连续时间循环神经网络：概述及其在重症监护室血糖预测中的应用

Oisin Fitzgerald[a, *]，Oscar Perez-Concha[a]，Blanca Gallego-Luxan[a]，Alejandro Metke- Jimenez[b]，Lachlan Rudd[c]，Louisa Jorm[a]

[a] 澳大利亚新南威尔士州悉尼，UNSW Sydney，AGSM 大楼 2 层，健康大数据研究中心，邮编 2052
[b] 澳大利亚昆士兰州赫斯顿 Herston Road 296 号，STARS 大楼 7 层（外科治疗与康复服务中心），澳大利亚电子健康研究中心，邮编 4029[c] 新南威尔士州圣伦纳兹 Reserve Road 1 号，eHealth NSW 数据与分析部，邮编 2065 * 通讯作者：o.fitzgerald@unsw.edu.au

摘要

在许多以时间序列建模为关键统计工具的应用场景中（包括医学领域），不规则测量的时间序列十分常见。这给模型选择带来了挑战，往往需要采用插补或类似策略。连续时间自回归循环神经网络（CTRNN）是一种深度学习模型，它通过在观测值之间纳入隐藏状态的连续演化来处理不规则观测。这是通过使用神经常微分方程（ODE）或神经流层实现的。本文概述了这类模型，包括为应对持续进行的医疗干预等问题而提出的不同架构。此外，我们展示了如何在重症监护场景中，利用电子病历和模拟数据，将这些模型应用于血糖概率预测。实验结果证实，在测量不规则的情况下，添加神经常微分方程层或神经流层通常能提升自回归循环神经网络的性能。然而，若干种 CTRNN 架构的表现不及自回归梯度提升树模型（CatBoost）；只有长短期记忆网络（LSTM）和基于神经常微分方程的架构（ODE-LSTM）在连续排序概率评分（ODE-LSTM：0.118±0.001；CatBoost：0.118±0.001）、无知评分（0.152±0.008；0.149±0.002）和区间评分（175±1；176±1）等概率预测指标上取得了相当的表现。

关键词

深度学习；电子病历；时间序列；预测；血糖控制

引言测量时间不规则

不规则时间序列在许多应用场景中都很常见，在这些场景中，时间序列建模是关键的统计工具，医学领域也不例外。电子病历（EMR）等观察性医疗数据通常并非为研究目的而收集或设计，而是用于支持临床诊疗和提高行政效率 [1, 2]。在患者住院期间，不同患者之间，甚至看似相似的患者之间，测量的时间安排（或是否进行测量）可能存在很大差异 [2]。尽管存在这些局限，电子病历的丰富细节和日益普及，使其成为研究的重要组成部分，推动研究者开发数据驱动型工具，以用于精准／个性化医疗 [1, 3]。电子病历通常具有高维、稀疏的特点，特征之间还可能存在复杂的相互关系，因此，深度学习等数据自适应方法已成为现代医学预测建模的核心工具 [4, 5]。用于处理时间因素的关键方法包括循环神经网络（RNN）、卷积神经网络和注意力网络 [6]。通常，这些方法将时间视为离散变量，因此需要通过算法调整来处理测量之间具有信息意义的时间间隔 [7]。然而，深度学习的最新进展提供了一些方法，可通过对连续时间过程进行建模，自然地克服不规则时间序列带来的问题。

连续时间预测

循环神经网络及相关深度学习架构在序列数据建模方面有着悠久的成功应用历史 [8]。从宏观上看，循环神经网络通过引入一个隐藏（或潜在）状态来发挥作用，该状态充当对过去事件的记忆 [9]。随着新的观测值出现，隐藏状态会更新，以便更好地纳入新信息。然而，常见的循环神经网络模型通常假设时间步长是离散的，且感兴趣事件的各次观测之间的时间间隔不包含有效信息，同时隐藏状态在观测之间保持不变。近年来，人们提出了多种方法，例如神经常微分方程（常微分方程）和神经流 [10, 11], ，使模型能够明确考虑各次测量之间的时间间隔，并以可根据数据灵活学习的方式进行建模。

重症监护室的血糖控制

我们展示了如何运用这些方法对重症监护室（ICU）住院患者的血糖进行建模。血糖控制是重症监护室护理的重要组成部分 [12, 13]。患者入住重症监护室时，9-29% 的患者血糖偏高（应激性高血糖；高于 180 mg/dL），2-6% 的患者血糖过低（低血糖；低于 70 mg/dL） [14]。在检测到应激性高血糖的病例中，约 40% 会使用胰岛素 [14]。不同患者使用胰岛素的情况各异，加上血糖检测普遍采用床旁检测，导致重症监护室中的血糖测量频率及测量间隔差异很大。预测未来的血糖值可能有助于识别低血糖或高血糖风险患者。以往研究通常采用离散时间处理 [15] ，或改用床旁检测以外的方法（例如连续血糖监测）[16-19], ，这限制了研究结果在真实世界场景中的普适性。

文章的目标与结构

本文旨在概述用于连续时间序列建模的自回归循环神经网络。我们重点关注这类模型在医疗场景中常见的纵向数据结构中的应用，以及能够实时生成概率预测的模型。

文章结构如下：首先，我们概述神经常微分方程和神经流，以及它们如何将动力系统与深度学习联系起来。其次，我们介绍这些模型的一般计算形式和数学形式、该形式的各种变体，以及与离散时间模型相比，训练连续时间模型时需要考虑的其他因素。接下来，我们展示如何使用这些模型预测重症监护室中的血糖，并详细介绍用于评估概率预测的若干指标。最后，我们对这些方法进行批判性评估，并探讨未来的研究方向。

背景

问题陈述许多生理时间

序列（例如血糖）非常适合用随机微分方程建模，这是微分方程的连续时间噪声版本[20]。理论上，这些序列可以以任意高的频率进行测量，其数值持续变化；由于难以考虑影响其演变的所有因素，因此认为它们部分受噪声驱动。在研究和建模此类时间序列时，通常可以将受试者 i 的观测时刻 $\tilde{\boldsymbol{t}}_i$ 映射为离散值。例如，如果使用连续血糖监测仪每隔 30 秒采集一次血糖数据进行建模，那么每次测量后时间都可以增加一个单位。然而，我们考虑的是测量过程相当不规则的情况，因此，任何时间离散化（若不采用插补或相关方法）都难以说具有明确意义。我们介绍的方法通过将时间序列的分布建模为带跳跃的连续时间过程来解决这一问题。

我们的目标是：给定每位受试者 i 的 N 总人数中的 $K_i \times D$ 维患者病史 $\widetilde{\boldsymbol{h}}_i$，预测每位受试者连续目标变量 y_i（t）的分布。该病史在K_i 处观测，观测时间点间隔不规则，由观测时刻向量 $\tilde{\boldsymbol{t}}_i$ 指定，其最大值为 T_i。给定病史后，该分布是时间 $t > T_i$ 的函数

$$p_\theta\left(y_i(t)\middle|\widetilde{\boldsymbol{h}}_i\right) = p_\theta\left(y_i(t)\middle|\widetilde{\boldsymbol{h}}_i = \{\boldsymbol{y_i}[\tilde{t}_i], \boldsymbol{x_i}[\tilde{t}_i], \boldsymbol{x_{0,i}}\}\right)$$

其中，p_θ 是参数为$\boldsymbol{\theta}$的预测分布，$\widetilde{\boldsymbol{h}}_i = \{\boldsymbol{y}_i[\tilde{t}_i], \boldsymbol{x}_i[\tilde{t}_i], \boldsymbol{x_{0,i}}\}$是不规则采样的患者病史，其中记号 $\boldsymbol{y}_i[\tilde{t}_i]$ 表示在$\tilde{t}_i$中的每个时间点，目标y（t）此前观测值的拼接。此外，$\boldsymbol{x_i}[\tilde{t}_i]$ 包含潜在协变过程$\boldsymbol{x}$（t）（例如血压）的观测值，而$\boldsymbol{x_0}$、$_i$ 包含静态特征（例如患者的糖尿病状态）。正如我们将在下文实际看到的那样，在每次希望进行预测时，我们不会传入完整的病史，而是使用隐状态来学习对过去的必要依赖程度，只传入最近的一组观测值，记为$\boldsymbol{h}_{T_i}$、$_i = \{\boldsymbol{y}_i[T_i], \boldsymbol{x}_i[T_i]\}$。对$p_{\boldsymbol{\theta}}$（$y_i$（$t$）$\left|\widetilde{\boldsymbol{h}}_i\right.$）的建模，是通过使用神经网络对$\boldsymbol{\theta}$ 的每个元素进行建模来实现的。例如，如果假设y_i（t）服从正态分布，我们会生成如下$\boldsymbol{\theta}$（t、$\widetilde{\boldsymbol{h}}_i$）=（$F_\mu$（$\widetilde{\boldsymbol{h}}_i$、$t$）、$F_\sigma$（$\widetilde{\boldsymbol{h}}_i$、$t$）），其中$F_\mu$ 和F_σ 是分别对均值和方差进行建模的神经网络。为简便起见，下文通常省略主体下标。

自回归循环神经网络我们讨论的模型大体上可以看作是对

将自回归（带外生输入）循环神经网络扩展到连续时间设定[6, 21] （图1）。自回归模型会将结果变量$y(t)$的先前观测值（以及可能存在的外生变量）作为下一组预测的输入。循环指的是将先前的模型输出，或更常见地，将映射到输出的内部状态，作为下一组预测的输入。由于这种扩展

