****

硕士学位论文

|  |
| --- |
| 基于ICU动态数据的不良事件实时预警模型研究 |
|  |

|  |  |
| --- | --- |
| 作者姓名 | 肖龙慧 |
| 学科专业 | 概率论与数理统计 |
| 指导教师 | 夏立 教授 |
| 所在学院 | 数学学院 |
| 论文提交日期 | 年 月 日 |

**Research on Real-Time Early Warning Model for Adverse Events Based on Dynamic ICU Data**

A Dissertation Submitted for the Degree of Master

**Candidate：Xiao Longhui**

**Supervisor：Prof. Xia Li**

South China University of Technology

Guangzhou, China

**分类号：TP391 学校代号：10561**

**学 号：202220129898**

华南理工大学硕士学位论文

基于ICU实时数据的急性危重疾病早期预警模型研究

作者姓名：肖龙慧 指导教师姓名、职称：夏立、教授

申请学位级别：理学硕士 学科专业名称：概率论与数理统计

研究方向：统计学习及其应用

论文提交日期： 年 月 日 论文答辩日期：年 月 日

学位授予单位：华南理工大学 学位授予日期： 年 月 日

答辩委员会成员：

主席：

委员：

**华南理工大学**

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本论文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名： 日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

□保密（校保密委员会审定为涉密学位论文时间： 年 月 日），于 年 月 日解密后适用本授权书。

□不保密,同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文编入有关数据库进行检索，传播学位论文的全部或部分内容。

(请在以上相应方框内打“√”)

作者签名： 日期：

指导教师签名： 日期：

作者联系电话： 电子邮箱：

联系地址(含邮编)：

# 摘 要

现有问题的缺陷，针对问题提出方法//

一是提出基于不规则采样时间序列数据预测模型——TransLSTM模型

二是将TransLSTM模型应用于模拟数据集，

三是将TransLSTM模型应用于真实EHR数据集，

关键词：

# Abstract

英文摘要：

**Keywords:**

# 目 录

摘 要 I

Abstract II

目 录 III

第一章 绪论 1

1.1 研究背景与意义 1

1.2 研究历程与现状 2

1.3 论文架构 4

第二章 相关理论知识 5

2.1 数据定义 5

2.1.1 EHR的定义 5

2.1.2 ISMTS 数据定义 6

2.1.3 数据缺失机制 7

2.1.4 缺失值处理方法 7

2.2 机器学习方法 9

2.2.1 逻辑斯谛回归 9

2.2.2 随机森林 10

2.2.3 MLP模型 11

2.2.4 LSTM模型 12

2.3 可视化模型 12

2.3.1 attention机制 13

2.3.2 SHAP（SHapley Additive exPlanations） 13

2.4 本章小结 14

第三章 TransLSTM预测模型设计 15

3.1 引言 15

3.2 算法整体流程 15

3.3 数据预处理 16

3.4 模型预测 18

3.4.1 模型构建 18

3.4.2 模型训练与评估 20

3.5 决策支持 22

3.6 本章小结 23

第四章 TransLSTM模型在EHR数据集上的应用 24

4.1 引言 24

4.2 实验数据集 24

4.2.1 数据来源 24

4.2.2 时间窗口选择 28

4.3 实验结果与分析 29

4.3.1 神经网络架构消融实验结果 29

4.3.2 与传统机器学习对比试验结果 31

4.2 决策支持可视化 33

4.5 本章小结 36

总结与展望 37

参考文献 38

**攻读硕士学位期间取得的研究成果** 39

致谢 40

# 第一章 绪论

## 1.1 研究背景与意义

近年来，时间序列预测和分类任务引起了学术界和工业界的广泛关注，时间序列数据在各个领域被广泛应用，如金融市场股票预测、天气预报、医疗诊断等。然而，传统的预测模型大多假设时间序列数据是均匀采样且完整的[1]。但是在现实世界中，许多时间序列数据往往呈现出不均匀的采样特征，特别是在测量间隔上存在较大差异。这种不规则性通常是由于数据采集设备的限制、人为干预以及事件驱动等因素导致的，例如不同传感器或观测维度的采样频率差异，或是因患者生理状态变化、医疗资源分配不均等原因导致的不规则时间间隔。这类数据被称为不规则采样时间序列（Irregularly Sampled Time Series, ISTS）数据，其特点是观测值之间的时间间隔不恒定，且采样率往往随着时间变化。如何有效处理这种不规则的时间序列数据，尤其是医疗领域的不规则采样医疗时间序列（Irregularly Sampled Medical Time Series, ISMTS）数据，成为一个重要的研究问题。研究人员需要开发新的方法来处理这些数据，以便在不规则采样的情况下仍能准确地进行预测和分类。这不仅涉及到对数据的预处理和建模，还需要考虑如何在模型中有效地利用时间间隔信息，以提高预测的准确性和模型的鲁棒性。

