请参见第 5 节

总之，我们使用电子健康记录数据训练了卷积神经网络，并评估了每位患者在重症监护病房住院期间的感染风险评分。该评分仅根据生命体征分析得出，可简明地反映即将发生感染事件的可能性。随后，我们将风险评分与其他可用的解释变量相结合，再使用这组全面的预测因子训练 LM-CR 模型，作为生成感染事件发生动态预测的主要工具。

4 步骤 1：卷积神经网络的运作

本节介绍卷积神经网络模型，包括其结构、所用数据以及训练和测试方法。此外，还简要说明了如何评估卷积神经网络生成的评分。

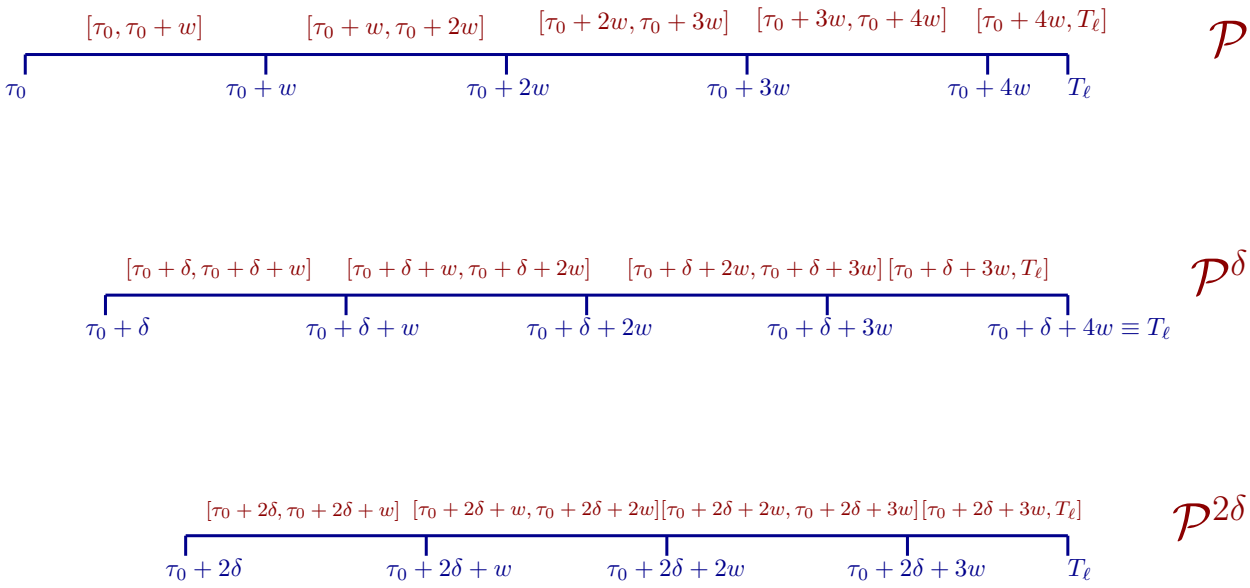


图1：为一名患者选取时间窗口的示例。 τ_0 表示患者的入院时间， w 表示预测窗口的长度。所有这些预测窗口构成的集合记为 $\mathcal{P}$。下图展示了以 δ为步长移动后选取的窗口；该窗口集合记为 $\mathcal{P}^\delta$。同样，图中也展示了以 2δ 为步长移动后选取的窗口

4.1 高频实例的选取与插补

我们所说的高频协变量，指的是我们可获取的五项高频生命体征，即心率（HR）、平均动脉血压（ABP）、脉压（PP）、功能性血氧饱和度（SaO2）和呼吸频率（RR）。如前所述，这些预测因子的采样频率为每分钟一次。这些数据被整理为多个24小时的时间序列（即每个时间序列包含1440条记录，每分钟一条）。

因此，我们按如下方式选取并提取时间序列实例：

1. 对于在重症监护病房住院期间死亡的患者，我们排除了其死亡前最后24小时的数据。此举旨在剔除可能包含非代表性信息的时间窗口，例如死亡前最后24小时内因医疗决策而停止治疗所导致的极端或异常记录。因此，使用这些记录可能会使卷积神经网络产生偏差，并增加分类任务的难度。

2. 从第 i名患者的入院时间 τ_0^i开始，我们将所有生理生命体征划分为宽度为 $w = 24$ 小时的时间窗口，直至患者记录的最终时间 T_ℓ^i（对于住院期间死亡的患者，该时间按第1点中的定义确定）。因此，我们获得了患者 i的区间集合$\mathcal{P}^i$：

$$\mathcal{P}_i := \bigcup_{k\geq 1} \left\{ [\tau_0^i + (k-1)w, \min(\tau_0^i + kw, T_\ell^i)] \right\}$$

同样地，我们将时间窗口集合定义为向后移动 δ：

$$\mathcal{P}_i^\delta := \bigcup_{k\geq 1} \left\{ [\tau_0^i + \delta + (k-1)w, \min(\tau_0^i + \delta + kw, T_\ell^i)] \right\},$$

前提是 $T_\ell^i \geq \tau_0^i + \delta$。因此，为患者 i选择的时间窗口是属于集合 $\mathcal{P}_i^{\text{total}}$ 的窗口：
$= \mathcal{P}_i \cup \mathcal{P}_i^{8\text{hrs}} \cup \mathcal{P}_i^{16\text{hrs}}$；见图1。 $\mathcal{P}_i^{\text{total}}$中的时间窗口集合（即连续的24小时时间窗口，以及将其平移8小时和16小时后得到的窗口）使我们能够以与第2步提取低频时变协变量相一致的方式，对信号的纵向演变进行连贯分段。我们将 $\mathcal{P}_i^{\text{total}}$中某个区间所对应的五种生命体征信号部分称为时间序列实例。

3. 对于每位 i 在重症监护病房住院期间未发生感染的患者，我们将其时间序列实例称为未感染实例。对于此类患者，我们考虑时间窗口位于 $\mathcal{P}_i^{\text{total}}$中的所有时间序列实例。

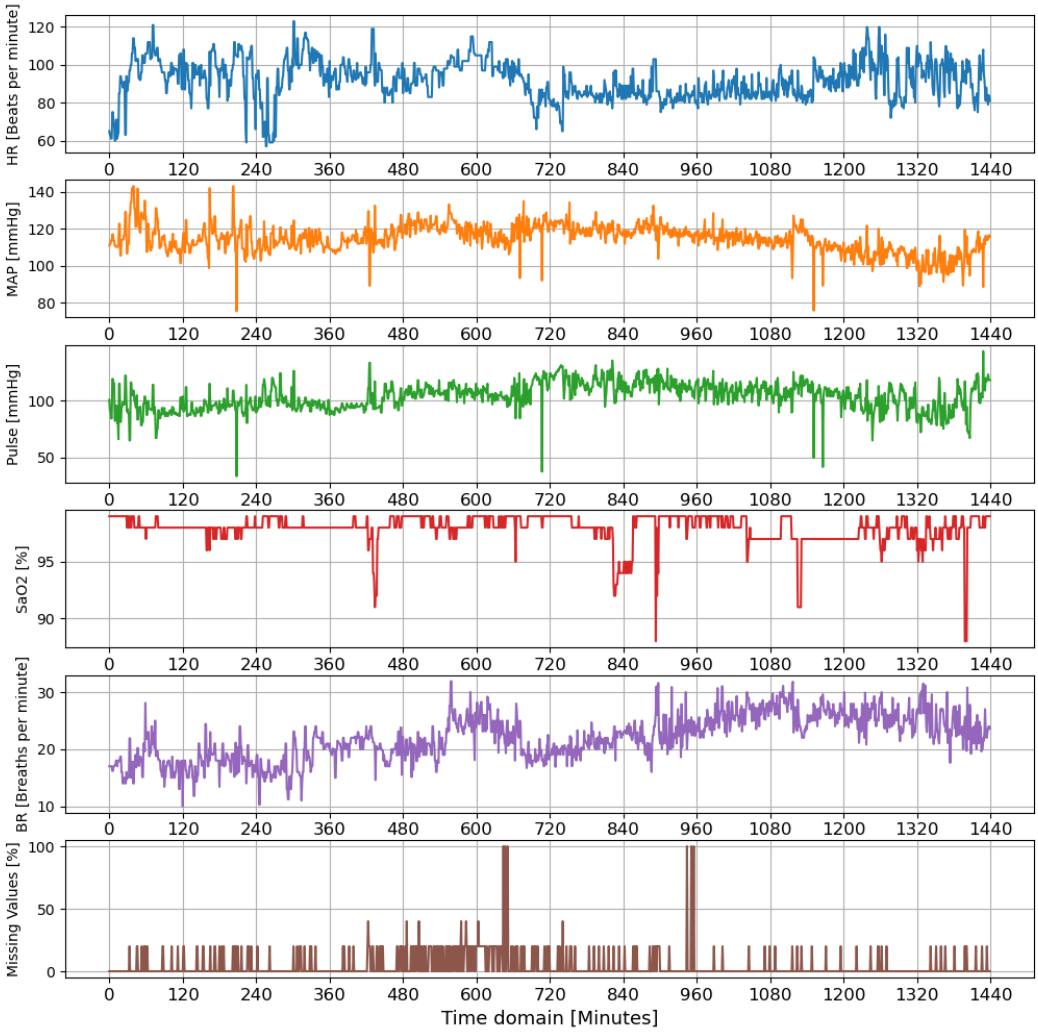


图 2：时间序列实例示例。x 轴：时间域（24 小时）。y 轴：各时间序列特征的取值。具体而言，蓝色表示心率，橙色表示动脉血压，绿色表示脉压，红色表示 SaO_2，紫色表示呼吸频率，棕色表示辅助时间序列（含缺失值发生率）。

4. 对于在重症监护病房住院期间 i 发生感染的每位患者，我们首先按照第2点（$\mathcal{P}_i^{\text{total}}$）中的方法划分其重症监护病房住院期间。接着，我们将发生ICU-AI事件的所有时间窗口标记为事件（即包含记录ICU-AI发作时间戳的时间窗口）。同样，我们还将包含ICU-AI事件起始时间的时间窗口之前的所有时间窗口标记为结局事件。这样，我们将能够预示ICU-AI发作被报告时刻之前最多24小时的时间序列实例标记为事件。此选择自然源于对所考察结局变量高度波动性的建模需求。与重症监护病房内的死亡等其他事件不同，除非临床状况恶化已十分明显，否则无法在某个确切时刻检测到获得性感染的发生。因此，其余未关联任何感染事件的时间窗口均视为未感染。

5. 我们只考虑患者首次 ICU-AI 发作，并舍弃同一患者的所有其他复发性发作。更确切地说，首次感染之后的所有实例均被舍弃。6. 我们为每个时间序列实例 配备了一个额外的时间序列，用于监测缺失值的存在情况：这一策略使我们能够追踪每个时间戳处缺失记录的百分比。

因此，每个 时间序列实例 均由一个 6×1440 矩阵描述，其行表示 时间序列特征 的类型（即，HR、动脉血压、脉压、血氧饱和度、呼吸频率和缺失记录），列表示时间域（注意，1440 对应一天中的记录总数，计算方式为 24 小时乘以每小时 60 分钟）。图 2 展示了一个样本时间序列实例 。

电子健康记录中的缺失值采用零阶样条进行插补，即使用最后一次观测值前向填充（LOCF）方法。尽管这种方法非常简单，但我们注意到它已应用于其他一些类似场景，例如 Gandin等（2021）和 Deng等（2023）。然而，在我们的案例中，通过为每个时间序列实例额外加入一个时间序列，以报告缺失的生命体征及其缺失区间，这种插补方法的简单性得到了弥补。这一策略有助于卷积神经网络模型更准确地识别经由前几层传递的模式所包含的信息。一般而言，第一卷积层负责处理完整生命体征的特征；此时，额外的时间序列不发挥作用，取值为零。然而，在处理缺失信号时，对额外时间序列的处理相当于加入一个额外项，用于重新调整激活函数的自变量。这种调整由一些适当的权重决定，这些权重在学习阶段不断优化。

除了理论考量外，选择 LOCF 插补方法也受一项涉及两种替代方法的比较分析所启发。第一种方法是多变量 k 近邻（k 最近邻）（Troyanskaya et al., 2001），原本用于矩阵数据，在本研究中则经过调整，以适用于时间序列实例。另一种方法下文称为*ICUAI*插补 方法，旨在更好地契合重症监护病房情境中缺失值的内在特点。具体而言，对于幅度超过 4 小时的缺失区间，以一个超出取值范围的常数（例如 100）替代；对于短于 4 小时的区间，则以空值常数进行插补。ICUAI 插补方法的设计借鉴了管理电子健康记录（EHR）的实际医疗经验。从实际情况来看，约 4 小时或更长的区间通常对应外科手术的持续时间，而较短的区间往往与重症监护室监测暂时中断有关，原因可能是患者无意间脱开设备，或设备错位。通过这种方法，我们系统地使用指定的占位值填补特定类型的缺失区间；如此一来，插补后的缺失区间模式还可以包含重症监护病房住院期间发生的、值得关注的不同临床事件。比较采用这些不同插补方法的卷积神经网络模型的表现后发现，LOCF 的表现最佳。更多相关分析见补充材料第 2 节。

将生命体征输入卷积神经网络模型之前，我们对其进行了预处理。具体而言，我们对所有时间序列特征应用了同一个线性变换，将其压缩到 [-1, 1] 范围内。我们为所有同类时间序列特征设计并应用了线性变换。因此，综合考虑可用生命体征的总体统计特性，我们为每个时间序列特征设计了线性映射，将其重新缩放至 [-1, 1] 范围内。有关预处理的更多信息，请参阅补充材料第 2 节。
需要指出的是，为了说明我们的方法，我们主要选择聚焦于 24 小时时间窗口。我们也使用 48 小时时间窗口重复了分析（做法见补充材料第 2 节）。然而，预测窗口越大，输入数据的维度也越高。

4.2 卷积神经网络的设计

过去十年表明，卷积神经网络（CNN）的预测能力在多种情境下解决各类任务时取得了巨大成功，例如图像识别（Liu，2018年；Zheng 等，2017；Lou 和 Shi，2020；Kagaya

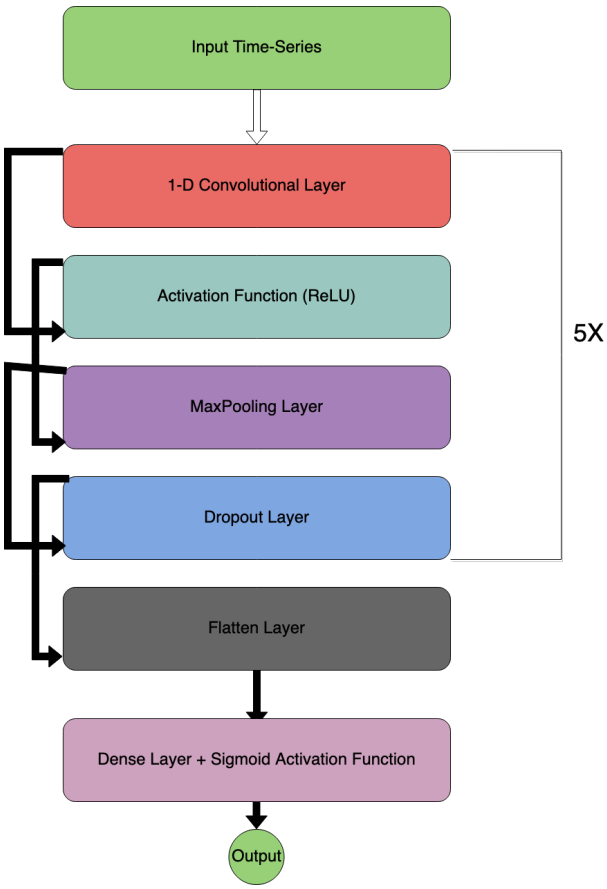


图 3：卷积神经网络模型示意图。输入信号由卷积层（128 个大小为 3 的滤波器）处理。在*ReLU* 函数之后应用最大池化运算，以缩小特征的尺寸。每个最大池化 层之后都接一个丢弃层 ，其丢弃率为 0.25。此隐藏层序列重复五次。随后，特征图被展平为一个数组（展平层），再通过一个全连接 层（稠密层），并采用 Sigmoid 激活函数。

等，2014）、异常检测（Kwon 等，2018年；Naseer 等，2018；Staar 等，2019年）和时间序列预测（Borovykh 等，2017；Selvin 等，2017；Livieris 等，2020；Guo-yan 等，2019）等。事实上，这类人工神经网络（ANN）专为处理网格结构数据而设计。它们在处理复杂多层级数据方面的强大能力，主要归功于卷积算子和最大池化算子的结合；这些算子能够对输入数据的多维域中模式的序列性进行编码。例如，在一维 CNN 中，对卷积层的卷积滤波器进行最优搜索，旨在获得输入时间序列实例最线性化的潜在表示。因此，为了评估电子健康记录（EHR）中感染的风险评分，我们选择使用卷积网络：其架构仅由卷积层、池化层和全连接层组成。选择 CNN 似乎顺理成章，因为我们要寻找的是可能出现在时间序列任意子区间中的平移不变模式。

为对这一推理提供定量依据，我们将 CNN 的准确率与其他传统机器学习模型和基于神经网络的模型进行了比较，包括逻辑回归（LR）、线性支持向量机（SVM）、多层感知机（MLP）以及 CNN-LSTM 网络（其中 LSTM 指长短期记忆）。具体而言，我们在一组精细划分的超参数网格上训练并验证了上述模型。此处说明，模型性能通过受试者工作特征曲线下面积（AUROC）进行评估，或更简单地说，使用 AUROC 分数进行评估。再次证明，这项指标适合用于衡量模型性能。各模型的 AUROC 分数列于表1。不过，需要指出的是，为确保候选模型能够稳健地考察不同关注时间尺度上的纵向变化，我们使用24小时和48小时预测窗口评估了它们的性能（即表1中的24小时实例模型和48小时实例模型。对于48小时实例，我们通过调整第4.1节的策略进行了选择）。表1的结果显示，CNN模型并非预测准确率绝对最高的模型

。不过，我们之所以选择它，是因为在同时考虑24小时和48小时实例时，它能够最好地模拟感染风险评分。尽管24小时 CNN-LSTM 模型的准确率可能略高于 CNN，但我们观察到，后者即使在48小时序列实例中也表现出更精确的预测性能。对于24小时模型，两种模型的 AUROC 差异极小；而对于48小时模型，这一差异则更加明显。最终，CNN模型在这两种情境下均展现出更强的能力，能够捕捉相关模式。

更多详情　　关于模型选择策略的详细信息，请参阅补充材料第 2 节　　.