从 离散时间设定扩展到连续时间设定，依赖于神经常微分方程和神经流的使用，因此我们先简要回顾这些方法。

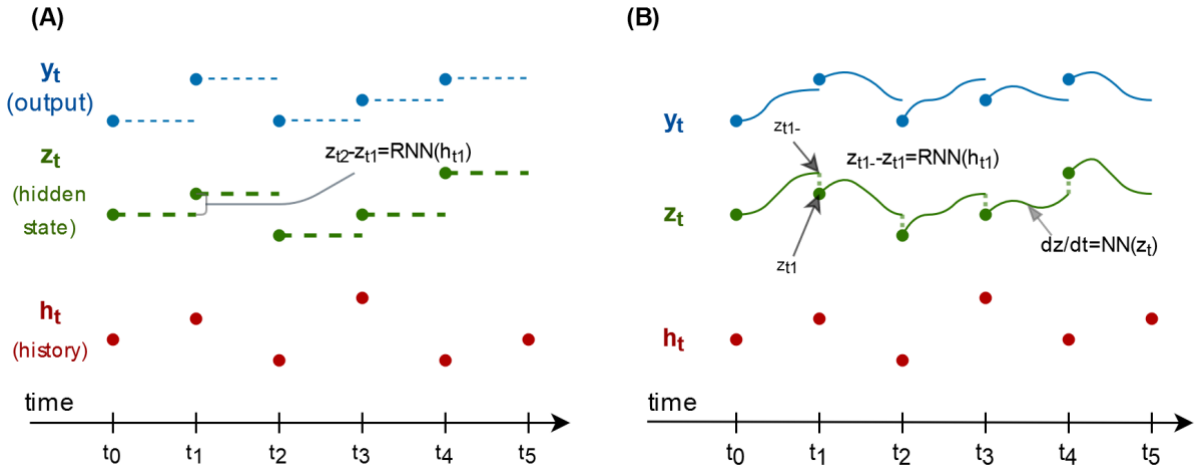

图1。（A）离散自回归循环神经网络。离散时间循环神经网络是一个映射 $\mathrm{Z} \times \mathrm{R}^{K_i \times D} \rightarrow \mathrm{R}$，其预测和隐藏状态在时间步之间未定义（实际上可假设为静态）。对于不规则测量的时间序列，寻找对齐（映射 $\mathcal{A}: \boldsymbol{t}_i \rightarrow \mathrm{Z}$）可能需要插补或插值。（B）连续时间自回归循环神经网络。隐藏状态 $z(t)$ 根据神经网络，在时间步之间连续演化。连续时间模型可视为将建模不规则时间序列所需的插值/插补步骤融合到单个模型中。注意：符号 y_t 是 $y(t)$ 的简写

神经常微分方程与神经流我们考虑的模型具有一个隐藏

状态 $z(t) \in \mathcal{Z}$，它会随时间连续演化，并可能在获得新的观测时发生不连续更新，详情如下文所述。在两个观测时刻（t_0, t_1）之间，隐藏状态根据一组流曲线 $F_\Theta: \mathcal{Z} \times \mathrm{R} \rightarrow \mathcal{Z}$ 演化，其中 Θ 表示固定或从数据中学习得到的参数。对这种演化进行参数化的方法有很多，下面介绍两种使用神经网络的方法。在神经常微分方程 [10] 中，我们使用神经网络 NN_{vec}对与流曲线相关联的向量场进行建模。为了在时刻 t_1 使用 $z(t)$ 进行预测，我们求解常微分方程初值问题。

$$\frac{dz}{dt} = NN_{vec}\big(z(t)\big), \qquad z(t_0) = z_0$$
$$z(t_1) = z_0 + \int_{t_0}^{t_1} NN_{vec}\big(z(s)\big)ds$$

实际上，需要使用数值方法来计算 $z(t_1)$，通常记作ODESolve$(NN_{vec}, z(t), (t_0, t_1))$（见图1）。ODESolve 的第二个参数（在本例中仅为 $z(t)$）就是 NN_{vec} 的输入，并且可能包含除隐藏状态当前值之外的其他信息，具体如下文所述。

作为神经常微分方程的一种替代方案，神经流 [11] 使用神经网络直接对 F_Θ 的流曲线进行参数化，即 $z(t_1) = NN_{flow}(z_0, (t_0, t_1))$。为确保所学习的神经网络满足初始条件 $NN_{flow}(z_0, (t, t)) = z_0$，必须谨慎选择网络架构。虽然按上述定义，神经常微分方程和流在两种情况下都仅依赖于 $z(t)$ 和 (t_0, t_1)，但我们可以向神经网络传入额外信息，例如正在进行的药物输注速率。接下来，我们定义一组流

$F_{\mathrm{x},\ \phi}: \mathcal{Z} \times \mathrm{R} \rightarrow \mathcal{Z}$ 由数据 $x \in X$ 和 NN_{flow} 或 NN_{flow} 参数 $\phi \in \Phi$ 索引，其中一部分是固定的（例如层数），另一部分则从数据中学习（例如权重）。

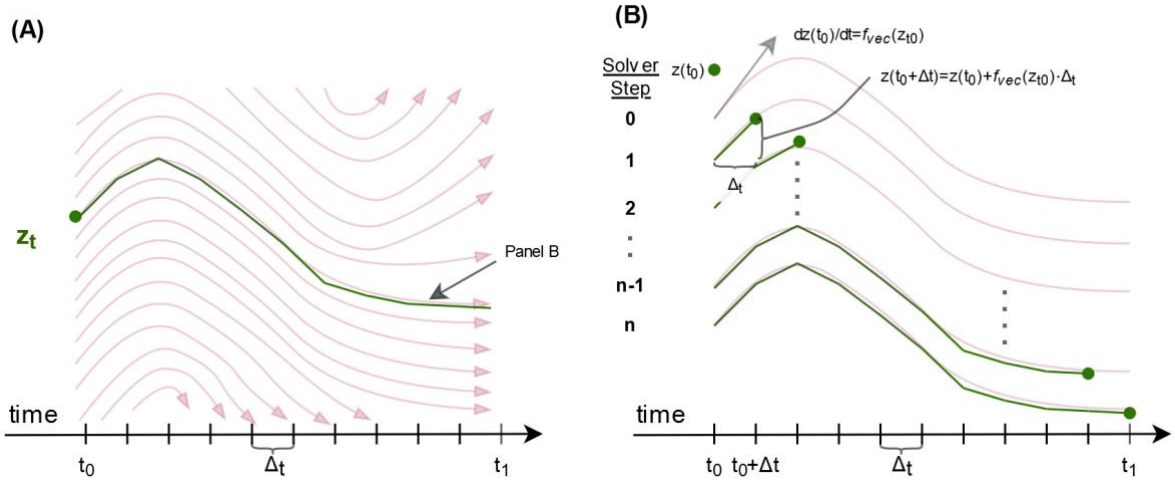

图 2.（神经）常微分方程求解过程示意图。（A）由神经网络 NN_{vec} 或 NN_{flow} 定义的流曲线。（B）用欧拉方法数值求解常微分方程初值问题的示意图——给定初始状态 $z(t_0)$ 和动力学 $dz/dt = NN_{vec}(z)$，这是求解神经常微分方程的一种方法。

连续时间自回归循环神经网络连续时间循环神经网络（CTRNN）已多次被提出

[22] 通常用于求解微分方程。近年来，其在医疗时间序列分析中的适用性受到关注 [10, 23]。它们是循环神经网络（RNN）[24] 的一种变体，在观测时刻之间使其隐藏状态连续演化（t_0, t_1），从而能够在任意时点进行预测。因此，它们非常适合我们所考虑的实时预测情形。需要注意的是，流不一定要由神经网络参数化，即 ϕ 可以定义一个指数衰减过程 [25] 或一组样条 [26]。给定结局分布 $P_{\boldsymbol{\theta}}$、神经网络：NN_{init}、NN_{jump}、NN_{out} 以及流 $F_{\mathrm{x},\ \phi}$，CTRNN 遵循通用的前向模式算法。

算法1. 连续时间自回归循环神经网络

i. 初始化： ii. 更新：使用新的观测向量 $\boldsymbol{h}_t$

set iii. 演化： iv. 预测：

$$z(t_0) = NN_{init}(x_0)$$

$$z(t) = NN_{jump}(\boldsymbol{h_t}, z(t-))$$

$$z(t + \Delta_t) = F_{\mathrm{x},\phi}(z(t), t + \Delta_t)$$

$$\boldsymbol{\theta} = NN_{out}(z(t + \Delta_t))$$

$$y(t + \Delta_t) \sim P_{\boldsymbol{\theta}}$$

请注意，我们将更新前一刻的隐藏状态值记为 $z(t-)$。从整体上看，该模型由以下几个部分组成：1）在出现新观测时对隐藏

状态进行不连续更新（更新）；2）使隐藏状态 $z(t)$ 随时间向前演化，并可能依赖于患者病史和时间信号的某些方面（演化）；3）根据需要，通过输出随时间变化的结局分布模型参数向量 $\theta(t)$，生成概率预测（预测）（图3）。这可以用函数复合表示为 $\boldsymbol{\theta}((t, t+\Delta_t), \boldsymbol{h}_i) = (NN_{out} \circ F_{\mathrm{x},\ \phi} \circ NN_{jump})(\boldsymbol{h}_i, (t, t+\Delta_t))$。

连续时间自回归神经网络的架构已有多种变体。它们的区别在于子网络的架构，以及传递给子网络的输入和隐状态。例如，在 [23] NN_{vec} 提出的 ODE-RNN 模型中，[23] NN_{vec} 是多层感知机，NN_{jump} 是循环神经网络（例如门控循环单元或普通循环神经网络）。相比之下，神经跳跃常微分方程 [27] 对 NN_{vec} 和 NN_{jump} 都使用多层感知机，其中 NN_{jump} 仅取决于输入向量，而不取决于当前隐藏状态。总体而言，有多种模型架构都符合算法1。下面我们将介绍并比较其中一些架构。

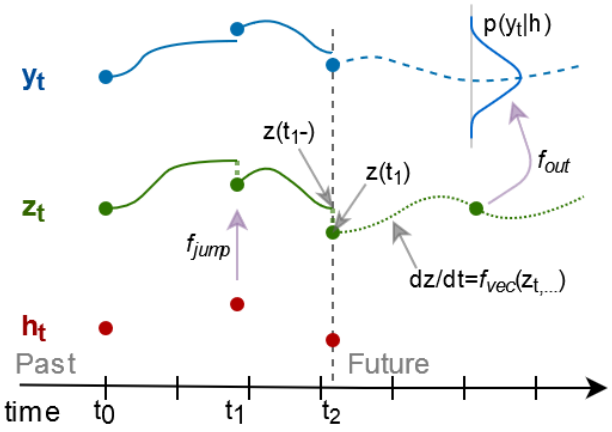

图3. 模型算法示意图当神经常微分方程.....。当历史记录 h_t 通过新的观测值更新时，NN_{jump} 会更新隐藏状态 z_t。在观测值之间，z_t 的状态根据 NN_{vec}演化。注：记号 y_t 是 $y(t)$ 的简写