在医疗领域，这种不规则采样的现象尤为常见，尤其是在处理电子健康记录（Electronic Health Records, EHR）时。EHR数据记录了患者的就诊历史、检查结果、用药情况等丰富信息，但这些数据通常是不规则采样的。例如，在重症监护病房（Intensive Care Unit, ICU），患者的生理指标可能会频繁监测，而在普通病房，监测频率则显著降低。此外，EHR中的用药数据包含多种药物，但每个患者的服药时间点、周期和间隔各不相同，表现出高度的不均匀性。这些基于事件的医疗时间序列数据不仅具有复杂的时间结构，还在采样的时序性和内容上呈现出高度的不均匀性。这些复杂的不规则采样医疗时间序列数据不仅具有时间结构上的不均匀性，还涵盖了多模态、高维、噪声和缺失等特点。

尽管如此，ISMTS 数据蕴含着丰富的信息，为疾病风险预测提供了重要的研究契机。例如，MIMIC - III 和 CINC 等真实世界医疗数据集记录了各种疾病的临床信息，包括生理指标、实验室检查结果、药物使用情况等，这些丰富的临床数据为开发精确的疾病风险预测模型奠定了坚实的基础。然而，现有研究在处理 ISMTS 数据时仍面临诸多挑战。一方面，ISMTS数据不仅包含序列内部的复杂关系，还存在不同序列之间的关联性，需要同时考虑局部和全局信息。例如，某些生理指标之间可能存在相互影响，药物使用可能会影响多个生理指标的变化。另一方面，研究方法需要在任务预测准确性与模型复杂性之间进行权衡，同时兼顾模型的通用性和可解释性。不同医疗机构的数据格式、采集标准可能存在差异，这要求模型具有良好的泛化能力，能够处理来自不同来源的医疗数据。模型的预测结果需要具备可解释性，以便医护人员理解和验证预测依据，这对模型设计提出了额外的要求。

近年来，随着深度学习技术的发展，研究人员针对 ISMTS 数据提出了多种建模方法。其中，基于时间嵌入的注意力机制模型通过引入时间编码，能够有效捕捉不同时间点之间的关联性。时间间隔感知的循环神经网络（Recurrent Neural Network, RNN）变体，如T-LSTM和GRU-D，通过在网络结构中显式地建模时间间隔信息，提高了对不规则采样数据的处理能力。基于Transformer架构的改进方法则利用自注意力机制来建模长期依赖关系，并通过位置编码来保留时序信息。这些方法在处理不规则采样数据方面展现出显著优势。但在实际应用中，仍需要考虑数据特征的多样性、模型的泛化能力以及临床可解释性。​研究如何有效地处理 ISMTS 数据并将其应用于疾病风险预测任务具有重要意义，比如在急性肾损伤预测中，通过监测肾功能相关指标的变化趋势，可以提前预警肾功能损害。不仅能够推动医疗数据分析技术的创新，还能为临床实践提供更精准、高效的决策支持。这些预测任务的实现不仅能够推动医疗数据分析技术的创新，还能为临床实践提供更精准、高效的决策支持。本文在现有的预测方法的基础上，着重从两个方面进行创新：一是通过改进模型架构提高预测精度，二是增强模型的可解释性，使预测结果更具临床参考价值。

## 1.2 研究历程与现状

近年来，ISMTS的建模逐渐成为研究热点，研究者们提出了多种建模方法，主要可分为以下两大类：基于原始数据的视角（raw data-based perspective）和基于缺失数据的视角（missing data-based perspective）。基于原始数据的方法直接处理原始的不规则采样数据，可以充分利用不规则采样的时间信息，例如，Baytas等人提出的T-LSTM通过引入时间衰减机制，使得模型能够根据时间间隔的长短调整记忆单元的状态。Che等人提出的GRU-D则通过设计特殊的门控机制来处理不规则采样和缺失值问题。但是当数据是多变量时，基于原始数据的方法需要对不同时间序列进行对齐，又会重新引入缺失数据问题。