模型	曲线下面积（24小时模型）	曲线下面积（48小时模型）
LR	0.59 ± 0.01	0.59 ± 0.01
SVM	0.57 ± 0.01	0.57 ±0.01
MLP	0.63 ± 0.01	0.61 ± 0.01
CNN	0.72 ± 0.01	0.68 ± 0.02
卷积神经网络-长短期记忆网络	0.74 ± 0.01	0.59 ± 0.01

表1：模型选择摘要：报告了各个受研究模型在验证阶段达到的最高性能，以曲线下面积衡量。显示曲线下面积分数的列分别对应 24 小时实例模型或 48 小时实例模型。曲线下面积分数已四舍五入至小数点后两位。误差使用标准误均值进行评估；若误差过小，则以最小误差 0.01 替代。

我们选择了卷积神经网络设计，因为它具有较高的准确率，并且可以进行第 6 节所述的显著性图分析。

最终选定的卷积神经网络架构如下：

1. 卷积层：每层的滤波器数量为 128，每个滤波器的大小为 3（像素）。我们将滤波器作用于前一层后得到的输出称为特征图。2. 激活层：每个卷积算子之后都会应用 ReLU 函数（即 $\text{ReLU}(x) := \max(0, x)$）。在特征图上应用非线性激活函数后，便得到激活特征图。3. 最大池化层：通过池化大小为 2 的最大池化算子对激活特征图进行重采样（子采样）。

此外，每个最大池化层之后都会添加一个丢弃层，丢弃率为 0.25。这一系列隐藏层重复五次。最后一个特征图会被展平为一个数组，然后传入具有 Sigmoid 激活函数的全连接层（稠密层）。该激活函数会返回一个介于 0 和 1 之间的正数，即风险评分。图 3 示意了所选卷积神经网络的架构。

4.3 卷积神经网络的训练与总体评估

在使用电子健康记录数据训练模型之前，我们从可用的时间序列实例中进行了筛选。实际上，我们选择对时间序列实例的总量进行欠采样。此选择有两个原因：首先，我们主要希望避免大多数风险评分感染与卷积神经网络在训练或验证阶段获得的输出相重合。换言之，我们旨在将一个源自此前从未通过卷积神经网络传播的时间序列实例的风险评分纳入 Landmark 模型。第二个原因仅与卷积神经网络在类别高度不平衡的数据集上进行充分训练的内在困难有关。病例组中的时间序列实例数量（即代表 ICU-AI 事件的实例）不到对照组时间序列实例总量的二十分之一（即不代表 ICU-AI 事件的实例）。因此，我们选择在由时间序列实例 组成的群体上拟合卷积神经网络模型，使对照组与病例组的比例为 8:1（即对照组中的时间序列实例数量是病例组的 8 倍）。需要特别指出的是，我们仅对对照组应用了随机欠采样。

模型拟合旨在通过 ADAM 算法（Kingma 和 Ba，2015）优化二元交叉熵损失函数。因此，我们训练卷积神经网络来解决二分类任务。Deep-LM-CR 与 LM-CR 模型之间的曲线下面积差值（即分别包含和不包含 CNN 评分的两个 Landmark 模型；参见第 5.2 节）用于评估这两个模型的相对优劣。此外，我们还将 Brier 分数（Brier，1950）作为评估模型预测能力的另一项指标，并使用 Brier 技巧评分（Wilks，2011）评估 Deep-LM-CR 相对于 LM-CR 的预测性能相对增幅

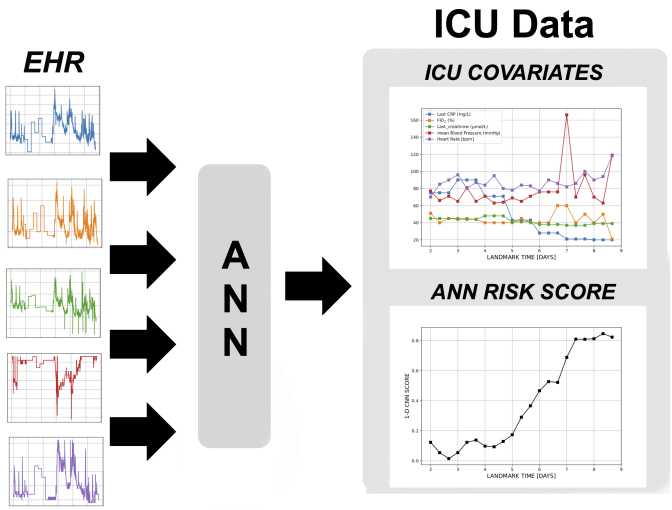


图4：将基于 CNN 的风险评分 $Z_{\text{CNN}}(t_{LM})$ 纳入 ICU队列数据的示意图。

。虽然我们主要关注的并非卷积神经网络本身的预测结果，但也需要确保卷积神经网络模型能够对时间序列实例 进行分类，并编码有关 ICU-AI 即将发生的信息模式。我们采用五折交叉验证 进行内部验证；在验证卷积神经网络模型作为二分类器的性能时，数据被划分为 5 个不同的折。随后，整体曲线下面积取这 5 折的平均值。我们通过在由超参数构成的精细网格上验证模型，寻找最优配置。所采用的验证策略同样是五折交叉验证。更多细节见补充材料 第 2 节，其中我们深入探讨了三个关键超参数对卷积神经网络模型曲线下面积的影响

4.4 CNN风险评分

CNN评分的提取及其纳入 LM-CR 模型是本文的新颖之处。感染风险评分通过 CNN 进行评估，其架构已在第4.2节中讨论，训练阶段则在第 4.3 节中讨论。

因此，评估风险评分的步骤如下：

1. 考虑患者的生命体征 i （HR、动脉血压、脉压、血氧饱和度和呼吸频率）以及标记缺失记录的时间序列。2. 从重症监护病房入院时间开始，通过 8小时滑动时间窗提取 24小时时间序列实例（见第4.1节），分别对应于以下时间区间 $\mathcal{P}_i$。3. 将时间序列实例 输入拟合后的卷积神经网络模型的隐藏层，并评估风险评分。4. 将风险评分分配给相应的时间戳（即日-月-小时-分钟）。

我们如何将风险评分纳入 ICU预测变量的方案如图 4 所示：对于单个患者，在每个 LM 时间 t_{LM} 计算评分。在每个 t_{LM}，还会报告其他时变协变量的数值（例如，C反应蛋白、FiO_2,肌酐水平、平均血压、平均心率）。

5 第二步：深度LM-CR 模型

5.1 符号与 LM-CR 模型

In t本节将按照 Nicolaie 等人（2013）使用的符号介绍 LM模型).

˜我们考虑一个由 N 名研究对象组成的队列，并用 T 表示失效时间，用 C 表示删失时间，用 D表示失效原因，以及用$\mathbf{Z}(\cdot)$和一个协变量数组表示。在一般框架下，一个受试者只能经历一种

J相互排斥的竞争失效原因之一；发生时， D取值为 {1,.....、J}，分别对应当前考察的失效原因。或者，当尚未发生任何失效原因时， D 始终取值为0。对于第i名受试者，元组(T_i， Δ_i，$\mathbf{Z}_i(\cdot)$)分别表示观测时间 $T_i = \min$（T_i， C_i）（即失效时间与删失时间中的较早者）、失效原因 $\Delta_i = \mathbf{1}$（$T_i < C_i$）D_i （其中$\mathbf{1}(\cdot)$为指示函数），以及截至时间 T_i的协变量$\mathbf{Z}_i(\cdot)$。请注意， $\Delta_i = 0$ 表示患者尚未经历任何失效原因；其临床病史已被删失。我们将采用下标 j 指代竞争失效原因，其中 $j \in \{1, \ldots, J\}$。

我们希望推导出在某个时限时间（t_{hor}）内、原因 j 导致失效时间的概率分布函数的动态预测，该预测以受试者存活且未发生事件为条件，并基于固定时间 t_{LM}（标志时间）可获得的信息。具体而言，给定预测窗口 w （使得 $t_{hor} = t_{LM} + w$），我们希望估计生存概率以及原因 j的累积发生函数（CIF）：

$$S_{LM}(t_{hor}|\mathbf{Z}(t_{LM}), t_{LM}) := \mathbb{P}(T > t_{hor}|\mathbf{Z}(t_{LM}), t_{LM}), \tag{1}$$

$$F_{j,LM}(t_{hor}|\mathbf{Z}(t_{LM}), t_{LM}) := \mathbb{P}(T \leq t_{hor}, \Delta = j|\mathbf{Z}(t_{LM}), t_{LM}). \tag{2}$$

LM方法包含两个步骤:

1. 首先，我们将观测的时间域 $[s_0, s_1]$ 划分为 n 个等间距标志点，记为$\{t_{LM}^k\}_{k=1}^n$，其中 $t_{LM}^1 \equiv s_0$ 且 $t_{LM}^n \equiv s_1$。我们固定预测窗口的宽度 w （即提前时间），然后针对每个LM时间 t_{LM}^k，通过选取在时间 t_{LM}^k仍处于风险中的所有研究对象，并在时间 $t_{LM}^k + w$ （时限时间）进行行政性右删失，来创建一个数据集。因此，对于时变协变量向量$\mathbf{Z}(t)$，在第 k个数据集中仅考虑$\mathbf{Z}(t_{LM}^k)$在 t_{LM}^k 时的取值。最后，我们将每个标志时间 t_{LM}^k 提取的所有数据集堆叠起来，创建一个大型数据集（LM超级数据集）。

2. 第二步是对堆叠后的 LM超级数据集拟合*LM-CR* 超级模型（Nicolaie 等人，2013）。由于在每个 t_{LM}^k 时，向量 $\mathbf{Z}(t_{LM}^k)$ 都被视为随时间恒定的协变量向量，因此可以使用标准生存分析方法分析该数据集。

在*LM-CR* 超级模型中，我们实际上对原因特异风险 λ_j拟合了一个 Cox 比例风险模型:

$$\lambda_j(t|t_{LM}, \mathbf{Z}(t_{LM})) = \lambda_{0j}(t|t_{LM}) \exp[\beta_j^T(t_{LM})\mathbf{Z}(t_{LM})], \tag{3}$$

其中， $\lambda_{0j}(t|t_{LM})$ 表示（未指定的）基线风险， $\beta_j(t_{LM})$ 表示区间 $[t_{LM}, t_{LM} + w]$ 内第 j 种原因对应的一组回归变量。我们假设系数 β 以平滑的方式依赖于 t_{LM}，即 $\beta_j(t_{LM}) = f_j(t_{LM}, \beta_j^{(0)})$，其中 $\beta_j^{(0)}$ 是回归参数向量， $f_\beta(\cdot)$ 是时间的参数函数，例如样条函数。我们选择使用二次函数:

$$\beta_j(t_{LM}) := \beta_j^{(0)} + \beta_j^{(1)} t_{LM} + \beta_j^{(2)} t_{LM}^2.$$

$\lambda_{0j}(t|t_{LM})$ 的估计可通过 Breslow 型估计量获得；我们可以将这种依赖关系建模为

$$\lambda(t|t_{LM})_{0j} = \lambda_{0j}(t) \exp(\gamma_j(t_{LM})). \tag{4}$$

对于系数 β，我们假设式（4）的系数 γ 参数化地依赖于标志时间，例如通过二次样条

$$\gamma_j(t_{LM}) := \gamma_j^{(0)} + \gamma_j^{(1)} t_{LM} + \gamma_j^{(2)} t_{LM}^2.$$

使用 Breslow 偏似然处理并列观测来拟合该模型，等价于最大化伪偏对数似然，如 Nicolaie 等人，2013 中所示。随后，可直接对堆叠数据集应用简单的 Cox 模型来拟合标志点超模型。因此，在估计系数和原因特异性基线风险后，我们得到生存概率（即 $S_{LM}(t_{hor}|\boldsymbol{Z}(t_{LM})\ t_{LM})$ ）以及原因 j 的累积发生函数（即 $\hat{}$ $F_{j,LM}(t_{hor}|\boldsymbol{Z}(t_{LM})\ t_{LM})$ ）的代入估计量。这些估计量的显式形式如下:

$$\hat{S}_{LM}(t_{hor}|\boldsymbol{Z}(t_{LM}), t_{LM}) = \exp\left(-\sum_{j=1}^{J} \exp\left(\boldsymbol{Z}(t_{LM})\hat{\beta}_j(t_{LM}) + \hat{\gamma}_j(t_{LM})\right)\left[\hat{\Lambda}_{0j}(t_{hor}) - \hat{\Lambda}_{0j}(t_{LM})\right]\right), \tag{5}$$

and

$$\hat{F}_{j,LM}(t_{hor}|Z(t_{LM}), t_{LM}) = \sum_{t_{LM} < t_i \leq t_{hor}} \hat{\lambda}_{0j}(t_i|\boldsymbol{Z}(t_{LM}))\hat{S}_{LM}(t_{hor}|\boldsymbol{Z}(t_{LM}), t_{LM}). \tag{6}$$

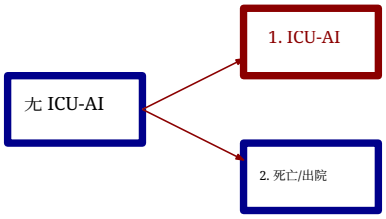


图5：ICU-AI 的竞争风险模型。

该（*6*）的估计原因特异性基线为

$$\hat{\lambda}_{0j}(t_i) = \frac{\#(t_{LM} \le t_i \le t_{hor}, \Delta_i = j)}{\sum_{t_{LM}:t_{LM}\le t_i\le t_{hor}} \sum_{t_k:t_{LM}\le t_i\le t_k\le t_{hor}} \exp\left[\boldsymbol{Z}_k(t_{LM})^T \hat{\beta}_j(t_{LM}) + \gamma_j(t_{LM})\right]}, \tag{7}$$

而估计的原因特异性累积基线很简单

$$\hat{\Lambda}_{0j}(t) = \sum_{t_i \le t} \hat{\lambda}_{0j}(t_i).$$

5.2 LM-CR for ICU-AI

在 ICU-AI 的动态预测背景下，我们采用了一个包含三种失效原因的竞争风险模型：*ICU-AI*、重症监护病房内死亡 和出院；参见图5。数据中不存在右删失，因为没有患者在出院或死亡前离开重症监护病房。

˜按照第5.1节使用的记号，我们用 T 表示失效时间，用 D 表示失效原因（即， $D = 1$表示 ICU-AI，而 $D = 2$ 表示出院或死亡），并用$\mathbf{Z}(\cdot)$表示协变量数组。对于第 i个受试者，三元组˜$(T_i, \Delta_i, \mathbf{Z}_i(\cdot))$表示观测时间 $T_i \equiv T_i$、失效原因 $\Delta_i \equiv D_i$以及协变量向量$\mathbf{Z}_i(\cdot)$。
本文中，预测窗口设为 $w = 24$ 小时。时间域为 $[s_0, s_1]$，其中包含 $s_0 = 48$小时和 $s_1 = 240$ 小时；我们考虑 $n = 25$ 个 LM 时点 t_{LM}，即相邻两个 LM 时点之间相隔 8 小时。

如果用 $Z_{\text{CNN}}(t)$ 表示时间 t 的 CNN风险评分（见图 4），用 $\mathbf{Z}(t)$ 表示时间 t 时 LM-CR 模型中所有其他协变量组成的向量，我们关注两个模型的动态预测：

1. $\pi_1 := F_{1,\ LM}(t_{hor}|\mathbf{Z}(t_{LM})\ t_{LM})$：即在存活至时间 t_{LM} 的条件下，并基于<当时的低频协变量（ LM-CR 模型）得到的感染累积发生函数；2. $\pi_2 := F_{1,\ LM}(t_{hor}|\mathbf{Z}(t_{LM})\ Z_{\text{CNN}}(t_{LM})\ t_{LM})$：即在存活至时间 t_{LM} 的条件下，同时基于<低频 协变量和 Z_{CNN} （Deep-LM-CR 模型）得到的感染累积发生函数。

通过比较 π_1 和 π_2 的准确性，我们可以衡量 CNN评分带来的额外预测能力。我们将第一个模型称为 LM-CR，将第二个模型称为 Deep-LM-CR。

5.3 LM-CR模型的评估

我们使用AUROC指标评估每个单独标志时间上的预测结果。若要考虑总体测量指标，评估全局AUROC时需考虑动态过程的时变特征。与Spitoni等人（2018年）提出的预测误差估计量类似，评估总体AUROC时需要考虑风险集大小随时间的变化。由于不存在删失，我们可以简单地估计总体AUROC评分：

$$\text{AUROC}_{\text{global}} = \frac{\sum_{k=1}^{n} R(t_{LM}^k)\,\text{AUROC}(t_{LM}^k)}{\sum_{k=1}^{n} R(t_{LM}^k)}, \tag{8}$$

其中， t_{LM}^k 为第 k个标志时间， n 为标志时间总数， R（t_{LM}^k）为时间t_{LM}^k的风险集大小。同样，我们估计了总体 Brier 分数：

$$\text{BS}_{\text{global}} = \frac{\sum_{k=1}^{n} R(t_{LM}^k)\,\text{BS}(t_{LM}^k)}{\sum_{l=1}^{n} R(t_{LM}^l)}. \tag{9}$$

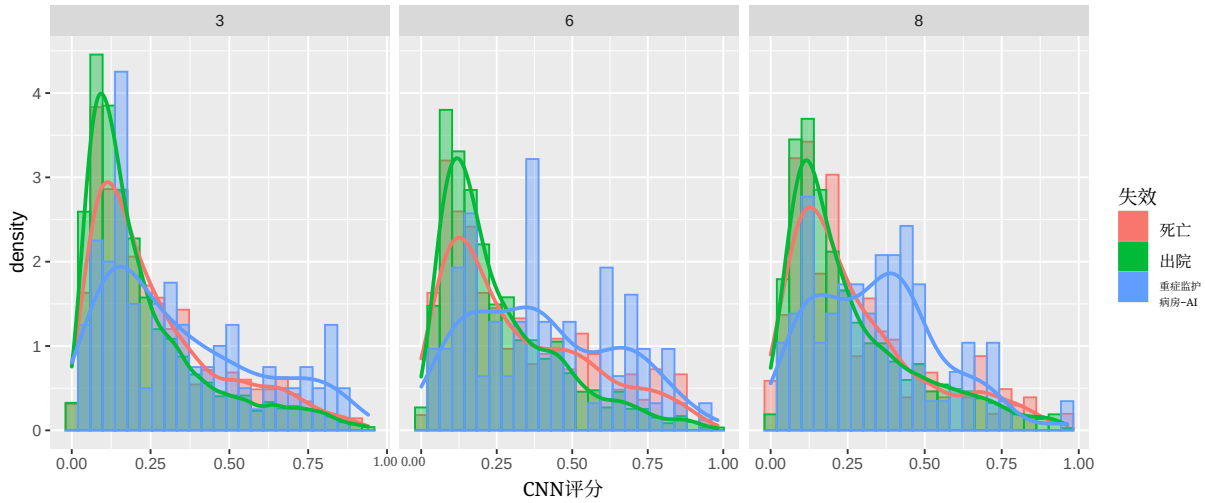


图 6：按失效原因分层的三个不同标志点（$t_{LM}^{k} \in \{3,6,8\}$ 天）处的CNN风险评分分布。

我们通过热图直观呈现各预测因子对预测结果的影响。我们计算包含所有预测因子的模型与移除某一预测因子的模型之间总体曲线下面积的相对变化。因此，我们构建了一幅热图，展示在标志时间 t_{LM}移除单个预测因子所导致的曲线下面积相对变化。

最后，我们指出，内部验证采用了*10*折交叉验证方法。总体曲线下面积$_{\text{global}}$ 以及在各时间点 t_{LM}^{k}评估的曲线下面积（t_{LM}^{k}），均为10折的平均值。对于CR-LM模型和Deep-CR-LM模型，我们报告了95%自助法置信区间。

5.4 结果

本节将展示CNN风险评分 Z_{CNN} 如何为模型提供额外的预测信息，而这些信息并未包含在标准协变量中。

在图 6 中，我们绘制了三个标志点（即 $t_{LM} \in \{3,6,8\}$）处 Z_{CNN}（t_{LM}）的经验分布，并按失效原因进行了分层。正如预期，感染患者的 Z_{CNN} 分布右偏程度更高：第三天这一现象较轻微，而第6天和第8天，密度分布的偏度则明显得多。