目标函数我们的模型估计较大

随后是 [28] 和 [27],，可视为与卡尔曼滤波器中的更新/滤波和预测步骤类似。如 [21] 所述，分布 P_θ应符合数据的统计特性。采用适当的 P_θ 后，我们着手最小化演化/预测步骤中的误差，并最大化更新步骤中的拟合效果。因此，结局 $y(t)$ 的每个观测都会被使用两次。我们将更新前隐藏状态的值记为 $z(t-)$（即，在以自回归方式将数据点重新输入模型之前），更新后的值记为 $z(t)$（图3）。预测步骤的损失函数是数据的负对数似然

$$\mathrm{Loss}_{pred}(\boldsymbol{y}_i, \boldsymbol{\theta}) = -\sum_{i=1}^{N} \sum_{t=t_{0,i}}^{T_i} \log p_{\boldsymbol{\theta}}(y|z(t-))$$

其中 $z(t-)$ 是更新步骤之前的最终隐藏状态值。为确保更新步骤之后由新的隐藏状态值所隐含的结局分布 与观测到的数据点 $y(t)$ [27, 28] 所对应的损失函数为

$$\text{Loss}_{jump}(\boldsymbol{y}_i, \boldsymbol{\theta}) = -\sum_{i=1}^{N}\sum_{t=t_{0,i}}^{T_i} m_i \log p_{\boldsymbol{\theta}}(y|z(t))$$

其中，m_i 是一个二元掩码，当需要进行更新步骤时（例如，由于$x(t)$发生变化），但$y(t)$尚未测量或以其他方式缺失时，该掩码很有用。如果观测过程存在测量误差，我们可以使用库尔贝克-莱布勒散度作为损失函数来处理这一问题

$$\text{Loss}_{jump}(\boldsymbol{y}_i, \boldsymbol{\theta}) = -\sum_{i=1}^{N} \text{KL}(p_{\boldsymbol{\theta}}(y|z(t)) || p_{obs}(y; \sigma^2))$$

其中，$p_{obs}(y;\ \sigma_e^2)$表示给定观测值 y 时响应$y(t)$的假定分布，并假设测量误差的均值为零、方差为σ_e^2。

由此得到一个包含可调参数λ 的最终目标函数，该参数用于考虑损失幅度可能存在的差异。

$$\text{Loss} = \text{Loss}_{pred} + \lambda\, \text{Loss}_{jump}$$

模型训练连续时间自回归

循环神经网络可以使用基于自动微分软件包构建的标准深度学习训练算法进行训练。正如[29] 所指出的，包含神经常微分方程的模型可以通过时间反向传播（BPTT）[24] 或伴随方法[10]与网络的其余部分联合训练。伴随方法通过一个新的常微分方程——伴随方程——计算损失相对于NN_{vec} 参数的导数，并将其传递给 ODESolve（并“反向”运行）。由于无需存储 ODESolve 前向计算过程中的中间状态，其内存开销低于 BPTT。

协变量与干预如上所述，我们通常会有

随时间变化的协变量或静态协变量作为输入，以及自回归分量。此外，其中一些协变量最好视为干预（例如医疗治疗），旨在控制 $y(t)$ 的取值。随时间变化的协变量过程 $\boldsymbol{x}(t)$ 可视为 $\boldsymbol{x}(t) = (\boldsymbol{x}_{input}(t), \boldsymbol{x}_{tv}(t))$ 的组合，其中 $\boldsymbol{x}_{input}(t)$ 是干预（例如胰岛素），而 $\boldsymbol{x}_{tv}(t)$ 是与 $y(t)$ 的水平相关、但不会控制其水平的时变特征。先前的研究指出，对治疗过程 [30] 进行非常明确的建模可能会带来更好的性能，我们将在下文对此进行探讨。

其他考量加入连续 dy

将动态过程加入自回归循环神经网络模型，相较于离散时间情形会带来额外的复杂性。与 $F_{x,\ \phi}$ 相关的向量场必须是 Lipschitz 连续的，才能真正描述连续动力学[11]。可能影响这一点的因素包括 ODE 求解器的选择，以及NN_{vec} 或NN_{flow}的架构。本文不讨论 ODE 求解器方法的综述（参见[31, 32]），但会简要说明几点。ODE 求解器的选择会影响稳健性以及训练和评估所需的时间。例如，过度离散化可能会

将解转化为离散残差网络（ResNet），而其性质不同于连续向量场[33]。关于NN_{vec}，我们通常考虑其为多层感知机或门控循环单元的情况。鉴于一次 ODESolve 调用的输出是对NN_{vec}的多次调用进行加权求和，建议将最后一层的输出限制在一定范围内（例如使用双曲正切非线性），以避免数值上溢或下溢。对于神经流模型NN_{flow}，使用门控循环单元或可逆残差网络作为神经流模型，可确保满足初始条件和连续性要求[11]。

模型架构连续时间自回归

循环神经网络（算法1）支持多种特定架构。下面我们介绍几种不同的架构，然后比较它们在血糖预测任务中的表现。

基于神经常微分方程的架构ODE-RNN 是一种连续时间
循环神经网络，其隐藏状态通过神经常

微分方程在观测值之间演化。与离散时间情形一样，循环神经网络的结构可以比简单的循环神经网络更新方程更复杂。在连续时间情形中，门控循环单元（GRU）[34] 和长短期记忆网络（LSTM）单元[35] 均已得到应用。事实上，研究表明，基础循环神经网络在学习长期依赖关系时表现不佳这一经验性问题，在连续时间情形中同样存在[29]。LSTM 还具有额外的隐藏（单元）状态$c(t)$，该状态在更新之间保持不变[29] 。

先前作者[23] 提出的 ODE-RNN 中，NN_{vec} 的结构是一个通用前馈神经网络。另一种方法是将离散时间 GRU 视为差分方程，并在时间步长趋于零时取极限[28] ，从而推导出基于门控循环单元的常微分方程。

IMODE 是一类围绕干预（例如医疗治疗）的显式建模而设计的模型。与上述方法不同，它们会增加一个额外的隐藏状态 z_a ，用于反映正在进行的干预（例如胰岛素输注）（算法2）。我们改编了先前工作 [30] 中的方法，以移除更新独立的隐藏状态；对于稀疏测量数据，这种状态是必要的。

算法2。IMODE 的更新与演化步骤

ii. 更新：获得新的观测向量 $\boldsymbol{h}_t$ 后，设定

$$z_h(t) = NN^h_{jump}(\boldsymbol{h_t}, z_x)$$
$$z_a(t) = NN^a_{jump}(x_{input}, z_a)$$

iii. 演化：

$$z_h(t+\Delta_t) = \text{ODESolve}(NN^h_{vec}, (z_x, z_z), (t, t+\Delta_t))$$
$$z_a(t+\Delta_t) = \text{ODESolve}(NN^a_{vec}, z_a, (t, t+\Delta_t))$$

我们评估了以下常微分方程循环神经网络架构：对长短期记忆网络和 IMODE 模型，使用多层感知机作为 NN_{vec} ；对门控循环单元模型，使用门控循环单元作为 NN_{vec} 。

- 常微分方程门控循环单元
- ODE-LSTM
- IMODE

神经流架构流式 RNN 是一种使用神经网络的连续时
间循环神经网络，

神经流模型，用于在观测值之间演化隐藏状态。我们使用门控循环单元作为神经流模型，评估以下常微分方程循环神经网络架构。

- 流式门控循环单元
- 流式长短期记忆网
络

替代架构一种替代的

TRNN 方法是避免使用神经网络来参数化 F_Φ。一种方法是令 $\frac{dz}{dt} = -Az$，其中 $A \in R$ 是众所周知的指数衰减方程，其解为 $z(t + \Delta_t) = z(t)e^{-A\Delta_t}$。为作说明，我们在采用门控循环单元更新的仿真研究中展示了这种方法。

- 衰减门控循环单元

血糖预测的应用

基准方法此前我们介绍了两种

用于重症监护室血糖概率预测的梯度提升树方法 [15]。这两种方法基于 CatBoost 库 [36]，预测的对象可以是 $\boldsymbol{\theta} = (\mu, \sigma)$，并假设 $y \sim N(\mu, \sigma)$；也可以是分位数 q_{p_j} 对应的 $\boldsymbol{\theta} = (q_{p_1}, \ldots, q_{p_n})$。为了考虑患者病史和时间间隔，模型将 $\widetilde{\boldsymbol{h}_t}$（完整病史）和 Δ_t 的滞后值及汇总值（例如滑动平均值、最小值和最大值）作为额外变量，构成输入向量 $\boldsymbol{h}_t^+ = (\boldsymbol{h}_t, \Sigma(\widetilde{\boldsymbol{h}_t},), \Delta_t)$。为便于与深度学习方法直接比较，我们采用假设正态性的这一方法，因为我们假设血糖对数服从正态分布。作为更简单的基准，我们使用线性模型，仅利用最近一次观测到的信息来建模 μ和 σ。在线性模型中，σ^2 的目标结果是 μ 的预测值与观测值 y[37]之差的平方。此外，作为一种消融研究，我们还考虑了将不规则时间间隔作为额外输入变量的标准自回归循环神经网络，得到以下基准方法：

- CatBoost
- 线性模型
- 时间间隔门控循环单元
- 时间间隔-长短期记忆网络

数据集

仿真研究先前的工作 [38]
propo

我们使用非齐次 Ornstein-Uhlenbeck 随机微分方程（SDE）[20] 作为危重症患者血糖的简单模型。我们利用该模型生成包含不规则测量的血糖值，以及时间、胰岛素输入和葡萄糖输入等协变量的合成数据集。模型形式如下：

$$dG = \gamma(G - G_b) + \beta(t)m(t) + g(t) + \sqrt{2\gamma\sigma^2}dW$$

其中，G 表示以 mg/dL 为单位的血糖，G_b 表示“基线”血糖，$\beta(t)$ 表示胰岛素敏感性（未观测到，且在大多数情景中假设其对同一受试者为常数），$m(t)$ 表示胰岛素输入，$g(t)$ 表示糖摄入量，dW 表示维纳“白噪声”过程。