基于缺失数据的视角认为，每个时间序列在理想情况下应具有均匀的时间间隔。因此，当将不规则的时间间隔转换为规则的时间间隔时，会出现缺失数据。这类方法的核心在于如何处理这些缺失值，以使数据能够适配传统的均匀采样模型。具体方法包括：（1）统计值填充：根据数据的分布情况采用均值、中位数、众数等进行填充。（2）数据插值法填充：例如随机插值、线性插值、牛顿插值等方式。（3）基于模型填充：采用机器学习模型对缺失值进行预测填充。例如基于循环神经网络（RNN）的插补方法，利用RNN及其变体，如长短期记忆网络(Long Short-Term Memory, LSTM)和门控循环单元(Gated Recurrent Unit, GRU)来插补缺失值。GRU-D通过引入时间衰减机制，将缺失值的上下文信息和时间间隔信息融入模型中，从而更准确地估计缺失值。基于生成对抗网络（Generative Adversarial Network，GAN）的插补方法，通过生成器和判别器的对抗训练来生成缺失数据。例如，GAIN通过引入掩码机制和提示向量，使得生成器能够根据已知数据和缺失模式生成合理的缺失值。基于缺失数据的视角在处理ISMTS数据时具有显著优势，尤其是在数据插补任务中表现出色。基于缺失数据的方法在处理不同领域的ISMTS数据时具有较强的适应性，尤其是在医学领域。并且插补后的数据可以直接用于下游任务（如预测、分类），减少了因数据缺失导致的性能下降。Li等人的研究表明，基于缺失数据视角的方法在多个医疗预测任务中都取得了优异的性能。

ISMTS数据，比如最为典型的电子健康记录 (EHR) 包含大量可以二次开发的信息和用途，例如临床事件预测、药物反应监测和慢性疾病管理。随着医疗信息化的发展，ISMTS数据量呈指数级增长，这为开发更实用的风险预测、疾病预后和慢性病管理模型提供了前所未有的机会。但是ISMTS数据的不规则性给数据分析和建模带来了诸多挑战。大多数传统的回归和机器学习方法无法有效地从包含多组重复变量的数据中提取时间模式，因为这些模型通常假设数据是静态的、样本之间互相独立同分布，且数据的统计特性不随时间变化。然而，时间序列数据本质上是动态的，数据点之间存在时间依赖性。因此研究者使用传统方法时依赖于从时间序列中提取聚合的单个值，例如平均值、中位数或其他聚合统计数据。这种简化处理虽然使得建模变得可行，由于无法利用数据的时间动态，它导致潜在有价值的序列信息的丢失。这些局限性使得传统方法在处理ISMTS数据时表现不佳。

近年来，随着人工智能的蓬勃发展，深度学习方法在准确性上逐渐超越了传统机器学习方法。鉴于其显著优势，许多研究者提出了专门针对ISMTS数据特性的深度学习模型，用于处理诸如疾病风险预测等下游任务。常见的预测疾病包括败血症（Sepsis）、急性肾损伤（Acute Kidney Injury，AKI）、心力衰竭（Heart Failure）等。在这些深度学习方法中，主要包括以下几类模型架构：循环神经网络（RNN）专门用于处理时间序列数据，能够捕捉时间依赖关系；卷积神经网络（Convolutional Neural Network , CNN），用于提取数据的空间特征和局部特征；以及图神经网络（Graph Neural Networks, GNN），可以表示和分析患者内部访问之间的复杂关系网络。深度学习算法具有学习能力、灵活性和通过复杂非线性实现的泛化能力，在建模ISMTS数据时展现出明显的优越性。例如，2019年Tomašev等人提出了一种基于循环神经网络的多任务架构，通过引入L1正则化的深度残差结构，实现了对急性肾损伤风险的连续预测。该模型能够提前48小时预测55.8%的急性肾损伤住院患者，以及90.2%需要透析治疗的重症患者。Lauritsen 等人提出了一种基于 CNN与LSTM结合的模型，该模型首先对输入事件进行向量映射，然后使用CNN提取EHR数据中的空间特征，最后通过 LSTM 捕捉时间依赖性并获得预测结果。实验表明，该模型的预测性能显著优于传统的多层感知机(MLP)和梯度提升模型。