在图7中，我们报告了 CNN风险评分与标志时间之前各个24小时时间窗口内生命体征信号平均值之间的皮尔逊相关系数（即 LM-CR 中包含的时变协变量）。虽然风险评分是相对于这些信号进行评估的，但相关性仅为轻度。我们的主要假设确实是，$Z_{CNN}(t_{LM})$包含其他协变量 $\mathbf{Z}(t_{LM})$所不具备的增量预测信息。

此外，对于感染的原因特异性风险而言，CNN风险评分最终成为最重要的预测因子：$\beta_1^{(0)}{}_{;\text{CNN}} = 4.8$（95%置信区间 3.05-6.72）。ICU-AI 的所有原因特异性风险完整列表见补充材料中的表3。

LM方法为 ICU-AI 累积发生函数的动态预测（2）提供了一种插件式估计量。为举例说明该模型所支持的动态预测，我们在图8中报告了 LM-CR 和 Deep-LM-CR 模型的累积发生函数，这些函数同时以标志时间和拟合线性预测因子的分位数组为自变量。给定标志时间 t_{LM}处的协变量值，任意 s下且 $s \in [t_{TM}, t_{LM}+w]$时的累积发生函数确实由式（2）的插件式估计量 $F_{1,\ LM}(s|\boldsymbol{Z}(t_{LM})\ t_{LM})$ 给出。

图8中的红色虚线表示感染累积发生函数的任意警戒水平（例如，8%）。我们可以看到，对于第四分位数 Q_4，在 LM 时间 $t_{LM} = 4$ 天时，Deep-LM-CR模型比 LM-CR 模型提前约 3 小时达到警戒阈值。

LM-CR 模型的整体曲线下面积为$_{\text{global}} = 0.69$（95%置信区间 0.68-0.70），而 Deep-LM-CR 的整体曲线下面积为$_{\text{global}} = 0.75$（95%置信区间 0.73-0.76）。在每个时间点 t_{LM}^{k} 评估的曲线下面积（t_{LM}^{k}）得分，$k \in$

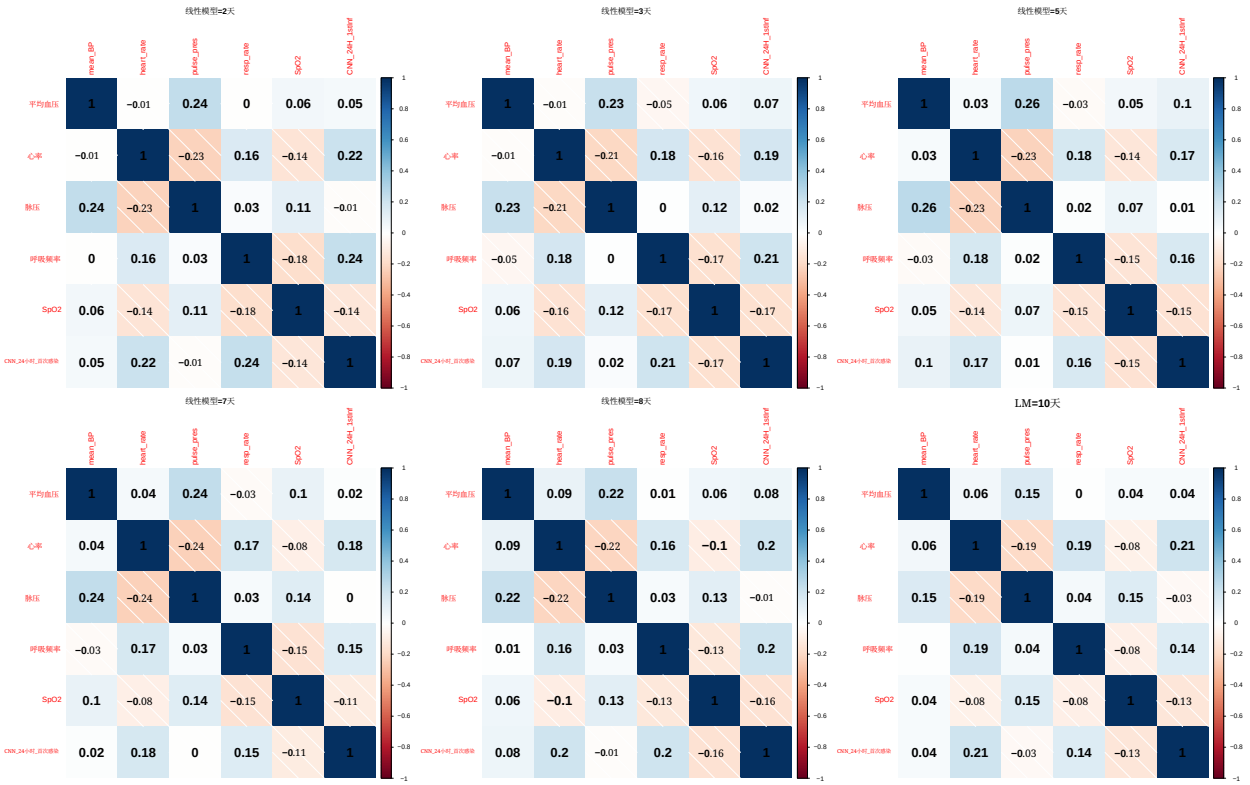


图7：相关　　性图：CNN风险评分与生命体征信号的关系（取标志时点前24小时的平均值）　　.

$\{1, \ldots, n\}$ 如图9所示。LM-CR 模型的预测性能始终低于 Deep-LM-CR。我们注意到，在重症监护病房住院期间的早期阶段（例如第3.66天）和第7天左右，卷积神经网络可将传统 ICU临床协变量的预测性能提高约 $8-9\%$，见图10。

Brier分数评估结果显示，LM-CR模型的总体布里尔分数为 $\text{BS}_{global} = 0.037$ （95%置信区间 0.036-0.039），而 Deep-LM-CR 的布里尔分数为 $\text{BS}_{global} = 0.036$ （95%置信区间 0.035-0.037）。各标志时间的布里尔分数 $\text{BS}(t_{LM}^k)$见图11。对于大多数标志时间，我们观察到 Deep-LM-CR 的准确度略高于另一模型；相比之下，LM-CR 在第 2 天和第 4 天的少数标志时间，以及第9天和第10天前后，表现得稍微更精确。图 12 也展示了这一结果，其中给出了布里尔技巧评分。最终，我们观察到总体布里尔技巧评分为 0.03，95%置信区间为 (0.01, 0.07)。所有置信区间均通过自助重采样计算（自助抽样总体为1000个样本）。

Deep-LM-CR模型中各解释变量 Z_j 的影响如图13所示。图中展示了去除协变量 Z_j 的Deep-LM-CR模型与完整模型（包含$\mathbf{Z}(t_{LM})$和 $Z_{CNN}(t_{LM})$）之间AUROC相对增幅的热图。当 $Z_j = Z_{CNN}$时，我们观察到AUROC的相对增幅至少为4%。

总之，我们还考察了LM-CR和Deep-LM-CR模型的整体性能，并采用了振幅更大的16小时和24小时滑动窗口。对于16小时滑动窗口，LM-CR的整体曲线下面积$_{\text{global}}$ 为0.70（95%置信区间0.69-0.71），而Deep-LM-CR的曲线下面积$_{\text{global}}$ 更高，为0.73（95%置信区间0.72-0.74）。同样，对于24小时滑动窗口，LM-CR的整体曲线下面积$_{\text{global}}$ 为0.69（95%置信区间0.68-0.70），而Deep-LM-CR的曲线下面积$_{\text{global}}$ 相近，为0.73（95%置信区间0.72-0.74）。将这些结果与采用8小时滑动窗口的模型结果进行比较后发现，后者更值得采用，因为整合风险评分时其曲线下面积最高。值得注意的是，8小时采样频率是低频ICU协变量中的最高采样频率。因此，本分析仅限于采样周期等于或大于8小时的数据集。

总而言之，我们已经证明，两步建模能够有效提高预测的准确率。额外的预测能力来自纳入了基于 CNN 的风险评分，它是一个摘要

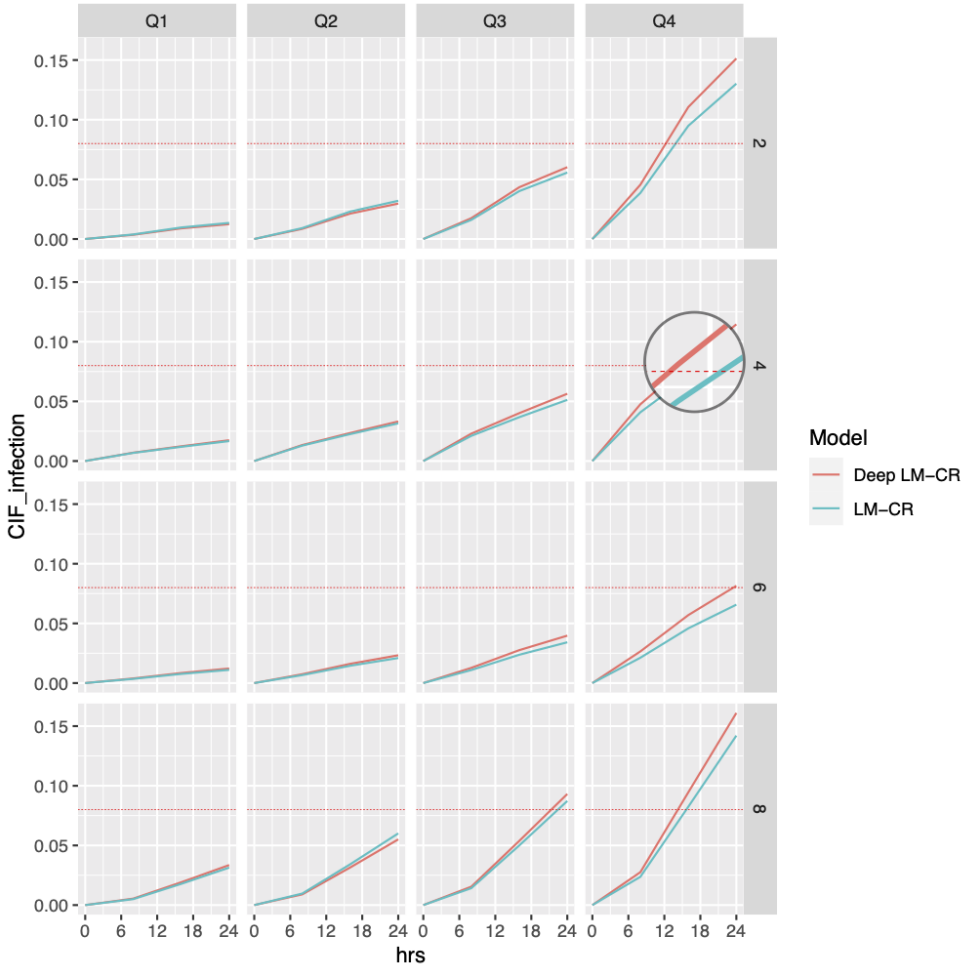


图8：模型 LM-CR 和 Deep-LM-CR 在不同标志时间（即 $t_{LM}^{k} \in \{2,4,6,8\}$ 天）的累积发生函数比较。

性度量，用于概括由仅使用五种生命体征信号（采样频率为 1 分钟）的数据训练的卷积神经网络模型所发现的预测模式。

需要指出的是，在我们的分析中未考虑复发性感染，而是仅关注 ICU-AI 的首次感染事件。

5.5 与完整人工神经网络模型的比较

为完整起见，我们希望将两步建模策略的预测结果与完整人工神经网络模型进行比较。具体而言，我们采用了双分支人工神经网络（TB-ANN），以便在一个人工神经网络模型中同时处理高频、低频和固定时间协变量。TB-ANN由两个不同的分支组成：第一个分支用于分析高频数据，另一个分支则同时分析低频数据和固定协变量。随后，两个分支连接起来，并经由预测层（即带有 Sigmoid 激活函数的稠密层）传播，输出与 CNN风险评分类似的分数。有关 TB-ANN 架构的更多细节，请参阅补充材料。因此，我们希望比较基于可解释量（如风险比）估计的模型（即 Deep LM-CR）与完全基于人工神经网络的模型（即 TB-ANN），以预测即将发生的感染事件。

与 Deep LM-CR 类似，我们在时间域$[s_0, s_1]$内以等间距标志时间训练 TB-ANN，其中 s_0 和 s_1 分别等于 48 和 240；任意两个相邻的标志时间相隔 8 小时。训练 TB-ANN 模型时，我们仅考虑每个标志时间可用的数据来预测

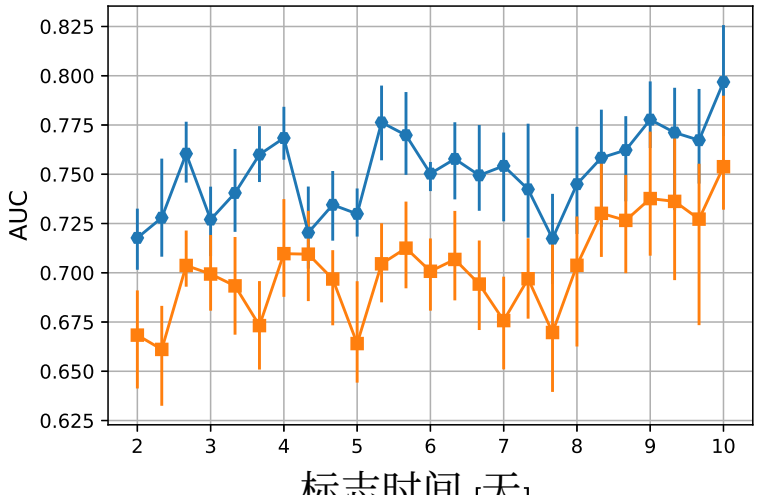


图9：AUROC 分数（y轴）随标志时间（x轴）的变化。两条曲线分别表示基础 CR-LM 模型（橙色）和 Deep-CR-LM 模型（蓝色）的预测性能。误差棒表示 95% bootstrap 置信区间。

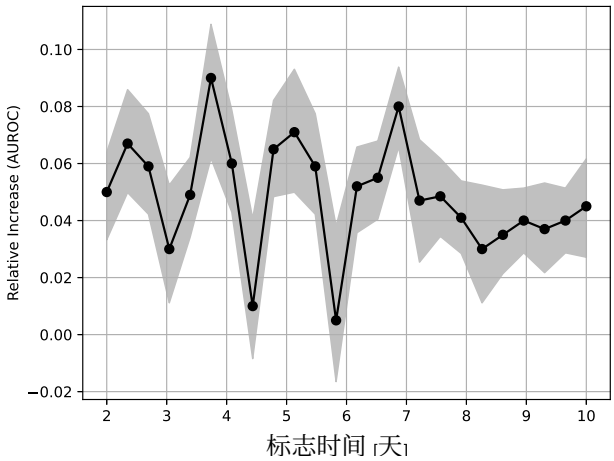


图10：纳入基于 CNN 的风险评分时，AUROC 分数总体相对增幅（y轴）随标志时间（x轴）的变化。95%置信区间以浅灰色区域表示。

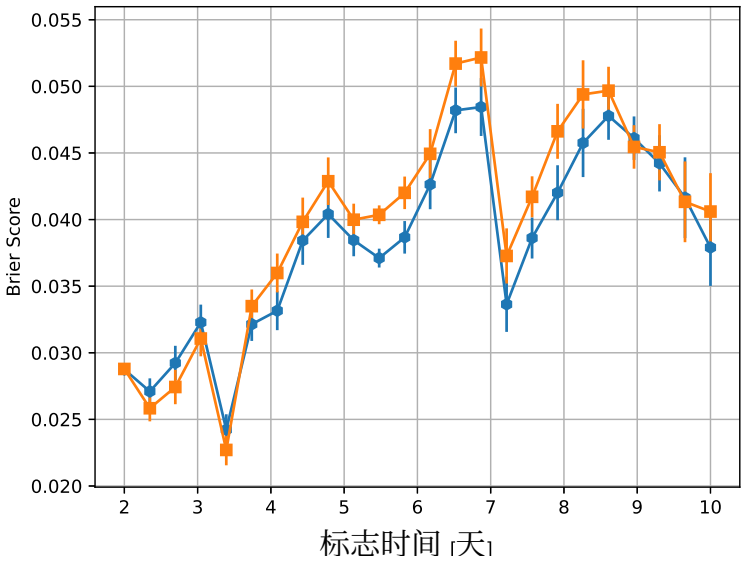


图 11：布里尔分数（纵轴）随标志时间（横轴）的变化。橙色表示 CR-LM模型的预测能力，蓝色表示 Deep-CR-LM。误差棒表示 95%自助法置信区间。

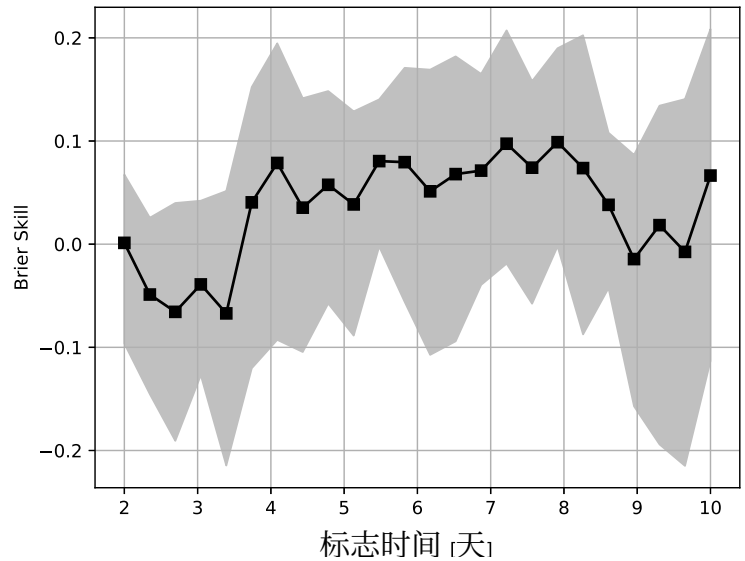


图 12：布里尔技巧评分（纵轴）随标志时间（横轴）的变化。95%置信区间以浅灰色区域表示。

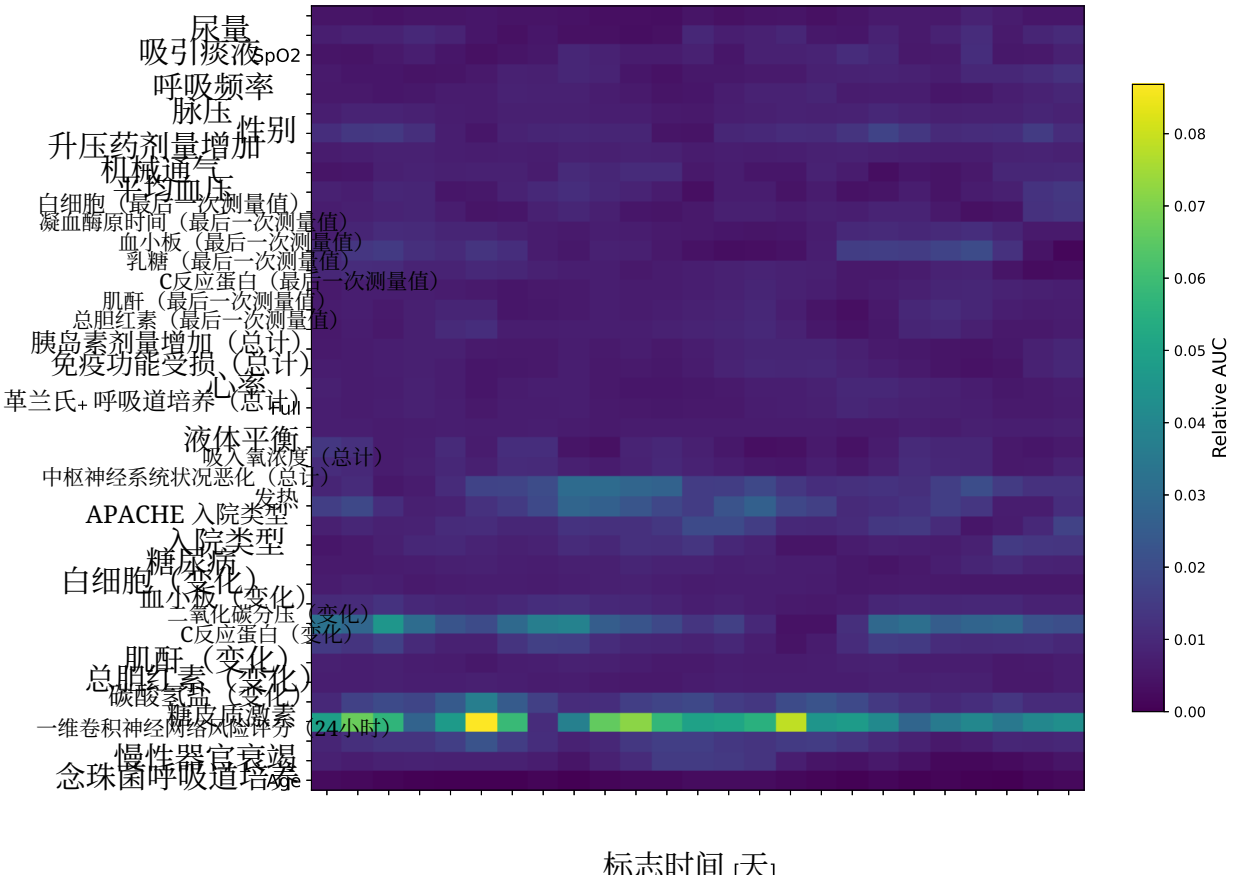


图13：评估 Deep-LM-CR 模型中各预测因子对 ICU-AI 预测影响的 AUC热图。每个像素的颜色表示某一协变量（纵轴）在 LM 时间（横轴）上的影响大小（AUROC 相对增幅）。