我们模拟 N 条轨迹，时间跨度为 24小时周期；每个案例中同时模拟$G(t)$、测量过程和治疗决策过程（后者根据图1确定 $m(t)$ 的取值），参照 [39]。测量过程旨在定性地模拟真实世界数据的特征（测量时间不规律；血糖较高且接受治疗的受试者被测量得更频繁），最终在24小时周期内平均观测到 7.5 个数值。

建模任务是预测直到下一次可用模拟测量值之前的血糖分布，同时纳入先前的血糖测量值以及营养和胰岛素输入。模拟研究考察模型性能如何随以下因素变化：1）样本量（N = 1 x 10^3, 5 x 10^3 以及 10 x 10^3 和 20 x 10^3 条轨迹）；2）观测值 G 的测量误差（t）；以及 3）胰岛素敏感性的非平稳性（β，t）。完整的数据生成算法和参数设置详见附录A。

电子病历我们的数据集构建自

公开可用的 MIMIC-IV数据库 [40]，使用与一个衍生的葡萄糖-胰岛素测量值数据集一同发布的代码 [41]。研究人员可通过以下网址申请访问该数据库：http://mimic.physionet.org。所得数据集包含与 44,334 条重症监护病房住院记录和 772,784 次血糖测量相关的去标识化健康数据。建模任务是预测直到数据库中下一次可用测量值之前的血糖分布，同时纳入先前的血糖测量值以及营养和胰岛素输入。队列及所用变量的更多详细信息见附录B。

模型与训练用于预测血糖

我们使用上述模型。对于血糖建模，我们将胰岛素和营养输入视为干预。

训练与超参数调优对于我们开展的模拟研究，我们进行 3 次运行

每种模拟设置均使用不同的随机种子进行每次运行。我们使用了一个包含 1,000 条轨迹的外部留出数据集，用于选择深度学习模型隐藏状态的维度、损失混合参数 λ 和测量误差方差 σ_e^2。对电子病历进行分析所得的结果，是通过在源数据 90% 的部分上进行 3 折交叉验证生成的。同样，我们使用一个留出数据集（交叉验证运行中排除的 10% 数据）来选择深度学习模型隐藏状态的维度、混合参数和测量误差方差 σ_e^2。对于深度学习模型和梯度提升树模型，均使用了内部留出数据集（每次交叉验证运行内部的数据集），以防止过拟合并确定算法何时收敛。

模型评估我们评估我们的概率♀♀♀♀

概率预测采用多种指标进行评估，并报告这些评分在测试数据集上的平均值。对于连续概率预测，理想情况下应使用一种标量汇总指标：当预测分布 P_θ （或其密度 p_θ）与真实分布 P_0相符时，该指标达到最小值；对于任何其他 $G \neq P_0$，其值都严格更高。这类指标称为严格适当评分规则 [42]，包括连续排序概率评分和对数评分（或无知评分）。连续排序概率评分定义为

$$\mathrm{CRPS}(P_\theta, y_i) = \int_{-\infty}^{\infty} \left(P_\theta(y_i) - \mathbf{1}_{\{\xi \geq y_i\}}\right)^2 d\xi$$

它相当于对所有可能的二元截断点 [43]的布里尔评分进行积分。如果 P_θ将其全部概率密度集中在一点 $\hat{y}_i$ ，则该评分等同于平均绝对误差。对数评分 $\mathrm{logS}(p_\theta, y_i) = -\log p_\theta(y_i)$衡量在假定分布 p下， y_i 出现的预期程度。显然，logS的均值也等于对数似然。此外，我们使用预测分布的均值计算均方根误差（RMSE），并评估预测的覆盖率和区间评分。对于下界为 l 、上界为 u 的 $100(1-\alpha)\%$ 预测区间，区间评分定义为

$$S(l, u, y_i) = (u - l) + \frac{2}{\alpha}(l - y_i)\mathbf{1}_{\{y_i < l\}} + \frac{2}{\alpha}(y_i - u)\mathbf{1}_{\{y_i > u\}}$$

分数越低，性能越好。对于一个分布 P_θ ，可以证明累积概率 $P_\theta(y < Y)$ 的分布应为均匀分布 [44]。在过度离散的情况下（不确定性过大 – 例如，对于均值固定的正态分布，预测方差过大），这些概率的方差小于$\frac{1}{12} = 0.83..$；而在离散不足的情况下，该方差则大于此值。按照以往研究 [44] 的做法，我们将这一指标称为概率积分变换的方差（Var PIT）。我们计算交叉验证各轮中这些指标的均值和标准差。

软件与可复现性该分析使用 Python 编写，
w

并使用 PyTorch [45] 编写模型，使用 Matplotlib[46] 绘制图表，使用 pandas 进行数据处理和汇总 [47] ，并使用 properscoring 计算若干模型指标。本文的代码可在 www.github.com/oizin/irregular-ts。

结果

模拟模拟结果

模拟研究结果如图4所示。随着数据集规模从1,000条轨迹增加到5,000条，所有模型的性能均有所提升。然而，当轨迹数量进一步从5,000条增加到10,000条时，各模型的提升幅度有所不同。在较小的数据集上，CatBoost模型明显表现更优；而在较大的数据集上，一些深度学习模型的表现达到或超过了它（图4A）。加入测量误差后，模型性能有所下降，但模型排名基本保持了标准设置下的情况。然而，令人意外的是，使胰岛素的影响呈非平稳性后，各模型的性能略有提升（图5A）。对标准设置和非平稳设置下的结果进行分析后发现，这是因为在非平稳设置下，胰岛素给药病例的表现有所改善。这是非平稳性带来的副作用：由于治疗平均持续时间延长了一倍，含有胰岛素治疗的训练样本数量也随之增加。

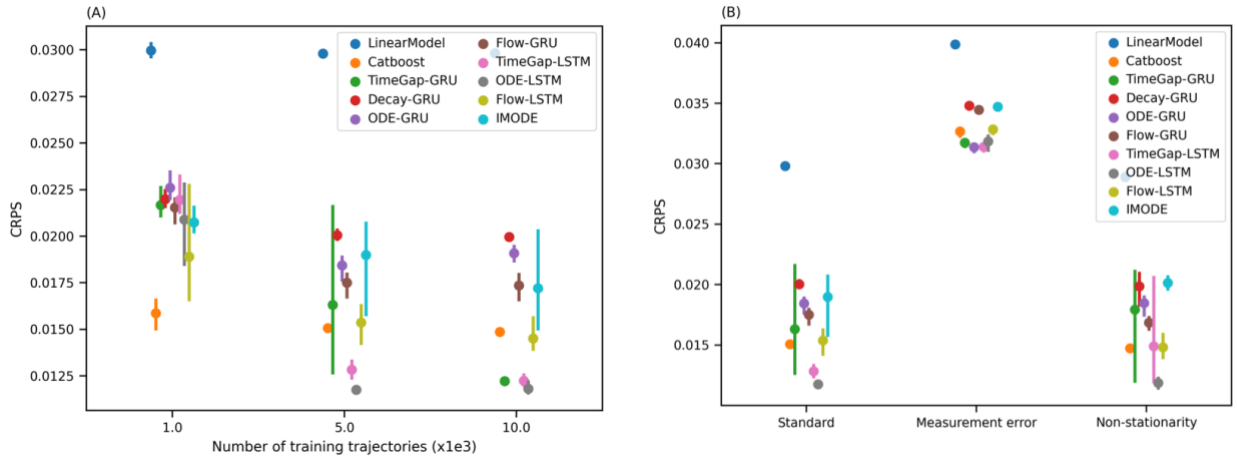

图 4. A）数据规模模拟中连续排序概率评分（CRPS）的均值、最小值和最大值；B）测量误差和非平稳性模拟中 CRPS 的均值、最小值和最大值（N=10,000 条轨迹）。请注意，实际数据点的数量是轨迹数量的 ~10 倍。

电子病历电子病历任务的结果可以

结果见表5（图示见附录 C）。在连续时间深度学习模型中，ODE-LSTM模型表现最佳。在点预测方面（均方根误差），所有连续时间循环神经网络模型的表现都优于加入时间间隔变量的离散时间对应模型，但在准确估计预测不确定性方面（例如考察区间评分）则未必如此。值得注意的是，除ODE-LSTM外，CatBoost模型的表现优于所有深度学习模型；可以说，这两个模型的表现相当。作为基线比较，使用上一次观测值进行点预测，在整个数据集上的均方根误差为45.4 mg/dL（与线性模型的预测结果一致）。

表 1. 血糖预测问题中各模型的比较，指标为各实验中的均值和标准差

模型	均方根误差（mg/dL）	连续排序概率评分（×100）	logS（×100）	Var PIT (x100)	95% 预测区间	
					实证覆盖面 *(%)**	区间评分
线性模型 44.1±0.8		13.3±0.1	-4.0±2.8	6.6±0.1	95.7±0.1	214±1
CatBoost	37.**1**±**0**.2	11.**8**±**0**.1	14.9±0.2	7.**7**±**0**.1	94.**3**±**0**.2	**176±1**
时间间隔-*GRU*	39.4±0.7	12.4±0.1	10.9±0.9	7.6±0.1	94.3±0.4	193±2
常微分方程门控循环单元	39.4±0.3	13.0±0.2	-23.5±6.6†	7.3±0.1	93.7±0.1	211±5†
流式门控循环单元	38.5±0.73	12.3±0.1	12.3±2.2	7.5±0.1	94.6±0.1	187±1
时间间隔-*LSTM*	39.4±0.9	12.4±0.1	10.9±0.1	7.6±0.1	94.7±0.2	181±1
ODE-LSTM 36.**9**±**0**.2		11.**8**±**0**.1	15.**2**±**0**.1	7.**6**±**0**.1	94.**5**±**0**.2	**175±1**
流式长短期记忆网络 37.6±0.2		12.1±0.1	5.8±2.6	7.4±0.1	94.6±0.3	184±1
IMODE	38.5±0.1	12.4±0.1	6.0±0.5	8.7±0.1	91.2±0.2	198±1