未来24小时内发生感染事件。为评估 TB-ANN 的预测能力，我们主要参考 AUROC指标；观察到总体 AUROC（按第5.3节的含义）为0.72（95%置信区间 0.55–0.9）。作为次要指标，我们还考虑了布里尔分数；总体布里尔分数为0.08（95%置信区间 0.06–0.11）。各标志时间的 AUROC 和 BS 平均值分别见图14和图15。误差棒表示95%置信区间

在大多数标志时间，TB-ANN 的曲线下面积分数约为 0.7 至 0.8。不过，在第 2.67、5、8.33 和 10 天出现了较大波动。尤其是后两个时间点，需要指出的是，较晚标志日的事件数减少可能会导致曲线下面积分数被高估。对于布里尔分数，我们观察到较早和较晚标志日的变化曲线有所不同。事实上，在第 2 至 6 天的区间内，布里尔分数围绕全局值 0.08 出现了明显波动。具体而言，在第 3.33 天和第 3.66 天，分数为 0.03；而在第 2.66 天和第 4.66 天，分数则高于 0.10。相比之下，第 6 至 10 天的区间更为稳定，波动幅度小得多，分数在 0.10 左右。

Deep LM-CR 的预测比 TB-ANN 更稳定，而准确率相近。不过，我们的两步策略能够立即解读每个低频协变量对预测的影响，并且可以采用方法来解读卷积神经网络的活动情况，详见第 6 节。

6 基于卷积神经网络的 ICU-AI 预测的可解释性

本节将介绍我们对卷积神经网络活动进行可解释化的尝试。如第 5.4 节所示，基于 CNN 的风险评分提升了 LM-CR 模型的预测能力。不过，目前我们还不了解卷积神经网络在训练过程中所选生命体征的显著性。这方面的知识或许

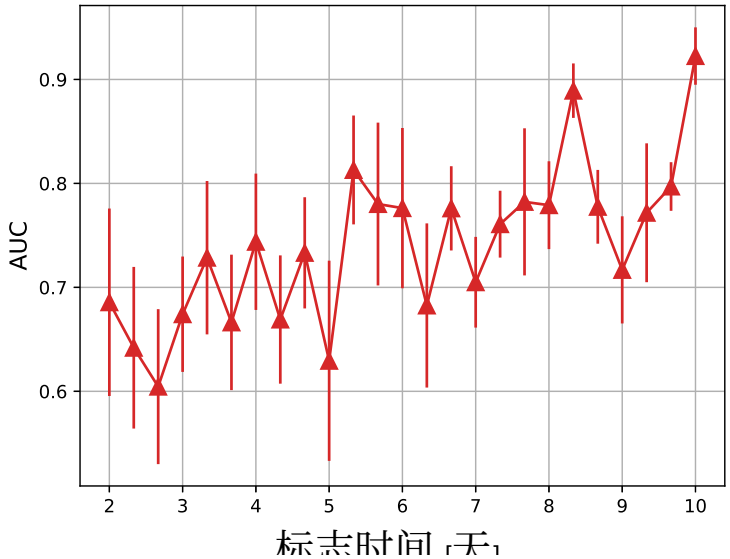


图 14：TB-ANN在各标志时间的曲线下面积均值。误差棒表示95% CI

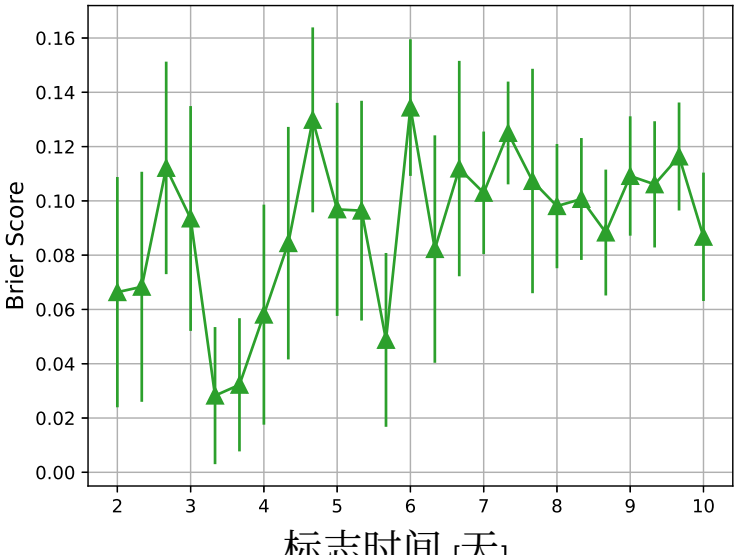


图15：TB-ANN在各标志时间的布里尔分数均值。误差棒表示95%置信区间

对于重症监护病房人工智能而言，了解网络的模式识别活动与患者医学状况之间的关系至关重要。

为了探究卷积神经网络所选模式的哪些特征，我们采用所谓的可解释人工智能 （XAI），即一类旨在理解人工神经网络技术所作决策和预测的方法（Phillips 等人，2020；Vilone 和 Longo，2021；Castelvecchi，2016）。在过去十年中，可解释人工智能已成为一项基础工具，使各种人工神经网络应用更加可靠、透明（Al-Najjar 等人，2022；Neubauer 和 Roy，2022；Jim´enez-Luna 等人，2020；Dave 等人，2020；Lancia 等人，2022，2023）。可解释人工智能的目的，正是应对许多用户在应用人工神经网络技术时普遍存在的黑箱 思维。

6.1 通过 SMOE 量表进行可解释性分析

显著图是作用于隐藏层中已激活特征的图，通常用于显示输入的哪些部分对网络的决策最为重要。本文采用的显著图顺序等价量表（SMOE）基于 Mundhenk 等人（2019年）开发的算法：这是一种高效的非梯度方法，基于对激活特征图的统计分析。有关 SMOE 量表的详细说明，请参阅补充材料第 3 节。

我们希望利用显著性图，从原始的 24小时时间序列中选出最相关的 8小时模式。我们强调，选用 8小时滑动窗口来识别最显著模式，目的是让读者清楚地了解该方法。如上一节所示，8小时最终被证明是研究患者临床动态的合适时间尺度。因此，这一选择旨在尽可能与深度 LR-CR 模型所展示的结果保持一致

所采用的方法如下：

1. 我们针对每个 $t_{LM}^{k} \in \{3, 7, 10\}$ 分别拟合三个不同的卷积神经网络。我们考虑使用三个不同的卷积神经网络，因为网络发现的预测模式可能因重症监护病房住院期间的不同阶段而异（例如，参见第 6.3 节中的讨论）。线性模型的时间点 *3* 天 代表住院早期，*7* 天代表住院中期，而 *10* 天 则代表住院较晚阶段。网络的设计与第 4.2 节所述相同。所有这些模型均通过五折交叉验证进行验证。2. 我们研究隐藏层执行的模式识别，并通过*SMOE* 量表使其具有可解释性。通过此方法，我们可以直观呈现输入数据中显著性最高的区域。具体而言，对于在每个线性模型时间点 t_{LM}^{k} 开发的模型，我们仅构建并可视化测试集的显著性图。我们对每个交叉验证折中的每个测试集重复此操作。3. 对于每张显著图，我们提取*8*小时区间 内累积显著性值最高的区间。从每个时间序列实例中提取最相关的 8 小时模式后，我们便可着重对其进行解释和聚类。图 16 展示了一个提取显著性最高的 8 小时模式的示例。

6.2 显著模式的数据驱动聚类

现在，我们将注意力集中于对第 6.1 节提取出的最显著模式进行聚类。我们确实希望回答以下问题：当 *ICU-AI* 即将到来时，如何将模式识别活动与某些医疗状况联系起来？我们采用以下策略来回答这个问题：

1. 我们按照第 6.1 节所述，将 SMOE 量表应用于时间序列实例，收集由此得到的、振幅为 8小时的最具预测性的模式集合。2. 我们考虑四种临床危重状况，即心动过速、低血压、氧饱和度下降和过度通气 （见表2），这些状况可预测 ICU-AI发作即将发生。这些医学状况反映了全身炎症反应综合征（SIRS）的主要症状，参见 Chakraborty 和 Burns（2019）。心动过速、低血压和过度通气在重症监护病房中相当普遍，通常也会在判定 SIRS 的一般指南中提及（Comstedt 等，2009）。表2所列标准参照 Comstedt 等（2009）；具体而言，氧饱和度下降的标准参照 Hafen 和 Sharma（2022）。3. 我们评估通过 SMOE 量表提取的每个最显著*8*小时模式的 HR、动脉血压、SaO_2 和呼吸频率的均值。根据所得数值（参见表2中的标准），我们检查是否存在这四种临床危重状况。因此，这些状况的组合产生了16种不同且值得关注的临床情形，如表3所示：它们代表了所提出的数据驱动

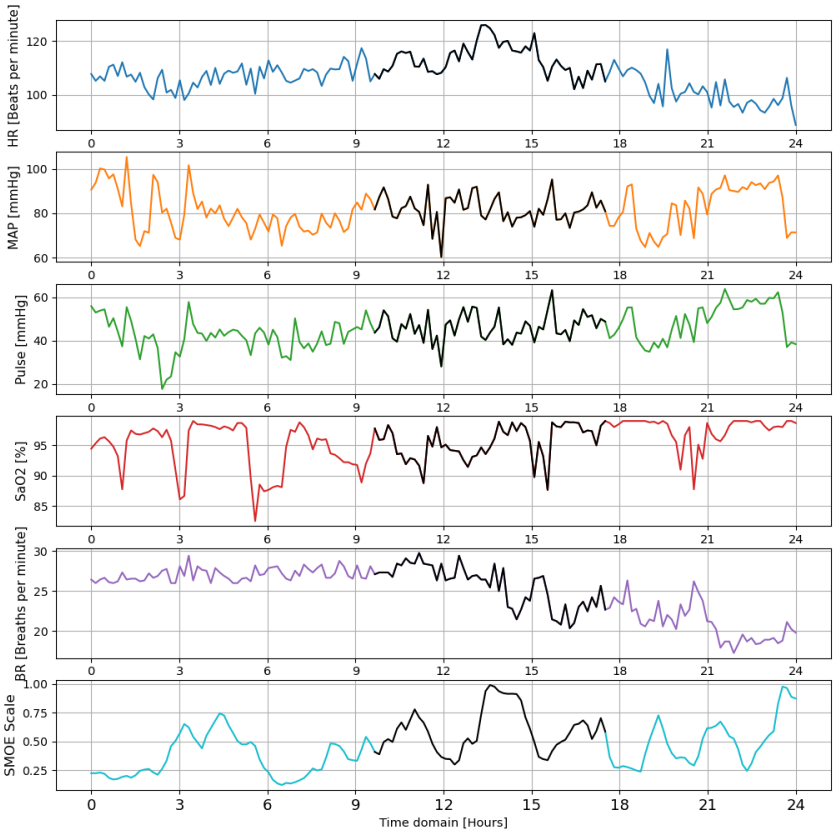


图 16：24小时时间序列样本中8小时最显著模式的示意图。从上至下，前几个信号分别代表所考虑的生命体征。青色线表示相应的SMOE图。黑色轮廓标出了平均SMOE值最高的8小时片段。

危重状况	标准
心动过速	心率 ≥90次/分钟
低血压	动脉血压（均值） ≤ 80毫米汞柱
氧饱和度下降	$SaO_2 < 95\%$
过度通气	呼吸频率 ≥ 24 次/分钟

表2：危重状况及其判定标准。

聚类的类别。图17将这16个不同类别表示为图中的节点（即一个四维超立方体（$\mathcal{Q}_4$））。

6.3 数据驱动聚类的结果

图18展示了16个数据驱动聚类簇的相对频率直方图。对于第3天（见图18(a)和18(b)），双样本柯尔莫哥洛夫-斯米尔诺夫检验（Hodges，1958）表明，未感染与感染实例所属类别的样本分布差异并不显著（p值=0.21）。然而，在第7天和第10天（见图18(d)-(f)），情况截然不同：双样本柯尔莫哥洛夫-斯米尔诺夫检验的零假设被拒绝（p值= 0.0003和p值= 10^{-10} ）。因此，本分析表明，不同的临床状况可能是卷积神经网络模式中的一项重要特征

类别	数据驱动聚类（临床状况）
0	None
1	心动过速
2	低血压
3	低血压、心动过速
4	氧饱和度下降
5	氧饱和度下降，心动过速
6	氧饱和度下降，低血压
7	氧饱和度下降，低血压，心动过速
8	过度通气
9	过度通气，心动过速
10	过度通气，低血压
11	过度通气，低血压，心动过速
12	过度通气，氧饱和度下降
13	过度通气，氧饱和度下降，心动过速
14	过度通气，氧饱和度下降，低血压
15	过度通气，氧饱和度下降，低血压，心动过速

表3：16种临床状况（聚类类别）列表。

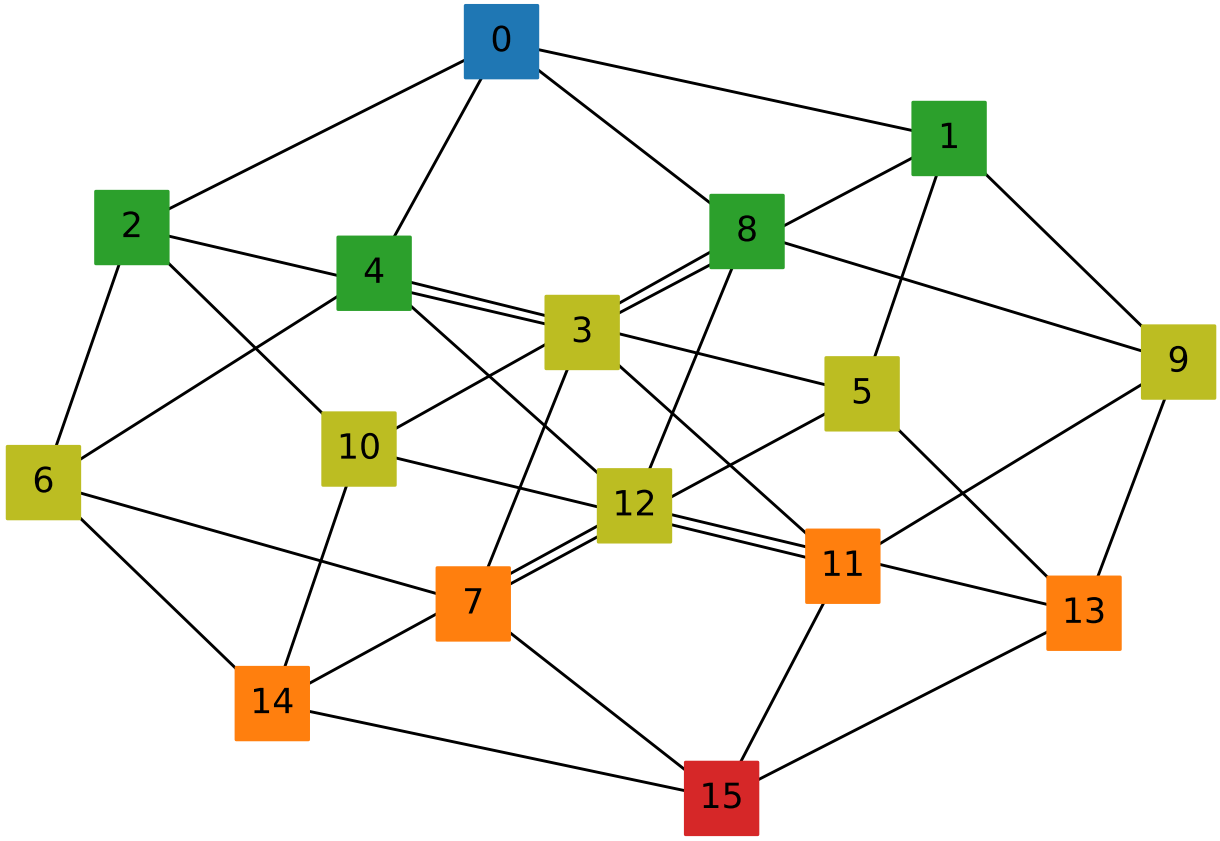
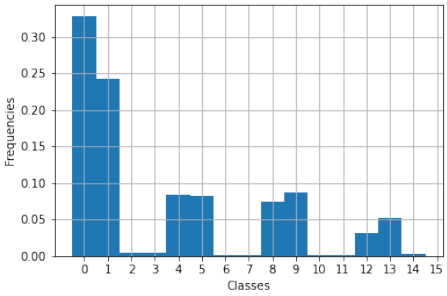
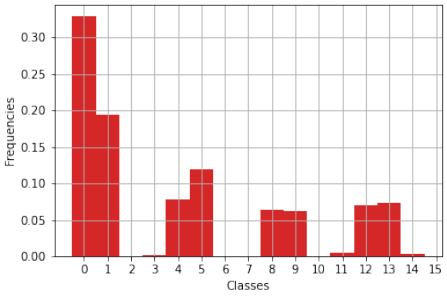


图17：包含聚类的16个类别的超立方体图（$\mathcal{Q}_4$）示意图。节点上的数字表示表3中所列的类别。每个节点的颜色反映相应临床状况的严重程度；（蓝色）无危重程度，（绿色）一种危重状况，（黄色）两种危重状况，（橙色）三种危重状况，（红色）全部四种危重状况

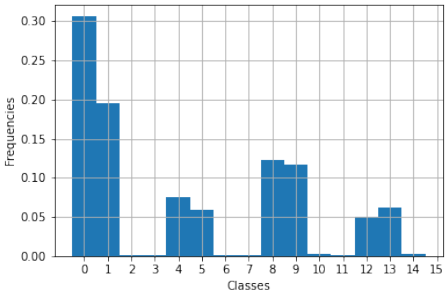
.



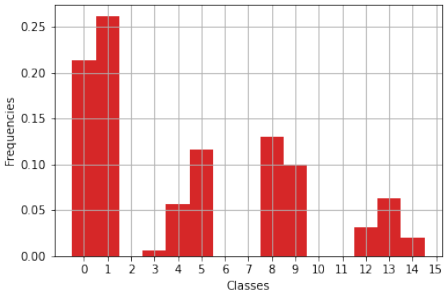
(a)



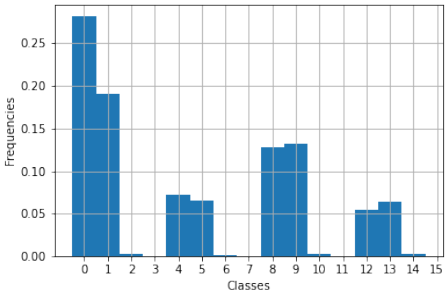
(b)



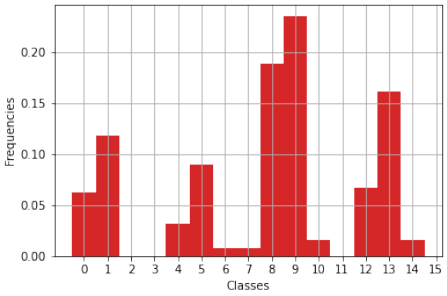
(c)



(d)



(e)



(f)

图18：数据驱动聚类方法的直方图。x轴上的区间代表16个类别。蓝色直方图表示未感染实例，红色直方图表示感染实例。第3天训练的卷积神经网络对应(a)和(b)，第7天训练的对应(c)和(d)，第10天训练的对应(e)和(f)。

模型在学习阶段捕捉到的内容。例如，对于感染实例，第10天至少出现这16种状况中的一种的患病率约为94%，而第7天为79%；见图18 (d)-(f)）。具体而言，第10天，过度通气事件出现在70%的样本中，与心动过速同时出现的比例为23%。而第7天的心动过速则更具相关性，出现在50%的感染样本中。因此，时间序列实例中最显著的8小时子区间可以与特定医学状况相关联，而这些状况已知与ICU-AI的存在有关。

7 结论

我们已证明，所提出的ICU-AI两步建模既是一种准确的预测工具，也是一种可解释模型。正如我们所讨论的，按照我们采用的定义预测感染是一项具有挑战性的任务：感染发生时间由抗生素治疗的开始时间决定。因此，无法确定实际感染时间，这本身就构成了基于高频数据构建高效预测模型的障碍。

然而，卷积神经网络能够通过分析五种生命体征信号的时间序列，捕捉可能具有预测价值的模式。这些模式包含额外的预测信息，且与生命体征信号的平均值仅有轻微相关；传统生存模型通常会纳入这些平均值。此外，我们还表明，SMOE 量表可能有助于医生对即将发生感染的患者进行聚类。

在本研究中，我们考虑了不含删失的生存模型，因为重症监护病房患者在住院期间均受到全面监测。然而，基于伪观测值的方法（Andersen 和 Pohar Perme，2010）是应对删失数据导致所需动态预测产生偏倚的一种可靠策略。在基于线性模型的生存动态预测领域，这种方法已有先例；例如，Nicolaie 等人（2013）的研究，以及 Cortese 等人（2013）以类似方式提出的、理论依据充分的标志点模型推广方法，能够估计基线效应和协变量效应如何影响所需的动态预测，并适用于左删失和右删失。同样，Zhao 和 Feng（2020）近期提出了将人工神经网络与生存预测相结合的初步尝试。尽管该研究仅采用简单的多层感知机架构来求解带有 logit 链接函数的广义模型，但它为开发新方法提供了有前景的思路，有望提高基于多路人工神经网络架构并利用删失数据所获得的生存预测准确率。与 TB-ANN 相比，我们发现，线性模型方法能够带来略为准确且校准良好的预测。尽管两者的总体准确率几乎相同，TB-ANN 的预测在不同标志时点天数上的敏感性往往高得多。据我们所知，这反映了在分析来自不同数据结构的大量信息时，人工神经网络分类器难以实现良好校准这一固有挑战。

我们在竞争风险框架下展示了这一方法。然而，LM方法近期已扩展至多状态模型，即使不满足马尔可夫假设也适用（Putter 和 Spitoni，2018年；Hoff 等，2019年）。因此，作为进一步的扩展，我们可以将复发性感染建模为非马尔可夫多状态模型中的新状态，其转移风险率可能取决于既往感染的发生顺序。此外，未来另一项具有挑战性的研究方向是对卷积神经网络进行某种反演，以识别并分类信号中预测能力更强的模式。这项分析或有助于更准确地对即将发生 ICU-AI 的患者进行聚类。

代码可用性

Python 代码和模块可在 GitHub 上获取：https://github.com/glancia93/ICUAI-dynamic-prediction/blob/main/ICUAI_module.py。

利益冲突声明

作者声明无任何需要披露的利益冲突。

科学写作中生成式人工智能的使用声明

作者声明，在撰写本稿件的过程中未使用生成式人工智能或人工智能辅助技术。

参考文献

Al-Najjar, H. A., Pradhan, B., Beydoun, G., Sarkar, R., Park, H.-J., and Alamri, A. (2022). 一种基于可解释人工智能（xai）的 Shapley 加性解释新方法，用于利用时间序列合成孔径雷达数据集进行空间滑坡预测。*Gondwana Research*。Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). 基于计数过程的统计模型。Springer。Andersen, P. K. and Pohar Perme, M. (2010). 生存分析中的伪观测值。*Statistical methods in medicalresearch*, 19(1):71–99。Borovykh, A., Bohte, S., and Oosterlee, K. (2017). 使用卷积神经网络进行条件时间序列预测。载于计算机科学讲义/人工智能讲义，第729–730页。Brier, G. W. (1950). 以概率形式表示的预测的验证。*Monthly weather review*, 78(1):1–3。Castelvecchi, D. (2016). 我们能否揭开人工智能黑箱？*Nature News*, 538(7623):20。Chakraborty, R. K. and Burns, B. (2019). 全身炎症反应综合征。Comstedt, P., Storgaard, M., and Lassen, A. T. (2009). 急性住院内科患者的全身炎症反应综合征（SIRS）：一项队列研究。*Scandinavian journalof trauma, resuscitationand emergencymedicine*, 17(1):1–6。Cortese, G. and Andersen, P. K. (2010). 竞争风险与时变协变量。*Biometrical Journal*, 52(1):138–158。Cortese, G., Gerds, T. A., and Andersen, P. K. (2013). 比较含时变协变量的竞争风险模型的预测结果。*Statisticsinmedicine*, 32(18):3089–3101。Dantes, R. B. and Epstein, L. (2018). 应对脓毒症：公共卫生视角。*Clinical infectious diseases*, 67(8):1300–1302。Dave, D., Naik, H., Singhal, S., and Patel, P. (2020). 可解释人工智能与医疗保健相遇：一项关于心脏病数据集的研究。*arXiv preprintarXiv:2011.03195*。Deng, Y., Ma, Y., Fu, J., Wang, X., Yu, C., Lv, J., Man, S., Wang, B., and Li, L. (2023). 一种用于预测体检人群非酒精性脂肪性肝病的动态机器学习模型：一项纵向研究。*Heliyon*, 9(8)。Ferrer, L., Putter, H., and Proust-Lima, C. (2019). 使用标志时间法和联合建模进行个体动态预测：估计量的验证与稳健性评估。*Statistical Methods in Medical Research*, 28(12):3649– 3666。Gandin, I., Scagnetto, A., Romani, S., and Barbati, G. (2021). 时间序列深度学习模型的可解释性：一项针对入住重症监护病房的心血管患者的研究。*Journal ofBiomedicalInformatics*, 121:103876。Guo-yan, X., Jin, Z., Cun-you, S., Wen-bin, H., and Fan, L. (2019). 基于卷积神经网络和mc的水文时间序列预测组合模型。*Computer andModernization*, (11):23。Hafen, B. B. and Sharma, S. (2022). 血氧饱和度。StatPearls, StatPearls Publishing。Hodges, J. L. (1958). 斯米尔诺夫双样本检验的显著性概率。*Arkivf¨orMatematik*, 3(5):469–486。Hoff, R., Putter, H., Mehlum, I. S., and Gran, J. M. (2019). 含协变量的非马尔可夫多状态模型中转移概率的标志时间估计。*Lifetime DataAnalysis*, 25(4):660–680。Ivanov, O., Molander, K., Dunne, R., Liu, S., Masek, K., Lewis, E., Wolf, L., Travers, D., Brecher, D., Delaney, D., et al. (2022). 利用机器学习与临床自然语言处理，在急诊分诊时准确检测脓毒症。*arXiv preprintarXiv:2204.07657*。Jim´enez-Luna, J., Grisoni, F., and Schneider, G. (2020). 利用可解释人工智能进行药物发现。*NatureMachine Intelligence*, 2(10):573–584。Kagaya, H., Aizawa, K., and Ogawa, M. (2014). 使用卷积神经网络进行食物检测与识别。载于第*22*届*ACM*多媒体国际会议论文集，第1085–1088页。Kingma, D. P. and Ba, J. (2015). Adam：一种随机优化方法。载于*ICLR*。Klouwenberg, P. M. K., Ong, D. S., Bos, L. D., de Beer, F. M., van Hooijdonk, R. T., Huson, M. A., Straat, M., van Vught, L. A., Wieske, L., Horn, J., et al. (2013). 美国疾病控制与预防中心用于危重患者感染分类的标准的观察者间一致性。*Critical care medicine*, 41(10):2373–2378。Kwon, D., Natarajan, K., Suh, S. C., Kim, H., and Kim, J. (2018). 一项关于使用卷积神经网络进行网络异常检测的实证研究。载于*ICDCS*，第1595–1598页。

Lancia, G., Durastanti, C., Spitoni, C., De Benedictis, I., Sciortino, A., Cirillo, E. N., Ledda, M., Lisi, A., Convertino, A., and Mussi, V. (2023). 用于对肿瘤亚型基因组DNA拉曼光谱进行分类的学习模型。*arXiv*预印本 *arXiv:2302.08918*。Lancia, G., Goede, I., Spitoni, C., and Dijkstra, H. (2022). 基于数据的方法对厄尔尼˜诺预测中的物理学的捕捉。《混沌：非线性科学跨学科期刊》，32(10)。Liu, Y. H. (2018). 基于卷积神经网络的特征提取与图像识别。载于《物理学杂志：会议系列》，第1087卷，第062032页。IOP出版社。Livieris, I. E., Pintelas, E., and Pintelas, P. (2020). 用于黄金价格时间序列预测的CNN–LSTM模型。《神经计算与应用》，32(23):17351–17360。Lou, G. and Shi, H. (2020). 基于卷积神经网络的人脸图像识别。《中国通信》，17(2):117–124。Maki, D. G., Crnich, C. J., and Safdar, N. (2008). 重症监护病房中的医院感染。*Critical Care Medicine*，第1003页。May, R., Dandy, G., and Maier, H. (2011). 人工神经网络输入变量选择方法综述。载于Suzuki, K., 编，人工神经网络，第2章。IntechOpen, 里耶卡。Mundhenk, T. N., Chen, B. Y., and Friedland, G. (2019). 用于可解释人工智能的高效显著图。*arXiv*预印本 *arXiv:1911.11293*。Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., and Han, K. (2018). 基于深度神经网络的增强型网络异常检测。《*IEEE Access*》，6:48231–48246。Neubauer, M. S. and Roy, A. (2022). 面向高能物理的可解释人工智能。*arXiv*预印本 *arXiv:2206.06632*。Nicolaie, M., Van Houwelingen, J., De Witte, T., and Putter, H. (2013). 竞争风险中的标志时间法动态预测。*Statistics in Medicine*，32(12):2031–2047。Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., and Przybocki, M. A. (2020). 可解释人工智能的四项原则。盖瑟斯堡，马里兰州。Proust-Lima, C. and Taylor, J. M. (2009). 利用治疗后PSA的重复测量开发并验证前列腺癌复发动态预后工具：一种联合建模方法。《生物统计学》，10(3):535–549。Putter, H. and Spitoni, C. (2018). 非马尔可夫多状态模型中转移概率的非参数估计：标志点Aalen–Johansen估计量。*Statistical MethodsinMedicalResearch*，27(7):2081–2092。Rizopoulos, D. (2011). 纵向数据与事件发生时间数据联合模型中的动态预测和前瞻性准确率。《生物统计学》，67(3):819–829。Rizopoulos, D. (2012). 纵向数据与事件发生时间数据的联合模型：*R*语言应用。CRC出版社。Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., and Soman, K. (2017). 使用LSTM、循环神经网络和CNN滑动窗口模型进行股票价格预测。载于*2017*年计算、通信与信息学进展国际会议（*ICACCI*），第1643–1647页。IEEE。Spitoni, C., Lammens, V., and Putter, H. (2018). 多状态模型中状态占据和转移概率的预测误差。*Biometrical Journal*，60(1):34–48。Staar, B., L¨utjen, M., and Freitag, M. (2019). 使用卷积神经网络进行工业表面检测中的异常检测。《*CIRP*会议录》，79:484–489。Topol, E. J. (2019). 高性能医学：人类智能与人工智能的融合。《自然·医学》，25(1):44–56。Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). DNA微阵列的缺失值估算方法。《生物信息学》，17(6):520–525。van Houwelingen, H. and Putter, H. (2011). 临床生存分析中的动态预测。CRC出版社。Van Houwelingen, H. C. (2007). 事件史分析中的标志时间法动态预测。《斯堪的纳维亚统计学杂志》，34(1):70–85。Vilone, G. and Longo, L. (2021). 可解释人工智能的可解释性概念与评估方法。《信息融合》，76:89–106。Vincent, J., Rello, J., and Marshall, J. (2009). 重症监护病房中感染患病率及结局的国际研究。《美国医学会杂志》，302(21):2323–9。Wilks, D. S. (2011). 大气科学中的统计方法，第100卷。学术出版社。

Zeng, Z., Hou, Z., Li, T., Deng, L., Hou, J., Huang, X., Li, J., Sun, M., Wang, Y., Wu, Q., et al. (2022). 一种用于预测机械通气脓毒症患者呼吸机参数的深度学习方法。*arXiv*预印本 *arXiv:2202.10921*。Zhao, L. and Feng, D. (2020). 使用伪值进行生存分析的深度神经网络。《*IEEE*生物医学与健康信息学杂志》，24(11):3308–3314。Zheng, H., Fu, J., Mei, T., and Luo, J. (2017). 用于细粒度图像识别的多注意力卷积神经网络学习。载于*IEEE*国际计算机视觉会议论文集，第5209–5217页。

arXiv:2301.11146v2 [stat.AP] 6 Mar 2024

补充材料

1 数据、协变量与风险

本研究在脓毒症分子诊断与风险分层（MARS）研究（ClinicalTrials.gov 标识符 NCT01905033）的框架内开展。该研究是一项前瞻性重症监护病房队列研究，机构审查委员会批准采用知情同意选择退出法（方案编号 10-056C）。模型中纳入的时间固定变量见表1，时变协变量见表2。

变量名称	变量描述
Sex	性别（男/女）
Age	重症监护病房入院时的年龄
免疫缺陷	免疫功能低下状态；定义为患有获得性免疫缺陷综- 合征、大剂量使用糖皮质激素（相当于泼尼松龙 >75 毫克/天，至少持续 1 周）、目前正在使用免疫抑制药物、目前正在使用 抗肿瘤药物、近期患有血液系统恶性肿瘤，或有体液或细- 细胞缺乏
再入院	本次住院期间曾入住重症监护病房
主要专科	重症监护病房入院的诊断类别（心血管、胃肠道、神经系统、 呼吸系统、移植后、创伤、其他）
糖尿病	糖尿病病史
慢性 皮质类固醇- 的使用	慢性用药：全身性皮质类固醇
慢性器官衰竭	存在慢性器官功能不全，并且病史中记录了以下任一情况： 病史中记录： • 慢性心力衰竭定义为病史中记录有纽约心脏病协会心功能分级 2-4级，或射血分数 <45%（重症监护病房入院前2年内超声心动图检查结果） 或端坐呼吸并长期使用利尿剂 • 严重心血管功能不全，定义为静息时或 轻微活动时出现心绞痛或呼吸困难（NYHA IV级） • 慢性肾功能不全，定义为血清肌酐持续升高 >177 μmol/L或长期透析 • 导致严重功能损害的慢性限制性、阻塞性或血管性肺部疾病 功能损害 • 慢性肝衰竭伴门静脉高压（肝活检结果呈阳性） 和/或因门静脉高压导致的上消化道出血，和/或 因肝衰竭病史而发作肝性脑病/昏迷 ure
入院类型	入住内科/外科三级重症监护病房

表1：包含所有基线预测因子的表。

拟合后的 Deep LM-CR 模型的感染原因特异性风险 $\beta_1^{(0)}$ 如表3所示。

2 高频生命体征的人工神经网络

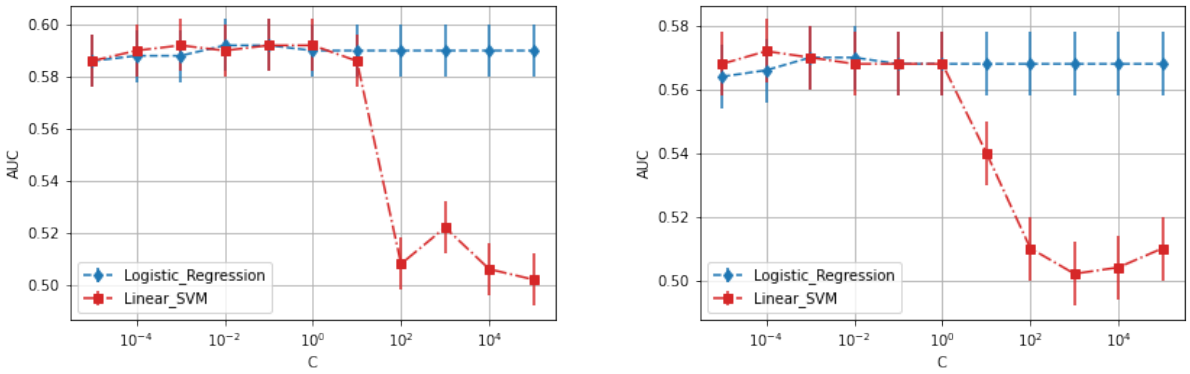
本节介绍用于高频生命体征的最佳人工神经网络设计的模型选择，即用于构建感染风险评分的五项电子健康记录。除此之外，我们还考虑了两种经典机器学习模型：逻辑回归和支持向量机。模型选择针对两个不同的预测窗口进行：24小时和48小时。为确定最佳模型，我们对超参数进行细致的网格搜索；根据选定的超参数组合，我们提出了准确率最高的最佳模型。在此阶段，准确率以经5折交叉验证模型的平均曲线下面积进行评估。