*应使用区间评分比较区间性能

†对于 60 个病例，该模型预测了标准差的极端离群值。

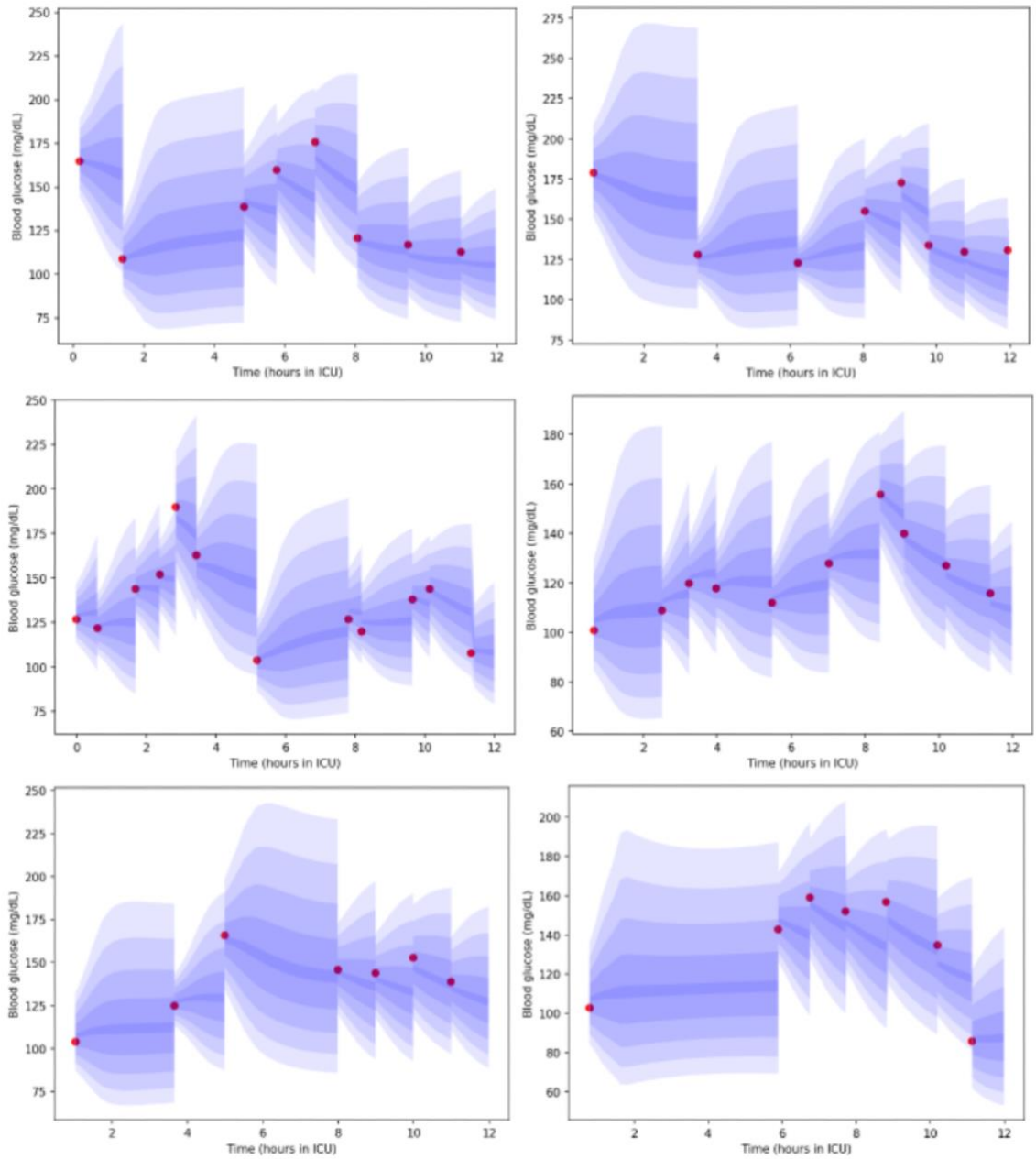

图 5. ODE-LSTM 模型的预测示例及其 95% 预测区间。红点表示观测值。所选患者是随机抽取的，且在 12 小时内至少有三次观测。

讨论

本文旨在介绍并评估连续时间自回归循环神经网络（CTRNN）模型，用于不规则测量时间序列情境下的概率预测。这些方法具有扎实的理论基础。它们以离散时间自回归循环神经网络为基础，通过加入连续时间层，使隐状态能够在观测之间演化。这种方法能够处理测量过程的变化以及药物输注等持续干预，从而实现实时预测。此外，这些模型具有很强的灵活性，并允许

广泛选择模型架构，具体取决于所要解决的问题，例如面向持续干预的 IMODE 模型。将这些方法应用于模拟数据和电子病历（EMR）中的重症监护室（ICU）血糖预测，结果表明，这些方法能够从稀疏且不规则测量的数据中学习。加入基于神经网络的连续时间层（神经常微分方程或神经流）后，在大多数指标上，其表现优于标准的门控循环单元（GRU）或长短期记忆网络（LSTM）架构。然而，这些模型仍容易受到测量误差等问题的影响，需要足够的数据，而且其表现往往不及梯度提升树（GBT）方法。

在电子病历场景中，连续时间模型的结果参差不齐，其中 ODE-LSTM 模型在大多数指标上的表现优于其他深度学习模型。加入连续的隐藏状态动态，通常能够提升相应离散时间模型的性能；比较 GRU 和 LSTM 模型的连续时间版本与时间间隔版本即可看出这一点（表5）。这并不出乎意料，因为胰岛素和葡萄糖输注等输入特征会产生随时间变化的影响（治疗效应）[48]。然而，仅仅加入连续动态并不能带来出色的表现。CatBoost 的表现优于除 ODE-LSTM 以外的所有深度学习模型，而且训练起来更简单、速度更快，这与先前的研究结果一致：梯度提升树方法在表格数据[49]，（包括层次化时间数据[50]）上可以取得更好的性能。此外，TimeGap-LSTM 的表现优于或不逊于若干基于连续时间 GRU 的模型，这凸显出，更新步骤和演化步骤中编码的假设对于构建性能良好的模型至关重要。据我们所知，目前还没有研究采用概率预测方法，在不规则时间间隔下进行血糖预测；这表明，为此任务建立基准可能值得一试，正如机械通气领域所做的那样[51]，。

仿真研究结果表明，数据规模对深度学习模型的训练至关重要。与 CatBoost 不同，随着数据规模从 1,000 条轨迹增加到 5,000 条，CatBoost 的性能仅略有提升，而深度学习模型的性能则显著提高。仿真还揭示了这些模型的局限性，并为电子病历与仿真之间的误差指标相差一个数量级提供了一些可能的解释（除了一般而言复杂度有所增加之外）。同时引入测量误差和未观测到的非平稳性会改变模型性能。测量误差是重症监护室血糖测量中的一个已知问题[52, 53]。测量误差的方差不太可能保持恒定，而是会随血糖值、测量设备或来源，以及血压等其他患者特征而变化。另有研究认为，危重症患者的胰岛素敏感性也会随时间发生显著变化 [54]。在仿真中，非平稳的胰岛素敏感性虽然提升了模型性能，但这是因为接受胰岛素治疗的观测值比例增加，从而使模型在这部分数据上的表现更加准确。这一点对于将这些模型用于治疗建议而言，是一项重要考量。

类似的效应可能解释了电子病历场景结果中的过度离散。在电子病历场景中，只有少数模型达到了 0.83 的最佳概率积分变换方差，过度离散才是常态。这可能是由于存在未观测到的因素–相关信息要么未记录在电子病历中，要么未纳入本次分析–，导致所有方法的模型不确定性都很高。

这项研究的优势在于评估了时间序列预测中的一项艰巨任务。可以说，我们的数据既不规则，时间上也较为稀疏。在一定程度上，这可能会限制研究结论在测量频率更高、但仍不规则的数据场景中的适用性。例如，据报道，原始 IMODE 模型虽然不适用于当前任务，但表现出色

在血压问题中表现良好[30]，而血压是在重症监护室中测量频率高得多的变量。模型和研究的局限性还凸显了未来研究的若干其他方向。一个关键问题是模型的可操作性——即模型学习到有意义的生理关系的程度，这些关系可用于指导治疗。反事实预测以及一定程度的模型可解释性，可能是此类应用的关键[6, 55]。当前的方法为进一步完善奠定了基础。当前研究中的一个直接例子，是开发胰岛素—葡萄糖剂量反应曲线模型。更广泛而言，持续发展能够考虑观测医学数据特殊性的机器学习和深度学习方法，或将推动医疗保健领域取得进展。

结论总而言之，

深度学习方法应用于不规则测量的时间序列预测，前景可期。然而，要通过梯度提升树方法（以及特征转换）等进行恰当的基准比较，才能突出新方法在表格数据场景中是否真正达到最先进水平。医疗数据在统计和因果层面都颇具挑战；如何处理测量误差、准确量化不确定性并生成反事实预测，是未来研究的关键目标。