变量名称	变量描述
心率	1 小时平均心率的中位数（次/分钟）
血压	1 小时平均血压的中位数，即有创平均动脉压测量值或无创袖带测量值（毫米汞柱）
血氧饱和度	1 小时平均血氧饱和度的中位数（%）
呼吸频率	1 小时平均呼吸频率的中位数（rpm）
脉搏	1 小时平均脉压的中位数（收缩压与舒张压之差，血压，毫米汞柱）
有创机械通气	最近一次观测到的机械通气状态
FiO2	过去 8 小时内最近一次观察到的吸入氧浓度值
长期皮质类固醇使用	长期用药：全身性皮质类固醇
发热	过去8小时内是否发热（>38摄氏度）
液体平衡	过去8小时的液体平衡（mL）
尿量	8小时内尿量总计（mL）
吸引痰液	在一个8小时的时间窗口
中枢神经系统状态恶化tus	意识水平下降（GSC运动评分降低或RASS评分恶化）或过去8小时内出现新的谵妄发作
C反应蛋白（最后一次测量值）	最近一次观测的C反应蛋白（mg/L）
C反应蛋白（变化量）	C反应蛋白相较于24小时前的单位变化量（mg/L）
白色血cell计数（最后一次测量值）	最近一次观测到的白细胞计数（x10^9/L）
白色计数cell计数（变化）	相对于数小时前白细胞（WBC）计数的单位变化（x10^9/L）
血小板计数（最近一次值）	最近一次观察到的血小板计数（x10^9/L）
血小板计数（变化）	相较于 24 小时前的血小板计数单位变化（x10^9/L）
凝血酶原time（最后一次测量值）	最近一次观测到的凝血酶原时间（秒）
肌酐（最后一个测量值）	最近一次观测到的肌酐（μmol/L）
肌酐（变化）	与 24 小时前相比，肌酐的单位变化量（μmol/L）
总胆红素（最近一次数值）	最近一次观测到的总胆红素（μmol/L）
总胆红素（变化）	总胆红素相较于 24 小时前的单位变化（μmol/L）
碳酸氢盐（变化）	碳酸氢盐相较于 24 小时前的单位变化（mmol/L）
乳酸（最后一次测量值）	最后一次观测到的乳酸（mmol/L）
血管加压药速率增加	相较于前一个8小时窗口，去甲肾上腺素剂量均值增加
增加胰岛素dose	相较于前一个8小时窗口，胰岛素剂量均值增加
革兰+呼吸系统培养	气道中培养出革兰氏阳性菌（最近一次培养的结果）
呼吸系统中的念珠菌-培养结果	气道中培养出的念珠菌属（最近一次培养的结果）

表2：包含所有时间依赖性预测因子的表。

协变量	$\beta_1^{(0)}$	β-CI
尿量	1	0.99-1.01
吸引痰液	1	0.99-1.02
SpO2	0.97	0.95-0.98
呼吸频率	1	0.99-1.01
再入院	0.92	0.85-1.11
脉压	1	0.99-1.01
性别（男性）	1.4	1.20-1.53
升压药剂量增加	1.4	1.25-1.49
机械通气	1	0.88-1.21
平均血压	1	0.99-1.01
白细胞（最后一次测量值）	1	0.99-1.01
凝血酶原时间（最后一次测量值）	1	0.99-1.01
血小板（最近一次测量值）	1	0.99-1.01
乳糖（最近一次测量值）	0.98	0.95-1.04
C反应蛋白（最近一次测量值）	1	0.99-1.01
胆红素（总量）	1	0.99-1.01
胰岛素剂量增加（总量）	1	0.99-1.01
免疫功能受损（总量）	1.2	1.01-1.54
心率（总计）	1	0.95-1.05
革兰氏+ 呼吸道培养（总计）	1.1	1.01-1.54
液体平衡	1	0.99-1.01
吸入氧浓度（总计）	1	0.99-1.01
中枢神经系统状况恶化（总计）	1.1	1.02-1.18
发热	2	1.86-2.75
APACHE创伤评分	1	0.92-1.34
APACHE-创伤	0.97	0.86-1.26
APACHE-呼吸	0.67	0.54-0.77
APACHE-其他	0.58	0.33-0.96
APACHE-神经	1.2	1.01-1.48
APACHE-胃肠	0.71	0.54-0.96
入院类型（外科）	1.3	1.16-1.44
糖尿病	0.82	0.73-0.99
白细胞（变化）	1	0.99-1.01
血小板（变化）	1	0.99-1.01
二氧化碳分压（变化）	1	0.99-1.01
C反应蛋白（变化）	1	0.99-1.01
总胆红素（变化）	1	0.99-1.01
碳酸氢盐（变化）	1	0.99-1.01
糖皮质激素	1.2	0.99-1.48
CNN风险评分	4.8	3.05-6.72
慢性器官衰竭	1.1	0.95-1.28
念珠菌呼吸道培养	0.82	0.72-0.91
age	1	0.99-1.01

表3：ICU-AI 的原因特异性风险



(a) (b)

图1：24小时预测模型（a）和48小时预测模型（b）的曲线下面积随逆收缩参数（C）变化的情况。红色曲线表示支持向量机模型，蓝色线条表示逻辑回归模型。

2.1 模型选择

我们首先评估了逻辑回归、支持向量机和多层感知机在沿时间域汇总时间序列后的准确率。因此，我们从时间序列实例的整个24小时（或48小时）范围内提取了这些实例的简单统计量（例如均值、标准差、偏度、峰度、最小值和最大值）；对五种生命体征中的每一种都进行了这项操作。我们总共获得了31个输入特征。需要指出的是，这些提取的输入特征均经过标准化（均值为零，标准差为单位值）。

逻辑回归和支持向量机均采用了正则化；这两个模型的权重均受到 L^2 范数惩罚。我们使用收缩参数的倒数（此处记为 C），它是这两个模型唯一的超参数。因此，我们通过优化 C 来寻找最优值，以曲线下面积为优化目标；结果见图1。然而，我们观察到，这两个模型的曲线下面积分数均无法超过0.59。24小时模型和48小时模型也得到了类似的较低结果。

与之前的模型不同，在考虑多层感知机时，模型要取得最佳准确率，离不开对更广泛的最优超参数组合进行搜索。确实，需要调整的参数更多：单元数、深度、丢弃率、学习率、激活函数以及批量大小。如前所述，我们通过寻找在五折交叉验证中返回最高平均曲线下面积的参数组合，来完成这些超参数的调优。需要注意的是，由于需要调优的参数数量庞大，将曲线下面积分数表示为超参数的函数并进行可视化，实际上并不方便。为此，我们先固定一个超参数，考察所有相应的配置（例如，所有深度为 3 或单元数为 16 的配置），再选出曲线下面积分数最高的配置（我们将其称为超参数的最大曲线下面积）。随后，我们绘制最大曲线下面积，以了解曲线下面积随单个超参数变化的情况。我们使用的多层感知机仅由一系列稠密层组成（我们将稠密层的数量称为深度），并配有非线性激活函数；每一层的正下方都设置了一个丢弃层以进行正则化。该多层感知机模型使用 ADAM优化器对二元交叉熵进行优化（学习率 和批量大小等超参数是优化器的功能参数）。图 2a 和 2b 展示了采用 ReLU 激活函数的多层感知机模型（即 $\mathrm{ReLU}(x) = \max(0, x)$）：该模型的曲线下面积分数未能超过 0.63。类似地，选择双曲正切函数（tanh）作为激活函数也得到相近的结果；见图 2c 和 2d。

我们分析的下一类模型是卷积神经网络。与多层感知机类似，我们采用相同的策略，在一组超参数范围内优化曲线下面积：卷积滤波器数量、卷积核大小、深度、学习率、丢弃率以及批量大小。在这种情况下，激活函数未纳入待调优的超参数；与多层感知机模型不同，我们只考虑了 ReLU 激活函数。做出这一决定的原因是，我们注意到，采用单次留出方法进行的多项测试显示，S形激活函数（例如 Sigmoid 函数或双曲正切函数）会影响模型的预测能力；在不同的容量组合（即滤波器数量 $times$ 丢弃率）、深度（即隐藏层数量）和感受野（即不同大小的卷积核与最大池化层的组合）下，我们得到的曲线下面积始终低于 0.60。

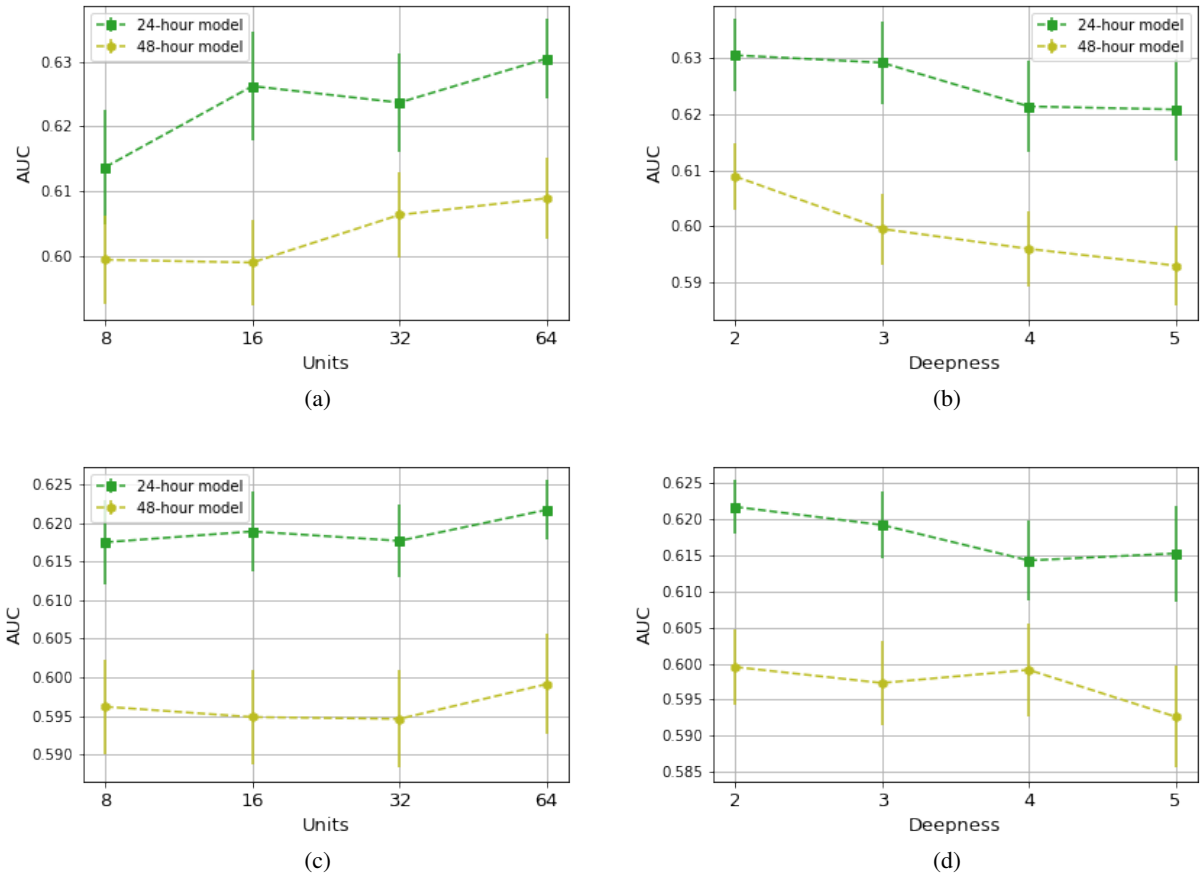
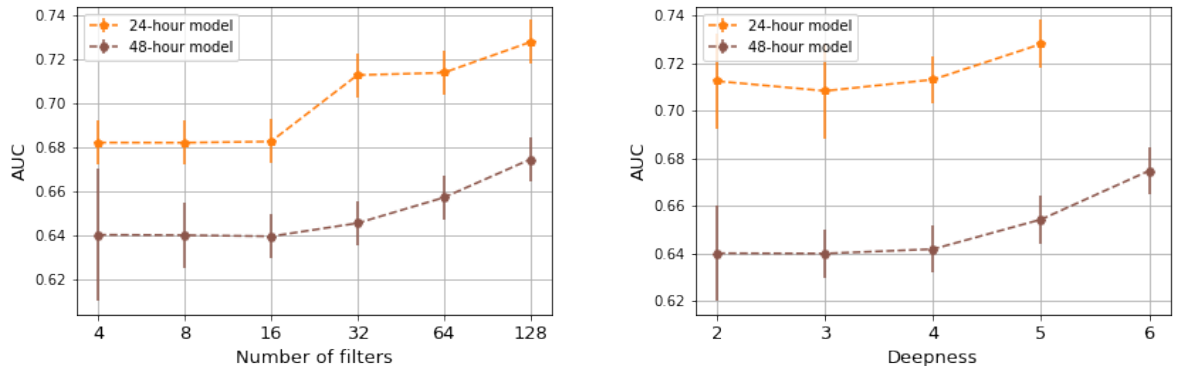


图 2：多层感知机模型。最大曲线下面积随超参数单元（即单元数）和深度的变化。每幅图均展示 24小时模型（绿色线）和 48 小时模型（黄色线）的表现。考虑以下情况：（a）单元数和 ReLU 激活函数，（b）深度和 ReLU 激活函数，（c）单元数和 tanh 激活函数，（d）深度和 tanh 激活函数。

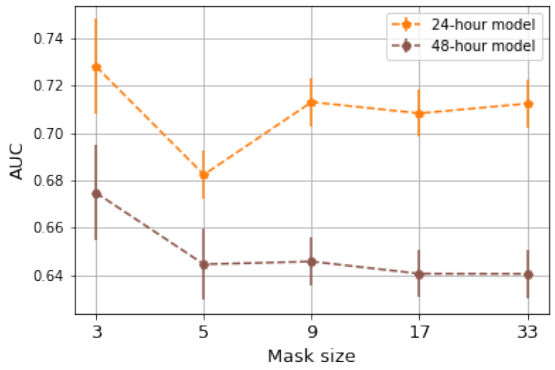
在将生命体征输入卷积神经网络模型之前，必须先进行一些预处理步骤。首先，对所有时间序列实例应用一次线性变换，使其在 [-1, 1] 范围内得到更紧凑的表示。具体而言，我们对所有实例应用同一类型的变换；根据需要重新缩放的生命体征类型，采用相应的线性变换。例如，我们使用相同的线性变换，对所有时间序列实例中包含的心率时间序列进行重新缩放；而对于其他所有生命体征（例如呼吸频率或其他指标），则设计并采用了另一种变换。因此，根据我们掌握的所有生命体征的总体统计数据，我们为每个时间序列特征构建了一个线性映射，将其重新缩放至 [-1, 1] 范围。例如，我们可以考虑时间序列实例中的心率预测因子；根据所有患者的全部心率指标的总体统计数据，我们记录到全局最小值和最大值分别为 41 次/分钟和 239 次/分钟。因此，以 $X_i^{HR}(t)$ 表示第 i 个时间序列实例的心率特征，我们对 $X_i^{HR}(t)$ 应用以下变换

$$X_i^{HR}(t) \to \frac{2X_i^{HR}(t) - 41\text{bpm} - 239\text{bpm}}{239\text{bpm} - 41\text{bpm}}.$$

不同于标准化（即规定所有时间序列特征都具有单位方差和零均值），应用这些基于数据的线性变换不会严重扭曲生命体征信号的固有特征，例如尺度（即均值）和能量（即经验二阶矩）。此外，数据采用分段近似聚合（PPA）方法进行处理（Keogh 和 Pazzani，2000年）。我们没有表示所有高尺度细节，而是在保持欧几里得空间中距离测量下界的同时，获得了生命体征信号的一种精简但信息丰富的表示（Chen 和 Qi，2019年）。因此，我们使用 PPA 聚合长度为 9 分钟的时间间隔。



(a) (b)



(c)

图 3：卷积神经网络模型。最大曲线下面积随超参数滤波器数量 (a)、深度 (b)和卷积核大小的变化。每幅图均展示了24小时模型和 48小时模型的表现。

在图 3a 中，我们可以看到，卷积神经网络的准确率随着滤波器数量的增加而提高：较多的滤波器（例如 128 个）使24小时模型和 48 小时模型都具有较高的准确率，其曲线下面积分别为 0.72 和 0.67。我们首先研究了一维卷积神经网络。由多个隐藏层组成是使模型具备良好性能的另一项关键特征。事实上，在图 3b 中可以观察到，5 个隐藏层对于24小时模型已经足够（曲线下面积范围为 [0.70、0.72]），而 48 小时模型则需要 6 个卷积层才能达到最高曲线下面积（0.67）。相反，卷积掩模的幅度会降低曲线下面积值，尤其是在考虑大小为 17 或 33 的掩模时；见图 3c。大小为 3 的卷积掩模使 24小时模型和 48 小时模型的曲线下面积值分别保持在 0.72 和 0.67。因此，我们对最佳配置的研究表明，应采用具有强大性能（即滤波器数量多）、层次较深且使用小尺寸卷积核的卷积神经网络模型。

我们比较了纯卷积方法（即一维卷积神经网络模型）与卷积神经网络-长短期记忆网络模型。后者是一种卷积神经网络模型，其架构与我们找到的最佳一维卷积神经网络完全相同，唯一不同的是用长短期记忆层替代了卷积神经网络模型的展平层。长短期记忆层的作用是在分析输入特征图沿时间域的连接关系时，生成数组结构的特征图。此外，长短期记忆层只有一个相关超参数，即表示数组结构编码维度的单元数量。因此，我们没有重新进行细网格搜索，而只关注通过用长短期记忆层替换展平层，能否改进找到的最佳一维卷积神经网络。尽管如此，单元数量仍是我们研究的唯一变量。图4显示，增加长短期记忆层的单元数量并未带来 AUROC 分数的整体提升。对于24小时模型，从单元数64开始进入的平台区表明，卷积神经网络-长短期记忆网络模型与卷积神经网络模型的准确度相同，即 AUROC 分数等于 0.72 ± 0.01。对于48小时模型，无论长短期记忆网络单元采用何种配置，AUROC 值都无法超过0.6。

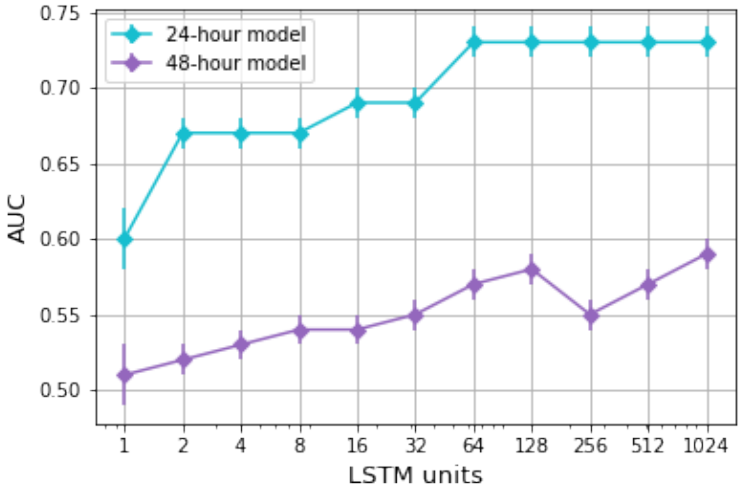


图 4：卷积神经网络-长短期记忆网络模型。AUROC 随长短期记忆层单元数量的变化而变化。卷积神经网络-长短期记忆网络模型中所有其他超参数的选择均与最优卷积神经网络模型的选择相同。

我们测试的最后一类卷积神经网络模型是二维卷积神经网络。尽管一维卷积层是分析时间序列结构数据最自然的选择，但只要将序列数据表示为二维形式，就始终可以采用基于二维卷积的方法。例如，Ye 等人（2019年）提出的方法可以将时间序列数据表示为二维形式，即进行二维分箱，其中每个二维分箱统计在时间域的特定时刻、特定取值范围内出现的记录数；示例见图5。电子健康记录的二维网格结构表示使我们能够探究使用二维卷积神经网络模型识别 ICU-AI 发病起始的可能性。

尽管二维卷积神经网络与一维卷积神经网络具有相似的结构，但二维卷积神经网络的实现需要调整更多的超参数。超参数数量的增加主要源于数据的二维结构；与一维情况不同，在二维情况下，不仅需要优化卷积掩模的宽度和高度，还需要优化表示每个时间序列特征的二维分箱的高度和宽度。与一维卷积神经网络模型一样，我们在精细的参数网格上对模型进行了最优调参：滤波器数量、卷积核大小（在两个维度上）、深度、丢弃、学习率、批量大小，以及二维分箱的宽度和高度。从图 6 可以看出，二维卷积神经网络模型架构的剧烈变化并不会导致最大曲线下面积分数的评估结果发生显著变化。也就是说，在滤波器数量（见图 6a）、深度（见图 6b）、卷积掩模大小（图 6c）以及二维分箱的高度和宽度（图 6d）方面选择多种配置，都无法使 24小时模型和 48小时模型获得大于 0.63 的 AUROC 分数。

2.2 插补方法的选择

为定量说明选择 LOCF 插补方法的依据，我们在此对两种替代方法进行比较分析。第一种方法是多变量 k 近邻（k 最近邻）（Troyanskaya et al., 2001）。k 近邻方法最初为矩阵数据设计；在这里，我们通过将张量数据（样本数、时间序列长度、时间序列特征）整理为矩阵数据（样本数、特征），展平时间序列特征，从而对该方法进行改造。因此，后者的特征包含了每个时间点上各条时间序列的信息。随后，我们采用 k 近邻方法插补缺失数据，利用单个实例中每条记录与其他所有记录之间学得的关联。第二种方法以下称为 *ICUAI*插补 方法，旨在从临床角度主要根据缺失值的性质提取信息。具体而言，持续时间超过 4 小时的缺失区间以一个超出取值范围的常数（例如 100）替代，而短于 4 小时的区间则以空值常数进行插补。ICUAI 插补方法的设计思路源于管理电子健康记录（EHR）的实际医疗经验。约 4 小时或更长的区间通常对应外科手术的持续时间，而较短的区间往往与重症监护室监测暂时中断有关，其原因可能是患者不慎使设备脱落，或设备发生错位。采用此方法，我们有系统地为特定类型的缺失区间填入指定的

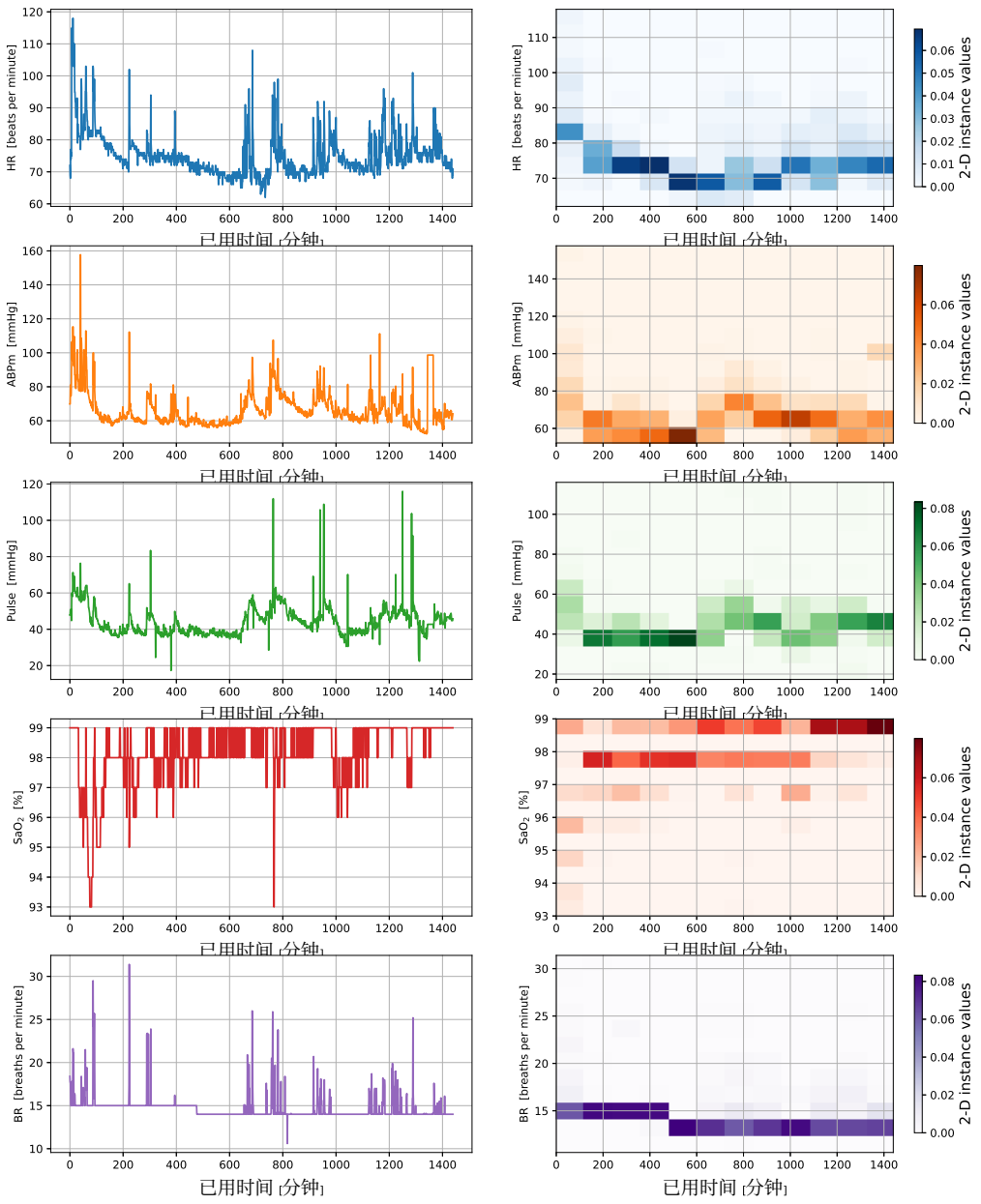


图5：示例时间序列实例的二维表示。左列为时间序列特征（电子健康记录），而右列为其二维表示

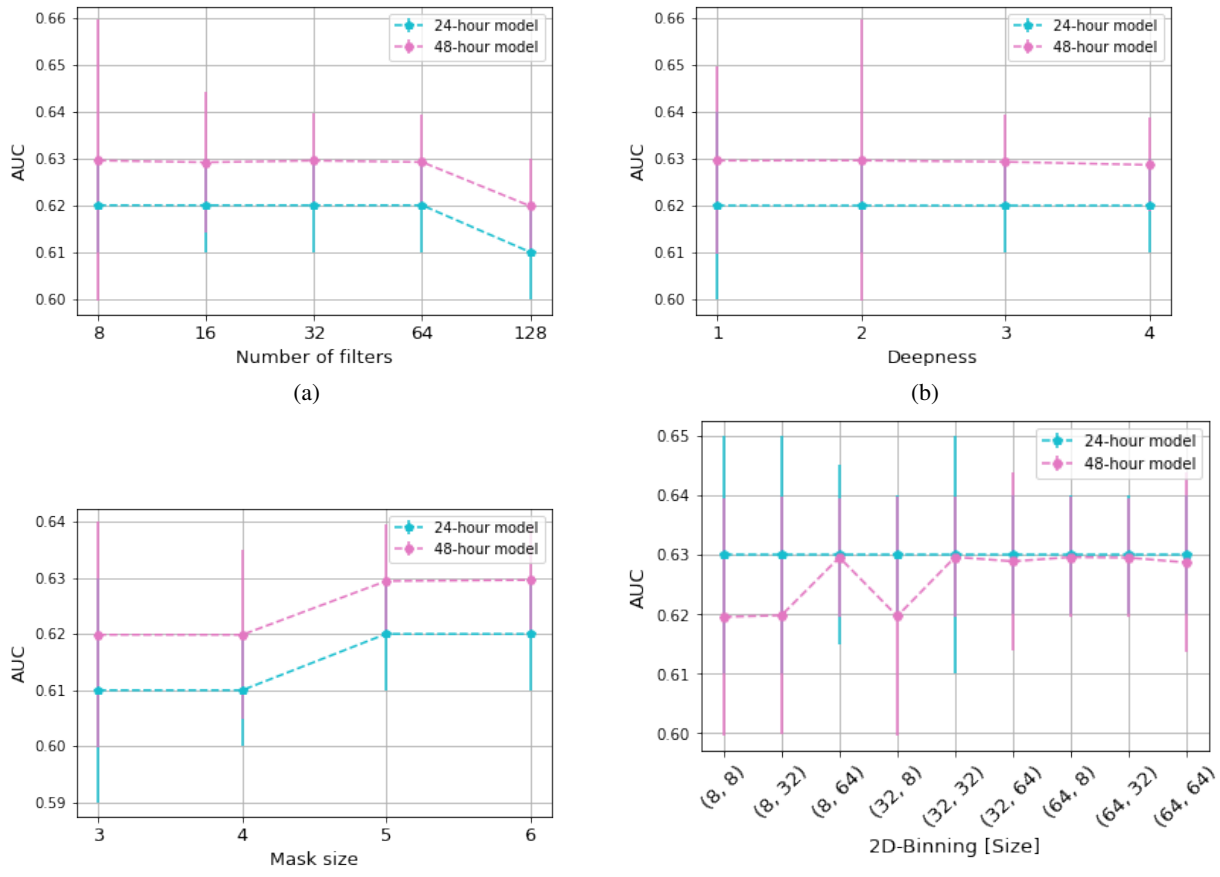


图 6：二维卷积神经网络模型。最大曲线下面积随超参数滤波器数量 (a)、深度(b)、卷积核大小 (c)以及二维分箱的维数 (d)的变化。每个图均展示了24小时模型（青色）和48小时模型（粉色）的表现。

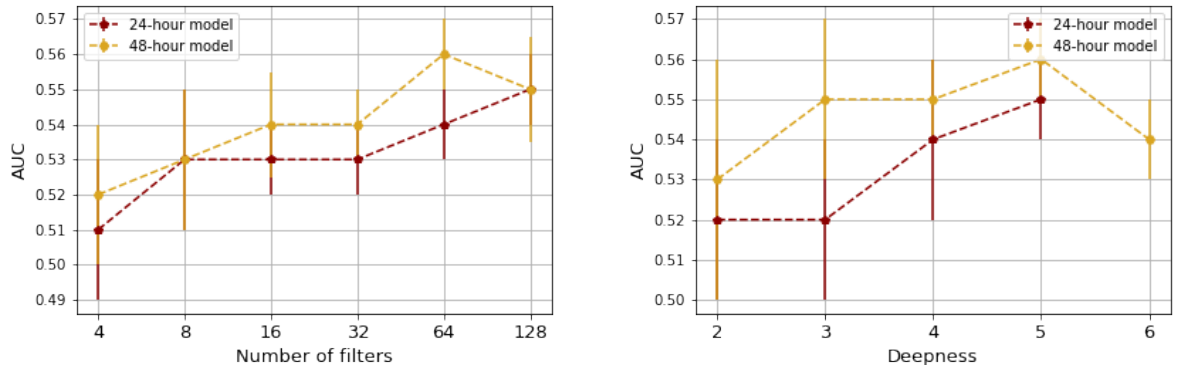
占位值；这样，缺失时间间隔的插补模式也可以包含重症监护病房住院期间发生的、值得关注的不同临床事件。与模型选择一样，我们对超参数进行了精细网格搜索。具体而言，我们展示了三个影响最大的参数的结果，即滤波器数量、卷积核大小（或掩码大小）和深度。图 7 和图 8 展示了最大曲线下面积。我们强调，对所有插补方法对卷积神经网络预测的影响进行考察时，采用了两种不同的预测窗口长度：24小时和48小时。我们选择这一策略，是因为我们旨在量化所考虑的插补方法在不同时间尺度下分析数据的纵向演变时是否确实有益。从表 4 中我们观察到，对于所考虑的两种时间尺度，即24小时和48小时（表 4 中我们考虑了24小时和48小时模型），末次观测值结转法最有前景。它与其他方法之间的差距很明显。对于24小时模型，k近邻和ICUAI插补方法的准确率均未能超过LOCF插补方法所取得的准确率。这一结果表明，末次观测值结转法是数据插补的首选方法。

2.3 选定的设计

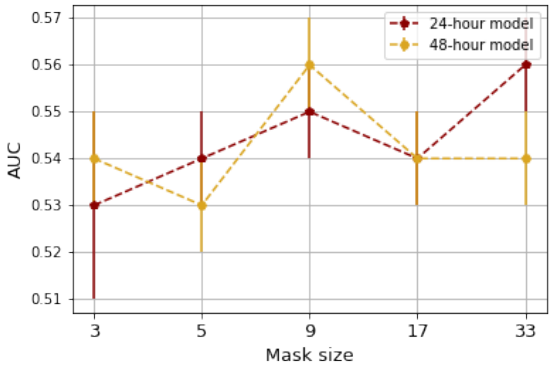
在所考虑的所有模型中，一维卷积神经网络模型并非预测性能最高的模型。尽管 24 小时卷积神经网络-长短期记忆网络模型的准确度可能略高于卷积神经网络模型，但我们观察到，即使使用 48 小时时间序列实例，后者的预测表现也更加精确。对于 24 小时模型，两种模型在曲线下面积方面的差异微乎其微；而对于 48 小时模型，这一差异则较为明显。此外，我们选择一维卷积神经网络模型，也是因为本研究旨在借助 SMOE 量表这一稳健的可解释人工智能方法，解释模式识别活动（见第 4 节）。

插补方法	曲线下面积（24小时模型）	曲线下面积（48小时模型）
LOCF	0.72 ± 0.01	0.68 ± 0.01
kNN	0.63 ± 0.01	0.65 ± 0.01
ICUAI插补	0.56 ± 0.01	0.56 ± 0.01

表 4：插补方法摘要：报告了各插补方法在验证阶段取得的最高性能，以曲线下面积衡量。显示曲线下面积分数的列分别对应 24 小时实例模型或 48 小时实例模型。曲线下面积分数已四舍五入至小数点后两位。误差采用标准误均值进行评估；若误差过小，则以最小误差 0.01 替代。



(a) (b)



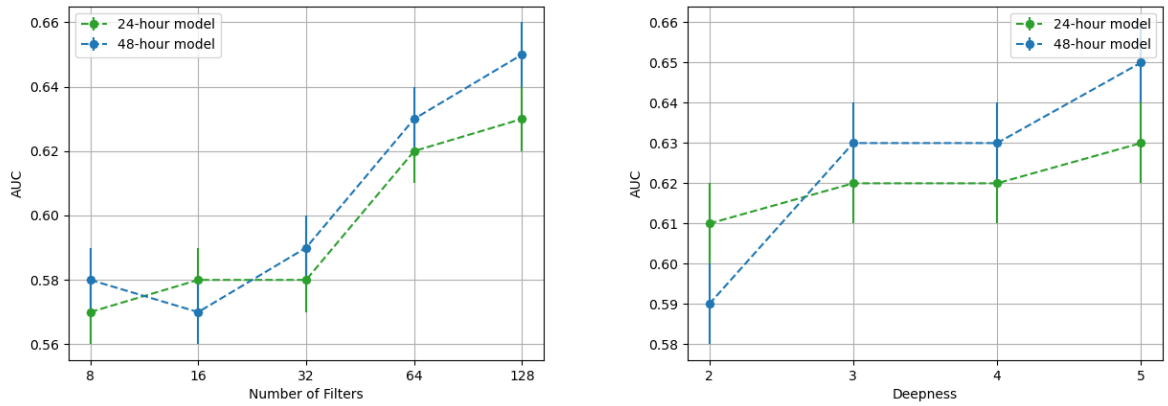
(c)

图7：由经 ICUAI 插补方法插补的电子健康记录数据输入的一维卷积神经网络模型。一维卷积神经网络模型的最大曲线下面积与其超参数的关系。(a) 滤波器数量，(b)深度 (b)，(c) 卷积核大小。每幅图均展示了24小时模型和48小时模型的表现。

对于 24小时模型的 m odel，我们提出以下最优架构 ：

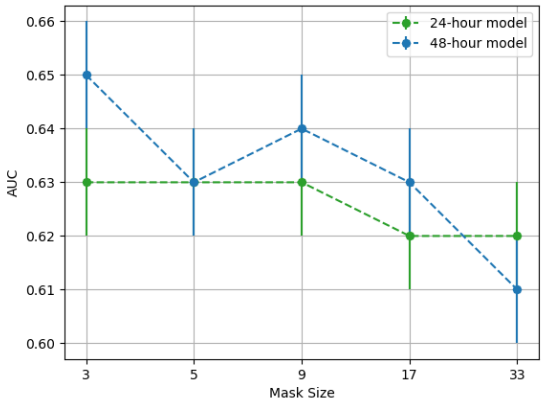
1. 卷积层：每层的滤波器数量为 128，每个滤波器的大小为 3（像素）。这些卷积的结果称为特征图。
2. 激活层：每个卷积算子之后应用 ReLU 函数。将非线性激活函数应用于特征图后，便生成激活特征图。3. 最大池化层：通过池化大小为 2 的最大池化算子对激活特征图进行重采样。

这组隐藏层重复五次。该架构还在每个最大池化层后设置一个丢弃层 。丢弃层的丢弃率为 0.25。最后一个特征图被展平为一个数组，然后传入一个全连接 层（稠密层），并使用 Sigmoid 激活函数。该激活函数



(a)

(b)



(c)

图8：采用 k近邻插补方法对电子健康记录数据进行插补后输入的一维卷积神经网络模型。一维卷积神经网络模型的最大曲线下面积随超参数变化的情况。(a) 滤波器数量，(b) 深度，(c) 卷积核大小。每幅图均展示了24小时模型和48小时模型的表现。

会返回一个介于 0 和 1 之间的正值，即风险评分，表示患者发生 ICU-AI发作的可能性。与通常情况一样，损失函数为二元交叉熵，优化器为 ADAM 算法。48小时模型的架构与24小时模型相同，但卷积层和最大池化层的序列重复了6次。

3 基于人工神经网络的预测：ICU-AI

除了本文提出的两步策略外，我们还考虑了一种完全基于人工神经网络的方法。我们希望了解完整的深度学习模型在预测获得性感染发生方面能发挥多大的预测能力，并比较这两种方法。