致谢

本研究获得了 eHealth NSW、悉尼新南威尔士大学和澳大利亚联邦科学与工业研究组织（CSIRO）的资助，资助项目为产业博士计划。

参考文献

1. Hulsen, T. 等，《从大数据到精准医学》。Frontiers in Medicine，2019年，6卷。 2. Agniel, D.、I.S. Kohane 和 G.M. Weber，《医疗系统内部流程导致的电子健康记录数据偏倚：回顾性观察研究》。BMJ，2018年，361卷。 3. Seymour, C.W. 等，《精准医学惠及所有人？将精准医学方法应用于重症的挑战与机遇》。Critical Care，2017年，21卷（1期）：第257页。 4. Rajkomar, A. 等，《利用电子健康记录实现可扩展且准确的深度学习》。NPJ Digital Medicine，2018年，1卷（1期）：第1–10页。 5. Esteva, A. 等，《医疗领域深度学习指南》。Nature Medicine，2019年，25卷（1期）：第24–29页。 6. Lim, B. 和 S. Zohren，《深度学习时间序列预测：综述》。Philosophical Transactions of the Royal Society A，2021年，379卷（2194期）：第20200209页。 7. Che, Z. 等，《针对含缺失值多变量时间序列的循环神经网络》。Scientific Reports，2018年，8卷（1期）：第1–12页。 8. Lara-Benítez, P.、M. Carranza-García 和 J.C. Riquelme，《时间序列预测深度学习架构的实验性综述》。arXiv预印本 arXiv:2103.12057，2021年。 9. Elman, J.L.，《发现时间中的结构》。Cognitive Science，1990年，14卷（2期）：第179–211页。 10. Chen, R.T. 等，《神经常微分方程》。arXiv预印本 arXiv:1806.07366，2018年。 11. Rhodes, A. 等，《拯救脓毒症运动：脓毒症和脓毒性休克管理国际指南：2016》。重症医学，2017年，43卷（3期）：第304–377页。 12. Vincent, J.-L.，《每天至少一次，给患者做个快速的 FAST HUG 检查》。Critical Care Medicine，2005年，33卷（6期）：第1225–1229页。 13. Baker, L. 等，《重症监护室血糖控制与胰岛素使用的真实世界特征分析》。Scientific Reports，2020年，10卷（1期）：第1–10页。 14. Fitzgerald, O. 等，《将真实世界证据纳入重症监护室患者血糖预测算法的开发》。Journal of the American Medical Informatics Association，2021年。 15. Pappada, S.M. 等，《基于人工神经网络的预测模型：支持优化住院患者血糖控制》。Diabetes Technol Ther，2020年，22卷（5期）：第383–394页。 16. Pappada, S.M. 等，《对危重外科患者血糖预测模型的评估》。PLOS ONE，2013年，8卷（7期）：第e69475页。 17. Pappada, S.M. 等，《开发用于外科重症监护环境血糖水平预测的神经网络模型》。Patient Safety in Surgery，2010年，4卷（1期）：第15页。 18. Van Herpe, T. 等，《用于预测危重患者血糖水平的自适应输入-输出建模方法》。Physiol Meas，2006年，27卷（11期）：第1057–1069页。 19. Särkkä, S. 和 A. Solin，《应用随机微分方程》。第10卷。2019年：剑桥大学出版社。 20. Salinas, D. 等，《DeepAR：利用自回归循环网络进行概率预测》。International Journal of Forecasting，2020年，36卷（3期）：第1181–1191页。 21. Funahashi, K.-i. 和 Y. Nakamura，《利用连续时间循环神经网络逼近动力系统》。Neural Networks，1993年，6卷（6期）：第801–806页。 22. Rubanova, Y.、R.T. Chen 和 D. Duvenaud，《针对不规则采样时间序列的潜在常微分方程》。arXiv预印本 arXiv:1907.03907，2019年。 23. Rumelhart, D.E.、G.E. Hinton 和 R.J. Williams，《通过反向传播误差学习表示》。Nature，1986年，323卷（6088期）：第533–536页。 24. Herrera, C.、F. Krach 和 J. Teichmann，《神经跳跃常微分方程：一致的连续时间预测与滤波》。arXiv预印本 arXiv:2006.04727，2020年。

25. De Brouwer, E. 等，《GRU-ODE-Bayes：对间歇观测时间序列进行连续建模》。arXiv 预印本 arXiv:1905.12374，2019年。 26. Lechner, M. 和 R. Hasani，《学习不规则采样时间序列中的长期依赖》。arXiv预印本 arXiv:2006.04418，2020年。 27. Gwak, D. 等，《用于干预建模的神经常微分方程》。arXiv预印本 arXiv:2010.08304，2020年。 28. Hairer, E.、S.P. Nørsett 和 G. Wanner，《常微分方程求解》。第1卷：非刚性问题。1993年：Springer-Vlg。 29. Hairer, E. 和 G. Wanner，《常微分方程求解 II》。第375卷。1996年：Springer Berlin Heidelberg。 30. Ott, K. 等，《毕竟还是残差网络：神经常微分方程及其数值求解》。载于《国际学习表征会议》。2020年。 31. Dormand, J.R. 和 P.J. Prince，《一类嵌入式龙格-库塔公式》。计算与应用数学杂志，1980年，6卷（1期）：第19–26页。 32. He, K. 等，《用于图像识别的深度残差学习》。载于《IEEE计算机视觉与模式识别会议论文集》。2016年。 33. Cho, K. 等，《利用循环神经网络编码器-解码器学习统计机器翻译的短语表示》。arXiv预印本 arXiv:1406.1078，2014年。 34. Hochreiter, S. 和 J. Schmidhuber，《长短期记忆网络》。神经计算，1997年，9卷（8期）：第1735–1780页。 35. Albers, D. 等，《人体葡萄糖 - 胰岛素系统预测的简易建模框架》。arXiv 预印本 arXiv:1910.14193，2019年。 36. Wilson, M.、J. Weinreb 和 G.W.S. Hoo，《重症监护中的强化胰岛素治疗：12种方案综述》。糖尿病护理，2007年，30卷（4期）：第1005–1011页。 37. Johnson, A.E. 等，《MIMIC-III：免费开放的重症监护数据库》。科学数据，2016年，3卷（1期）：第1–9页。 38. Arévalo, A.R. 等，《用于描述重症监护室血糖管理的数据驱动整理流程》。科学数据，2021年，8卷（1期）：第1–13页。 39. Gneiting, T. 和 A.E. Raftery，《严格适当评分规则、预测与估计》。美国统计学会杂志，2007年，102卷（477期）：第359–378页。 40. Zamo, M. 和 P. Naveau，《在信息有限情况下估计连续排序概率评分及其在集合天气预报中的应用》。数学地球科学，2018年，50卷（2期）：第209–234页。 41. Paszke, A. 等，《 PyTorch：一种命令式风格、高性能的深度学习库》。神经信息处理系统进展，2019年，32卷：第8026–8037页。 42. Hunter, J.D.，《Matplotlib：二维图形环境》。科学与工程计算，2007年，9卷（03期）：第90–95页。 43. McKinney, W.，《用于统计计算的 Python 数据结构》。载于《第9届科学领域Python会议论文集》。2010年。德克萨斯州奥斯汀。 44. Morrill, J. 等，《用于在线预测任务的神经控制微分方程》。arXiv预印本 arXiv:2106.11028，2021年。 45. Chase, J.G. 等，《新一代个性化模型驱动重症医学：计算机模拟虚拟患者模型、方法和队列的前沿综述，以及如何对其进行验证》。生物医学工程在线，2018年，17卷（1期）：第1–29页。 46. Egi, M.、S. Finfer 和 R. Bellomo，《重症监护室的血糖控制》。胸部，2011年，140卷（1期）：第212–220页。 47. Kulkarni, A. 等，《分析重症监护患者毛细血管和动脉血样本的血糖测量值》。重症医学，2005年，31卷（1期）：第142–145页。 48. Ferenci, T. 等，《危重症患者胰岛素敏感性变异的每日变化及其与诊断的关系》。PLOS ONE，2013年，8卷（2期）：第e57119页。 49. Ustimenko, A.、L. Prokhorenkova 和 A. Malinin，《通过集成方法处理梯度提升中的不确定性》。arXiv预印本 arXiv:2006.10562，2020年。

50. Gal, Y. 和 Z. Ghahramani。 将 Dropout 作为贝叶斯近似：在深度学习中表示模型不确定性。载于 国际机器学习会议。2016 年。PMLR。51. Krishnan, R.G.、U. Shalit 和 D. Sontag，深度卡尔曼滤波器。 arXiv 预印本 arXiv:1511.05121，2015。52. Prosperi, M. 等，面向可操作医疗保健的机器学习中的因果推断与反事实预测。《自然·机器智能》，2020 年。**2**(7)：第 369–375 页。53. Bellot, A. 和 M. van der Schaar，使用合成对照进行连续时间策略分析。 arXiv 预印本 arXiv:2102.01577，2021 年。

附录A：仿真研究

数据生成算法数据生成算法的特点

构建了三个相互关联的过程：一个因果随机微分方程（SDE）、一个测量过程，以及一个治疗决策算法。数据由以下 SDE 生成

$$dG(t) = (\gamma(G(t) - G_b) + \beta m(t, G(s)) + g(t))dt + \sqrt{2\gamma\sigma^2}dW$$

其中 $G(t)$ 表示以 mg/dL 为单位的血糖，G_b 表示"基线"血糖，$m(t, G(s))$ 表示根据时间 $s \le t$ 的血糖测量值 $G(s)$ 设定的胰岛素输入，$g(t)$ 表示糖分摄入，dW 表示维纳"白噪声"过程。我们使用欧拉–丸山方法从该模型中抽取 N 条轨迹，每条轨迹覆盖一个24小时周期。对于这些轨迹，我们还同时模拟测量过程和治疗过程。只有在测量时间，才能更改 $m(t, G(s))$ 的设置，以模拟临床就诊。从时间零开始，我们使用以下模型抽取下一个测量时间 t_{obs}

$$t_{obs} \sim LogNormal(m_{obs}, 0.2)$$

其中 m_{obs} 的值基于 G(t) 和 $m(t, G(s))$

$$\begin{cases} m_{obs} = b(m) \cdot \exp\left(\left(\frac{\mathrm{G(t)} - 120}{50}\right)^2\right) & \text{if G(t)} < 80 \\ m_{obs} = b(m) \cdot \exp\left(\left(\frac{\mathrm{G(t)} - 120}{140}\right)^2\right) & \text{if G(t)} \ge 80 \end{cases}$$

$$\begin{cases} b(m) = 3 & \text{if } m > 0 \\ b(m) = 5 & \text{if } m = 0 \end{cases}$$

该模型旨在定性模拟真实世界数据中观察到的模式：测量值越极端且正在接受治疗的受试者，测量频率越高。治疗算法依据一篇胰岛素方案综述中的图 X[39]。采用以下治疗算法，单位为胰岛素单位/小时

$$\begin{cases} m\big(t, G(t)\big) = 0.0 \; if \; G(t) < 140 \\ m\big(t, G(t)\big) = 3.0 \; if \; 140 \le G(t) < 160 \\ m\big(t, G(t)\big) = 10.0 \; if \; 160 \le G(t) < 200 \\ m\big(t, G(t)\big) = 20.0 \; if \; G(t) \ge 200 \end{cases}$$

We u对每条模拟轨迹使用以下参数设置和初始条件

$$G(0) \sim N(140,20)$$
$$G_b \sim N(140,5)$$
$$\gamma \sim N(0.5,0.01)$$
$$\sigma \sim N(20,2)$$
$$\beta \sim N(50,5)$$

对于存在测量误差的模拟，我们在观测值中加入 $e \sim N(0, X)$，建模时使用这个测量不准确的值 $\tilde{G}(t)$，而评估算法性能时使用真实值 $G(t)$。