因此，为了使基于人工神经网络的预测能够与（深度）LM-CR 得出的预测进行比较，我们选择在每个标志时间拟合一个人工神经网络模型，同时输入所有仍处于院内感染风险中的患者的各类数据（即高频、低频和基线数据）。与（深度）LM-CR 一样，我们只关注首次感染事件。此外，我们训练人工神经网络模型时采用了 24 小时的提前时间；也就是说，假设使用标志时间 τ 时可获得的数据来训练该人工神经网络模型，响应变量将是一个二分变量，用于标记在 $\tau + 24$ 小时发生院内感染。 如前所述，该人工神经网络模型旨在同时输入不同类型的数据，例如时间序列（高频数据）

以及数据矩阵（低频数据和基线数据）。为适应数据集的这一特点，我们选择开发一种双分支人工神经网络 （TB-ANN）。
因此，该模型的架构由两个不同的分支组成，分别用于分析时间序列或数据矩阵；最终，两者的潜在表示会传入同一个预测层。用于分析时间序列的分支与为 CNN 风险评分开发的一维卷积神经网络非常相似：

1. 输入时间序列依次经过一系列隐藏层，即卷积层、激活函数、最大池化和丢弃层。卷积层包含 128 个滤波器，卷积核大小为 3，最大池化大小为 2，丢弃率为 0.2。激活函数为有理激活函数（Boull´e et al., 2020）。
2. 随后，将这一系列隐藏层重复 3 次。
3. 特征图经过一个包含 32 个输出单元的长短期记忆层处理，这些单元表示时间序列的潜在表示。

类似地，处理数据矩阵的分支是一个包含 3 个全连接层的多层感知机，每层有 32 个输出单元；激活函数仍为有理激活函数。因此，数据矩阵经过这一系列全连接层后，即可得到数据矩阵的潜在表示，其为一个维度为 32 的数组。最后，将这两个潜在表示拼接起来，并传入一个具有单个输出且使用 Sigmoid 激活函数的稠密层。与其他人工神经网络一样，TB-ANN 通过 ADAM 算法进行训练，以优化二元交叉熵损失函数。图 9 展示了 TB-ANN 的示意图。
对于第 2 节中讨论的模型，我们也采用了细网格搜索来确定最终模型的架构。因此，我们优化了两个分支各自的参数，以及由这两个分支得到的潜在表示的维度；通过在标志时间范围内最大化总体曲线下面积，搜索最优超参数。也就是说，我们使用相同的架构来验证不同的 TB-ANN，每个 TB-ANN 都在一个不同的标志时间进行训练。我们以总体曲线下面积作为评估模型架构优劣的指标；该指标是不同标志时间下按风险集加权的曲线下面积均值（参见稿件中的相关章节）。为了将 TB-ANN 与 Deep LM-CR 进行比较，我们采用了与稿件中所述类似的策略。因此，我们通过 AUROC 指标评估了 TB-ANN 预测的准确性。作为次要指标，我们还考虑了布里尔分数。众所周知，布里尔分数需要使用预测概率，而非预测分数；因此，我们采用了 Platt 缩放（Platt 等人，1999），使 TB-ANN 的输出能够与概率相比较。与 Deep-CR-LM 模型类似，总体指标通过对每个标志时间的风险集取均值进行评估，95% 置信区间则通过自助重采样进行评估。

4 显著性图顺序等价（SMOE）尺度

本节介绍本文中用于估计显著图的算法。与基于梯度的方法（例如原始梯度（VG），可参见 Simonyan 等人（2013））不同，SMOE 量表（Mundhenk 等人，2019年）为估计卷积神经网络激活特征图的显著性提供了不同的视角。SMOE 量表关注这些特征图激活的统计特性。

该算法能够（相当忠实地）表示输入数据中包含的信息：特征图的整体激活越强，输入特征就越可能包含有用信息。考虑一个卷积神经网络模型，我们用 $\chi_{ij} \in \mathbb{R}^+$ 表示经 ReLU 激活函数处理后的激活特征图的取值，并分别用 i 和 j 表示空间域和深度（即时间序列特征的数量）。函数φ： $\mathbb{R}^+ \to \mathbb{R}^+$ 应用于空间域中的每个点，并遍历整个深度维度。因此，我们通过关系式 $S = \varphi(\chi)$ 得到显著图。我们假设激活特征图 χ 服从形状参数为 k 、尺度参数为 θ 的伽马分布。作此假设的原因在于，对于均值和熵固定的随机变量（拉格朗日乘子），伽马分布是最大熵概率分布。鉴于在我们的情境中，每个激活图都具有固定的均值（即激活图的尺度）和固定的熵（即特征图中捕获的信息），因此选择伽马分布来描述特征图似乎是合理的。因此，我们采用最大似然原理估计分布参数，即：

$$\hat{\theta}_i = \frac{\sum_{j=0}^{D} \chi_{ij}}{D\hat{k}_i}, \tag{1}$$

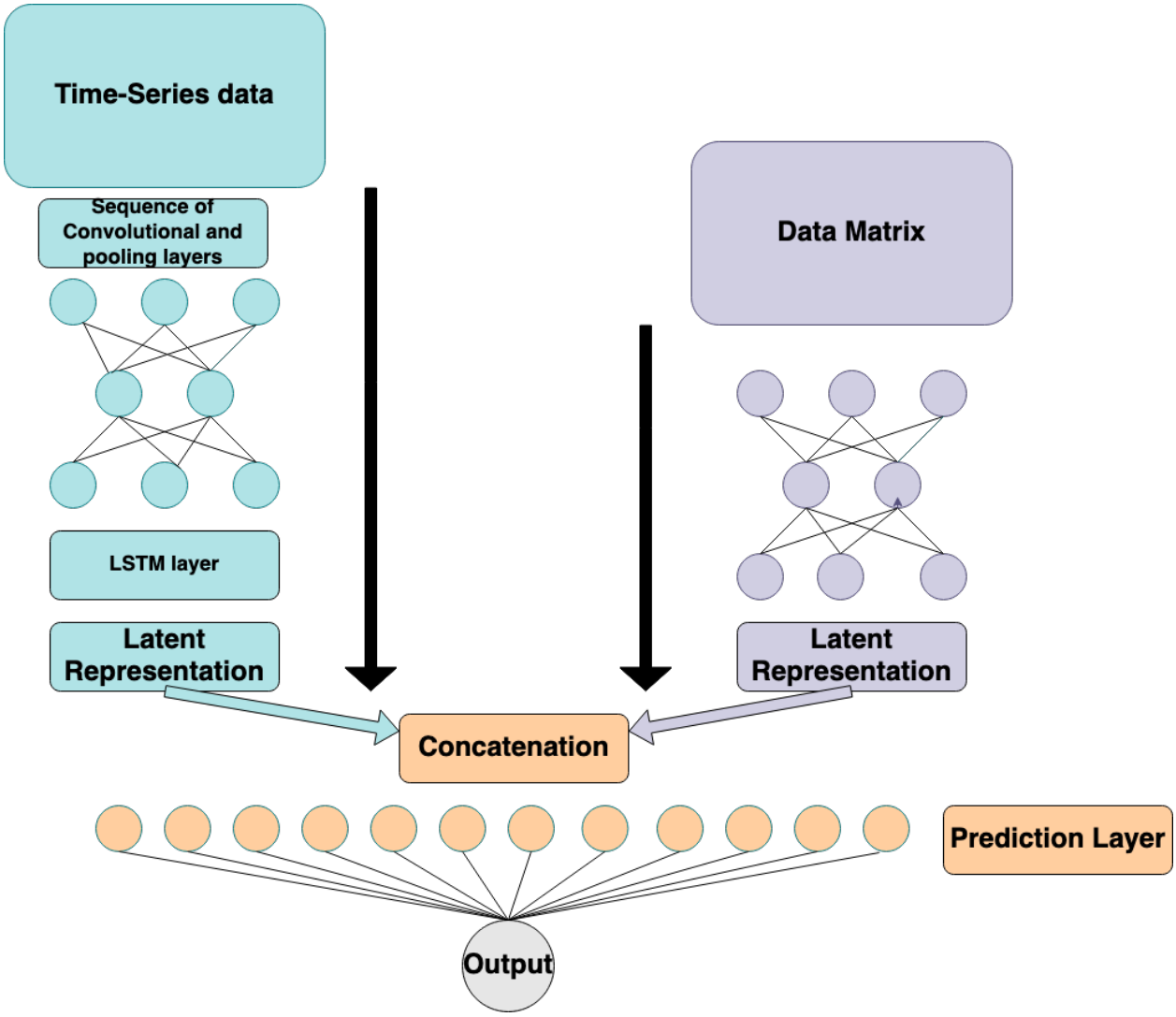


图 9：TB-ANN 示意图。右侧分支（浅蓝色）专门用于编码时间序列特征；与一维卷积神经网络类似，它由按顺序排列的卷积层和池化层组成。长短期记忆层紧随其后。左侧分支（浅橙色）专门用于编码数据矩阵，采用类似多层感知机的结构。每个分支都会生成一个潜在表示，并将其传递至拼接层。最后，经过预测层（带有 Sigmoid 激活函数的稠密层），即可得到所需的输出。

and

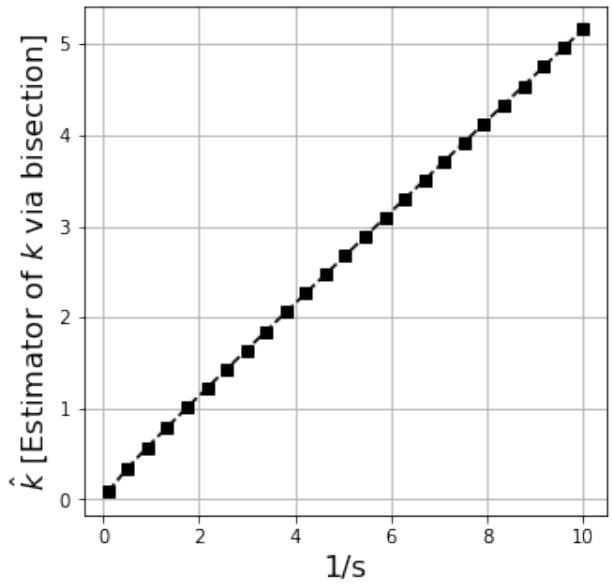
$$\log\left(\hat{k}_i\right) - \psi(\hat{k}_i) = \log\left(\frac{\sum_{j=0}^{D} \chi_{ij}}{D}\right) - \frac{\sum_{j=0}^{D} \log \chi_{ij}}{D};$$

求和范围为深度域（即输入特征的域），其中 D 为输入特征的数量， $\psi(x)$ 为双伽马函数（Silverman 等人，1972年）。我们回顾一下，双伽马函数定义如下：

$$\psi(x) = \frac{d \log \Gamma(x)}{dx},$$

其中 $\Gamma(x)$表示欧拉伽马函数（Silverman 等人，1972年）。

请注意，参数 θ_i 和 k_i 的估计仅限于激活特征图 χ的空间域中沿深度域的第 i个元素：我们提取的是沿深度域的激活稀疏性信息，而



ˆ 图10：通过二分法估计的 k 与数值 $1/s$的函数关系

不沿着　　空间域。不过，我们无法以闭式求得 k_i 的估计值，但可以令：

$$s_i = \log\left(\frac{\sum_{j=0}^{D}\chi_{ij}}{D}\right) - \frac{\sum_{j=0}^{D}\log\chi_{ij}}{D};$$

然后，我们使用双伽马函数的渐近展开（Abramowitz 和 Stegun，1964），得到以下近似结果：

$$\log x - \psi(x) \simeq \frac{1}{2x}\left(1 + \frac{1}{6x}\right).$$

ˆ 因此， k 的一阶近似为：

$$\hat{k}_i \simeq \frac{\frac{1}{4} + \frac{1}{2}\sqrt{1+3s_i}}{s_i}. \tag{2}$$

ˆ 然而，我们可以使用牛顿-拉夫森法（Ypma，1995）对 k_i 进行细化，并将（2）作为初始值。结果表明，ˆ k_i 似乎与 $\frac{1}{s_i}$ 相关；如图10所示。将（1）中的 $\frac{1}{s_i}$ 替换为ˆ k_i 后，我们得到：

$$\hat{\theta}_i = \left(\frac{\sum_j^D \chi_{ij}}{D}\right)\left[\log\left(\frac{\sum_j^D \chi_{ij}}{D}\right) - \sum_{j=0}^{D}\frac{\log\chi_{ij}}{D}\right],$$

最终可改写为：

$$\hat{\theta}_{\text{SMOE},i} = \frac{1}{D}\sum_{j=0}^{D}\langle\chi\rangle\log\frac{\langle\chi\rangle}{\chi_{ij}}, \tag{3}$$

其中

$$\langle\chi\rangle := \frac{1}{D}\sum_{m=0}^{D}\chi_{lm}.$$

因此，(3) 表示*SMOE* 量表，即计算显著性图所涉及的统计量：它通过项 $\langle\chi\rangle$ 与激活均值（沿深度方向）成正比。此外，我们对 $\log\chi_{ij}$ 在 $\langle\chi\rangle$处进行泰勒展开后可以看出，它还取决于方差：

$$\log\langle\chi\rangle - \langle\log\chi\rangle \simeq \frac{\langle\chi - \langle\chi\rangle\rangle^2}{2\langle\chi^2\rangle}.$$

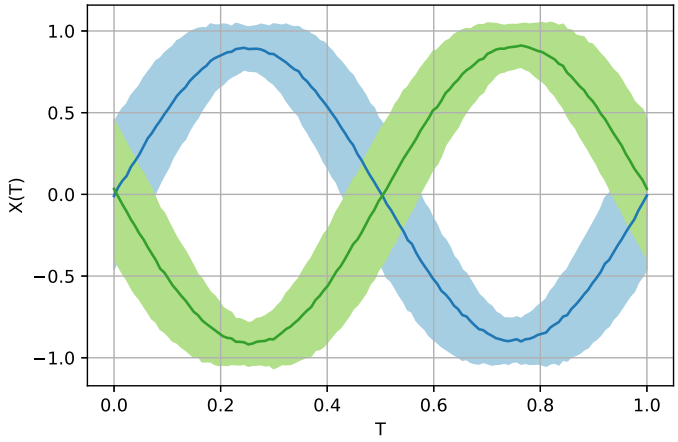


图 11：玩具正弦数据集示例。对于类别 0 和 1，均值样本分别以蓝色和绿色绘制。浅色区域表示均值样本的标准差值。

在估计伽马尺度参数时采用的简化方法，是相较于完整迭代尺度参数估计（Mundhenk 等，2019年）的 SMOE。

根据构造方式，我们只能将 SMOE 量表应用于单个激活特征图；也就是说，我们只能分别估计每个激活特征图的信息稀疏性。随后，我们可以将它们结合起来，获得每个时空位置的显著性总体测量值。

因此，所使用的 SMOE 量表是伽马分布的估计尺度参数。在主稿件的第 6 节中，我们利用这一假设推导了显著图。然而，我们通过多重柯尔莫哥洛夫-斯米尔诺夫检验并采用邦费罗尼校正，对该假设进行了检验：对于每个激活图以及给定的 ˆ 时间域值，我们检验特征值是否服从尺度参数为 θ_i（如公式 (3) 中所估计）的伽马分布。我们未以 $\alpha = 0.05$ 拒绝零假设。

在结束本节之前，我们想简要举例说明上文介绍的可解释人工智能方法。考虑以下用于时间序列二元分类的玩具数据集：类别 0 由以下方式生成：

$$X_0^{(n)}(t) = \sin\left(2\pi[t + \phi_n]\right),\ t \in [0, 1];$$

其中 $X_0^{(n)}(t)$表示 第 n个类别 0 实例，且 $\phi_n \overset{i.i.d}{\sim} U(-0.125,\ 0.125)$。同样，对于类别 1，我们设 t:

$$X_1^{(n)}(t) = -\sin\left(2\pi[t + \phi_n]\right),\ t \in [0, 1].$$

我们将此数据集称为玩具正弦数据集。玩具正弦数据集的示意图见图 11。我们采用单次留出法训练和测试一个卷积神经网络（即仅将数据集划分为训练集和测试集一次，并在此后评估模型的准确率），训练集大小为 75%（即使用数据集的 75% 来训练卷积神经网络模型）。结果显示，模型的曲线下面积为 0.99。

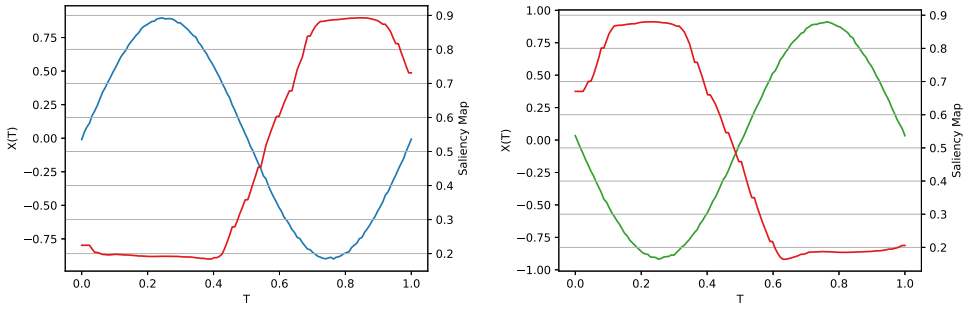
图 12 展示了 SMOE 量表的显著性图。我们注意到，这些显著性图在正弦振荡的波谷处检测到显著区域，即输入域中显著性图取值最高的那些区域。正如预期，波谷在输入域的两个不同区域中的位置$T \in [0, 1]$ 代表了卷积神经网络用于区分两个类别的关键特征。

参考文献

Abramowitz, M. 和 Stegun, I. A.（1964）。数学函数手册：含公式、图表和数学表，第 55 卷。美国政府印刷局。

Boull´e, N.、Nakatsukasa, Y. 和 Townsend, A.（2020）。有理神经网络。神经信息处理系统进展，33：14243–14253。

Chen, Y. and Qi, B. (2019). 用于心力衰竭风险预测的术中生命体征表征学习。*BMC*医学信息学与决策制定, 19(1):1–15.



(a) (b)

图 12：通过 SMOE Scale 方法获得的正弦玩具数据集的显著图。(a) 类别 0 的均值样本（蓝色）和均值显著图（红色）。(b) 类别 1 的均值样本（蓝色）和均值显著图（红色）。

Keogh, E. J. and Pazzani, M. J. (2000). 扩展动态时间规整在数据挖掘应用中的规模。载于第六届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，第 285–289 页。Mundhenk, T. N., Chen, B. Y., and Friedland, G. (2019). 面向可解释人工智能的高效显著性图。*arXiv* 预印本 *arXiv:1911.11293*。Platt, J. et al. (1999). 支持向量机的概率输出及其与正则化似然方法的比较。大间隔分类器进展, 10(3):61–74。Silverman, R. A. et al. (1972). 特殊函数及其应用。Courier Corporation。Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). 深入卷积网络内部：图像分类模型与显著性图的可视化。*arXiv* 预印本 *arXiv:1312.6034*。Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). DNA 微阵列的缺失值估算方法。《生物信息学》, 17(6):520–525。Ye, Y., Jiang, J., Ge, B., Dou, Y., and Yang, K. (2019). 使用网格表示和矩阵距离进行时间序列数据分类的相似性度量。知识与信息系统, 60(2):1105–1134。Ypma, T. J. (1995). 牛顿–拉夫森法的历史发展。*SIAM Review*, 37(4):531–551.