下述数据生成算法的概要如下所示（表 1）

<escape>表 1. 数据生成算法对于</escape>

	i in $1:N$
启动	样本 $G(0)$
采样轨迹	虽然 $t < 24$
	如果 $t = t_{obs}$: 设置 $m = m(t, G(t))$ 设置下一个测量时间 t_{obs} $G(t) = G(t - dt) + dG$ $\tilde{G}(t) = G(t - dt) + e$
返回	观测过程 完整流程

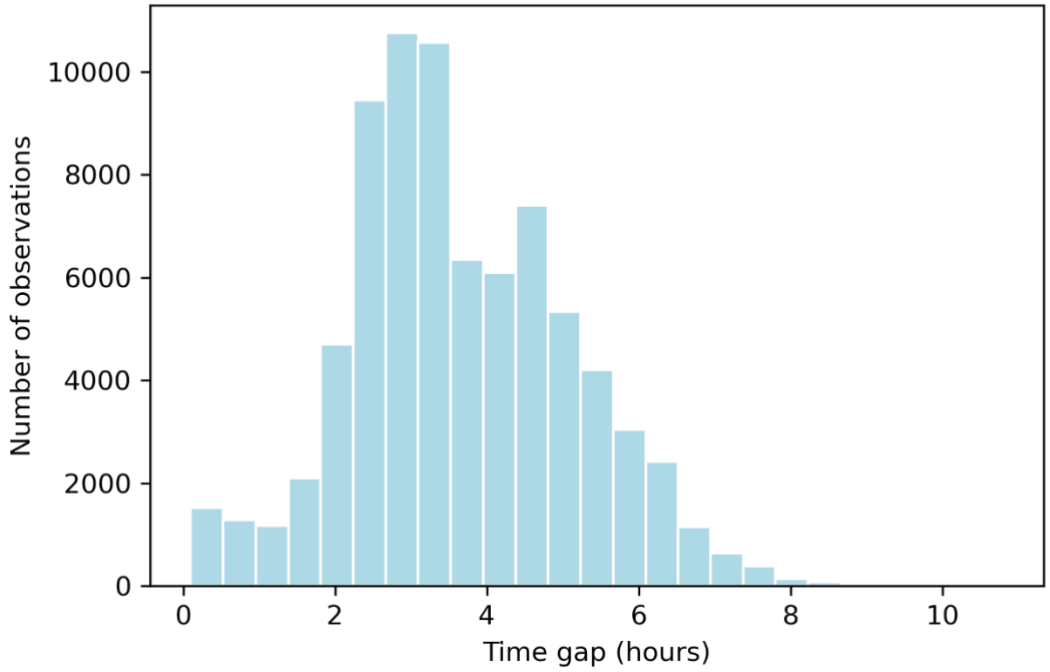

图 A1. 一次模拟运行中测量间隔时间的分布

参考文献

1. Hulsen, T. 等，从大数据到精准医学。《医学前沿》，2019 年。**6**。2. Agniel, D.、I.S. Kohane 和 G.M. Weber，由于医疗保健系统内部流程导致的电子健康记录数据偏差：回顾性观测研究。《英国医学杂志》，2018 年。**361**。3. Seymour, C.W. 等，精准医学适用于所有人吗？精准医学方法应对重症挑战与机遇。《重症监护》，2017 年。**21**（1）：第 257 页。4. Rajkomar, A. 等，利用电子健康记录实现可扩展且准确的深度学习。《NPJ 数字医学》，2018 年。**1**（1）：第 1-10 页。5. Esteva, A. 等，医疗保健领域深度学习指南。《自然医学》，2019 年。**25**（1）：第 24-29 页。6. Lim, B. 和 S. Zohren，基于深度学习的时间序列预测：综述。《英国皇家学会哲学汇刊 A 辑》，2021 年。**379**（2194）：第 20200209 页。

7. Che, Z., 等，《针对含缺失值多变量时间序列的循环神经网络》。Scientific Reports，2018。8(1)：第1–12页。8. Lara-Benítez, P., M. Carranza-García 和 J.C. Riquelme，《时间序列预测深度学习架构的实验性综述》。arXiv 预印本 arXiv:2103.12057，2021。9. Elman, J.L.，《发现时间中的结构》。Cognitive Science，1990。14(2)：第179–211页。10. Chen, R.T., 等，《神经常微分方程》。arXiv 预印本 arXiv:1806.07366，2018。11. Biloš, M., 等，《神经流：神经常微分方程的高效替代方案》。神经信息处理系统进展，2021。34：第21325–21337页。12. Rhodes, A., 等，《拯救脓毒症运动：脓毒症和脓毒性休克管理国际指南：2016》。重症医学，2017。43(3)：第304–377页。13. Vincent, J.-L.，《每天至少一次，给患者一个快速拥抱》。Critical Care Medicine，2005。33(6)：第1225–1229页。14. Baker, L., 等，《重症监护室血糖控制和胰岛素使用的真实世界特征》。Scientific Reports，2020。10(1)：第1–10页。15. Fitzgerald, O., 等，《将真实世界证据纳入重症监护室患者血糖预测算法的开发》。Journal of the American Medical Informatics Association，2021。16. Pappada, S.M., 等，《基于人工神经网络的预测模型：支持优化住院患者的血糖控制》。Diabetes Technol Ther，2020。22(5)：第383–394页。17. Pappada, S.M., 等，《危重症外科患者血糖预测模型的评估》。PLOS ONE，2013。8(7)：第e69475页。18. Pappada, S.M., 等，《用于预测外科重症监护环境中血糖水平的神经网络模型开发》。Patient Safety in Surgery，2010。4(1)：第15页。19. Van Herpe, T., 等，《用于预测危重症患者血糖水平的自适应输入－输出建模方法》。Physiol Meas，2006。27(11)：第1057–69页。20. Särkkä, S. 和 A. Solin，《应用随机微分方程》。第10卷。2019：剑桥大学出版社。21. Salinas, D., 等，《DeepAR：基于自回归循环网络的概率预测》。International Journal of Forecasting，2020。36(3)：第1181–1191页。22. Funahashi, K.-i. 和 Y. Nakamura，《用连续时间循环神经网络近似动力系统》。Neural Networks，1993。6(6)：第801–806页。23. Rubanova, Y., R.T. Chen 和 D. Duvenaud，《针对不规则采样时间序列的潜在常微分方程》。arXiv 预印本 arXiv:1907.03907，2019。24. Rumelhart, D.E., G.E. Hinton 和 R.J. Williams，《通过反向传播误差学习表示》。Nature，1986。323(6088)：第533–536页。25. Che, Z., 等，《针对含缺失值多变量时间序列的循环神经网络》。Scientific Reports，2018。8(1)：第6085页。26. Durkan, C., 等，《神经样条流》。神经信息处理系统进展，2019。32。27. Herrera, C., F. Krach 和 J. Teichmann，《神经跳跃常微分方程：一致的连续时间预测与滤波》。arXiv 预印本 arXiv:2006.04727，2020。28. De Brouwer, E., 等，《GRU-ODE-Bayes：对稀疏观测时间序列进行连续建模》。arXiv 预印本 arXiv:1905.12374，2019。29. Lechner, M. 和 R. Hasani，《学习不规则采样时间序列中的长期依赖》。arXiv 预印本 arXiv:2006.04418，2020。30. Gwak, D., 等，《用于干预建模的神经常微分方程》。arXiv 预印本 arXiv:2010.08304，2020。31. Hairer, E., S.P. Nørsett 和 G. Wanner，《常微分方程求解》。第1卷，非刚性问题。1993：Springer-Vlg。

32. Hairer, E. 和 G. Wanner，《常微分方程求解 II》。第375卷。1996：Springer Berlin Heidelberg。33. Ott, K., 等，《毕竟是残差网络：神经常微分方程及其数值求解》。载于国际学习表征会议。2020。34. Cho, K., 等，《使用循环神经网络编码器-解码器学习用于统计机器翻译的短语表示》。arXiv 预印本 arXiv:1406.1078，2014。35. Hochreiter, S. 和 J. Schmidhuber，《长短期记忆网络》。神经计算，1997。9(8)：第1735–1780页。36. Prokhorenkova, L., 等，《CatBoost：具有类别特征的无偏提升方法》。神经信息处理系统进展，2018。31。37. Fan, J. 和 Q. Yao，《随机回归中条件方差函数的高效估计》。Biometrika，1998。85(3)：第645–660页。38. Albers, D., 等，《人体葡萄糖-胰岛素系统预测的简易建模框架》。arXiv 预印本 arXiv:1910.14193，2019。39. Wilson, M., J. Weinreb 和 G.W.S. Hoo，《重症监护中的强化胰岛素治疗：12种方案综述》。糖尿病护理，2007。30(4)：第1005–1011页。40. Johnson, A., 等，《MIMIC-IV》。PhysioNet。在线网址：https://physionet.org/content/mimiciv/1.0/（访问日期：2021年8月23日），2020。41. Arévalo, A.R., 等，《用于描述重症监护室血糖管理的数据驱动整理流程》。Scientific Data，2021。8(1)：第1–13页。42. Gneiting, T. 和 A.E. Raftery，《严格适当评分规则、预测与估计》。美国统计学会期刊，2007。102(477)：第359–378页。43. Zamo, M. 和 P. Naveau，《在信息有限的情况下估计连续排序概率评分及其在集合天气预报中的应用》。数学地球科学，2018。50(2)：第209–234页。44. Gneiting, T. 和 M. Katzfuss，《概率预测》。Annual Review of Statistics and Its Application，2014。1：第125–151页。45. Paszke, A., 等，《PyTorch：一种命令式风格的高性能深度学习库》。神经信息处理系统进展，2019。32：第8026–8037页。46. Hunter, J.D.，《Matplotlib：一个二维图形环境》。科学与工程计算，2007。9(03)：第90–95页。47. McKinney, W.，《用于统计计算的 Python 数据结构》。载于第9届科学领域Python会议论文集。2010。德克萨斯州奥斯汀。48. Chase, J.G., 等，《新一代个性化模型驱动的重症医学：计算机模拟虚拟患者模型、方法和队列的最新综述，以及如何验证这些模型》。生物医学工程在线，2018。17(1)：第1–29页。49. Shwartz-Ziv, R. 和 A. Armon，《表格数据：深度学习并非你所需的一切》。Information Fusion，2022。81：第84–90页。50. Makridakis, S., E. Spiliotis 和 V. Assimakopoulos，《M5准确度竞赛：结果、发现与结论》。International Journal of Forecasting，2022。38(4)：第1346–1364页。51. Morrill, J., 等，《用于在线预测任务的神经控制微分方程》。arXiv 预印本 arXiv:2106.11028，2021。52. Egi, M., S. Finfer 和 R. Bellomo，《重症监护室中的血糖控制》。胸部，2011。140(1)：第212–220页。53. Kulkarni, A., 等，《分析重症监护患者毛细血管和动脉血样本的血糖测量值》。重症医学，2005。31(1)：第142–145页。54. Ferenci, T., 等，《危重症患者胰岛素敏感性变异的每日变化及其与诊断的关系》。PLOS ONE，2013。8(2)：第e57119页。55. Prosperi, M., 等，《面向可行动医疗保健的机器学习因果推断与反事实预测》。Nature Machine Intelligence，2020。2(7)：第369–375页。

附录B：电子病历

队列

该队列包含 44,334 个住院时段中的 772,784 个血糖测量值。

变量表 A1. Va

电子病历数据集中使用的变量表 A1. 结局

血糖
协变量/干预措施
糖尿病状态（并发/无并发症） Age 重症监护病房入院时的体重/身高 胰岛素 短效（注射） 短效（推注） 中效（注射剂） 长效（注射剂） 全肠外营养/输注 药物输注中的葡萄糖（g） 葡萄糖用量（全胃肠外营养；g） 肠内营养 肠内营养液体积（mL） 碳水化合物量（克；肠内） 葡萄糖用量（g；肠内） 脂肪量（g；肠内营养） 蛋白质摄入量（克；肠内） 纤维量（g；肠内） 热量（肠内营养）
持续干预
胰岛素 胰岛素输注
其他
距下一次预测的时间

附录 C：补充结果模拟

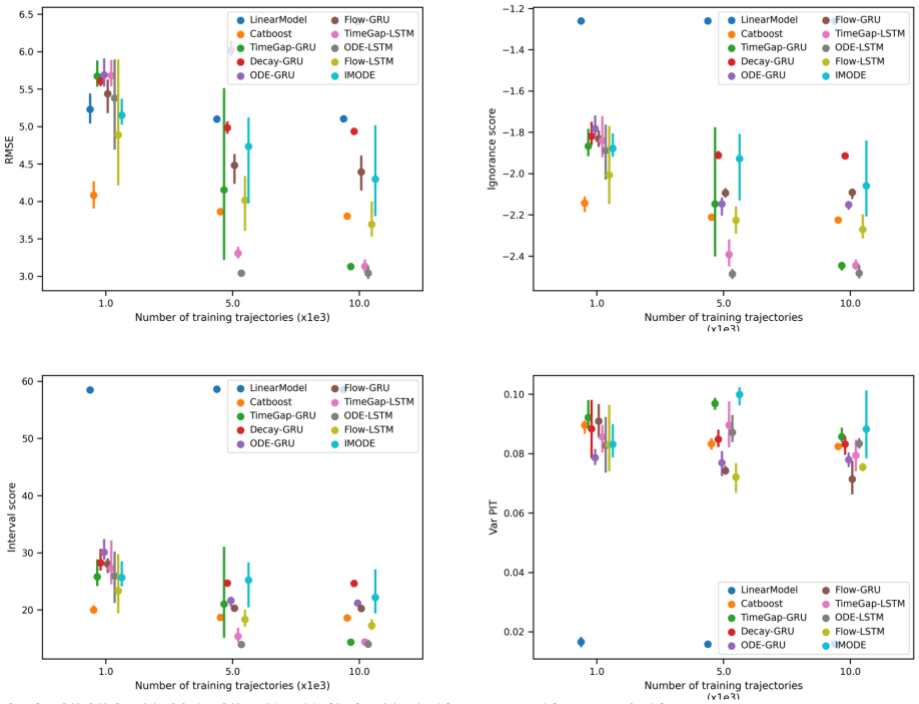

图 C1。数据规模模拟的其他模型评估指标的均值、最小值和最大值。

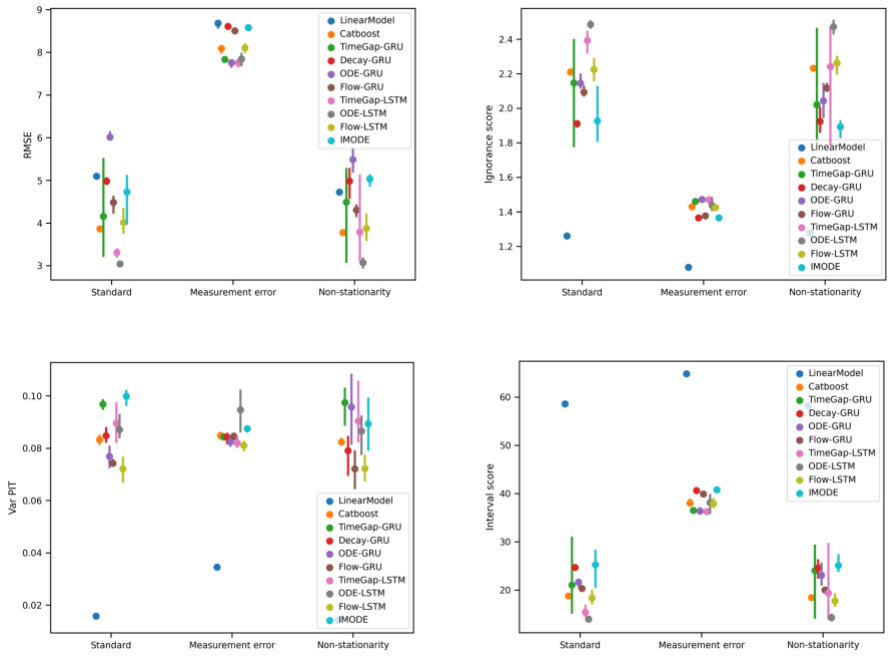

图 C2。测量误差和非平稳模拟的其他模型评估指标的均值、最小值和最大值。

电子病历下方（图 C3）以图形方式展示

了正文表 1 中的信息。

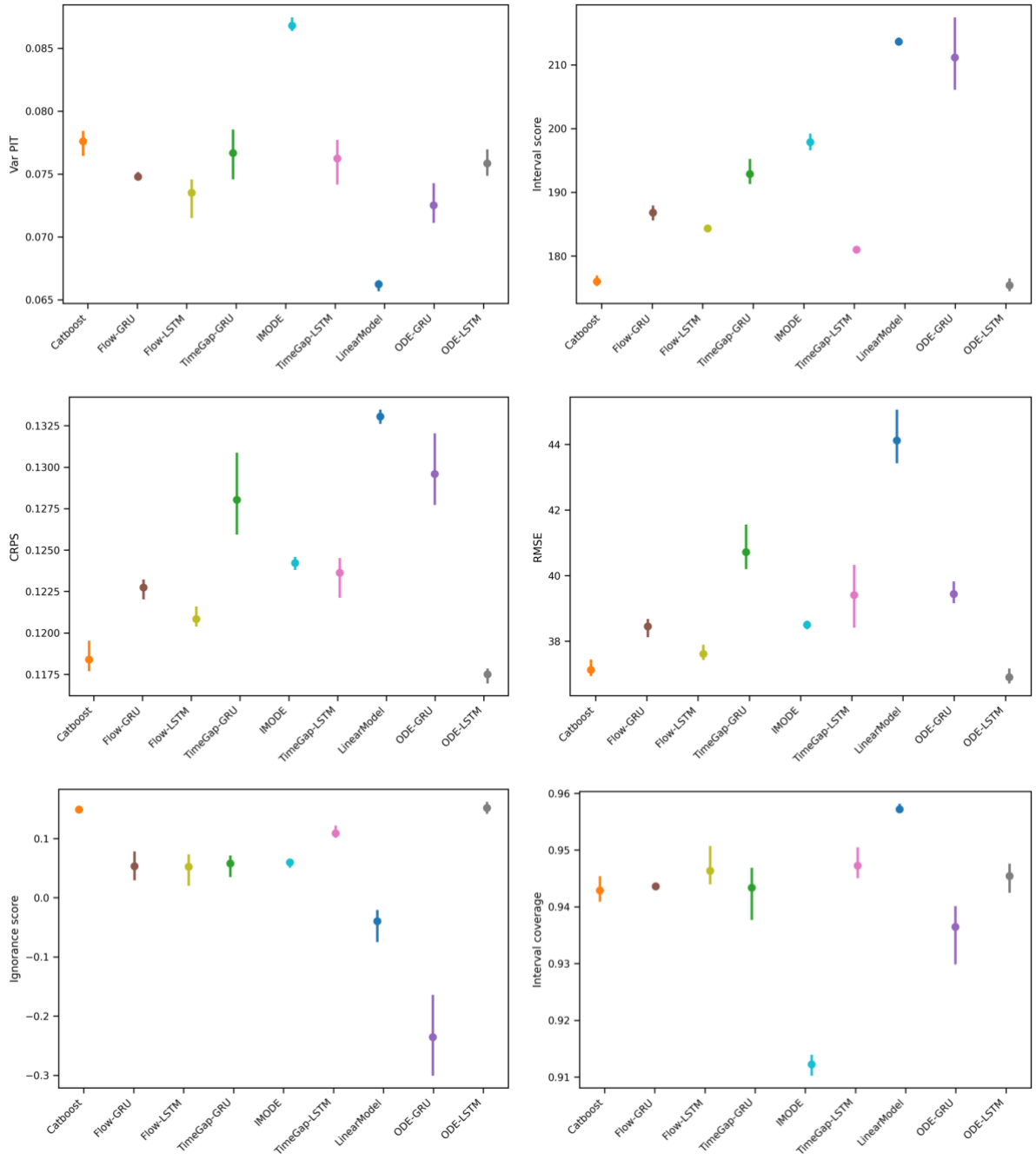

图　C3。电子病历数据集模型评估指标的均值、最小值和最大值
病历　数据集。