虽然模型的精度和性能一直是研究的重点，但是为了在临床应用时能给医生或者患者提供参考，模型可解释性的研究也是一大重点，尤其是深度学习模型一度被认为是黑盒模型。2017年，Vaswani 等人提出注意力机制使得构建可解释性的神经网络模型成为可能。该机制通过计算输入序列中每个元素之间的相关性权重，使模型能够动态地关注输入中的重要部分，从而提高了模型的性能和可解释性。基于注意力机制的提出，研究者们开发了多种可解释的深度学习模型。例如，2018年Ying Sha等人开发了一种基于门控循环单元（GRU）的循环神经网络，并结合层次注意力机制，通过可视化注意力权重，展示了模型在死亡率预测过程中会更关注近期的医疗访问记录以及与病情严重程度相关的诊断代码。在2019年，Shickel等人开发了名为DeepSOFA的框架，该框架结合门控循环单元（GRU）和自注意力机制，用于评估重症监护室住院期间疾病的严重程度，与传统的SOFA评分系统相比，DeepSOFA在预测院内死亡率方面的平均AUROC达到了0.90（相比之下SOFA为0.79），并且能够实时生成连续的急性评分，为临床决策提供支持。除了注意力机制，其他解释性模型也取得了显著进展。例如2020年Lauritsen等人提出了可解释的人工智能预警评分（xAI-EWS）系统，由时空卷积网络（TCN）预测模块和深度泰勒分解（DTD）解释模块组成，TCN用于从电子健康记录中预测急性危重症，预测准确率达到82.4%, 而DTD则能够以简单的可视化方式展示预测结果的依据，即哪些临床参数在预测过程中起到了关键作用。

在本文中，我们从缺失数据的角度出发，考虑不规则采样的EHR数据，并按照数据预处理、模型构建和模型评估的流程，构建了一个用于疾病预测的算法框架。基于此框架，我们开发了一种名为TransLSTM的综合疾病预测模型。TransLSTM不仅能够精确预测特定时间点急性危重疾病的发作，还能借助Transformer的自注意力机制，为预测结果提供简明的可视化解释。该模型将预测结果与相关的EHR数据相结合，帮助临床医生做出更明智的决策，从而显著增强了其在临床应用中的可行性。在MIMIC-III和MIMIC-IV数据集上的实验表明，我们的模型在预测准确性和可解释性方面都取得了令人满意的结果。

## 1.3 论文架构

本论文共分为五章，具体内容安排如下：

第1章为绪论。主要介绍了本论文研究课题的背景和意义，系统回顾相关课题的发展历程和现状，并说明本文的研究内容和创新点。

第2章为相关理论基础，重点介绍本研究涉及的核心概念和关键技术，包括ISMTS数据的定义与特征、数据预处理方法，以及模型构建所需的主要算法，包括逻辑斯蒂回归、随机森林、LSTM、注意力机制等。

第3章为本文预测模型的设计思路及层次结构概述。详细阐述本文所提出模型的整体架构和工作流程，包括数据预处理模块、模型构建模块和评估模块的设计原理与实现方法，以及如何将模型预测结果转化为可用的临床决策支持信息。

第4章是介绍了实验数据集，实验方法以及本文的研究结果展示和分析。

第5章是总结和展望，对本论文研究进行总结，并对现有方法的不足提出未来可研究深化的方向。

# 第二章 相关理论知识

在基于不规则采样医疗时间序列数据的建模中，预测任务不仅依赖于患者的历史医疗数据，还需要充分考虑时间因素对疾病发展的影响。因为患者的身体指标变化可能随着时间的推移而逐步显现。因此，本文中的预测任务的时间序列特性使得对时间的处理尤为重要。

进行时间序列预测时，通常需要对原始数据进行一系列的预处理操作，以确保数据的质量和一致性，并为后续建模做好准备。这些预处理步骤通常包括数据提取、数据清洗、数据填充和特征工程等。从而形成可以用于建模的时间序列数据。本文中的预测任务旨在通过分析时间序列数据，预测患者是否会发生急性肾衰竭这一事件，属于离散型预测任务。因此，本研究的预测任务不仅需要准确捕捉患者的历史健康信息，还需充分考虑时间因素对身体指标变化的影响，以提高预测的准确性和可靠性。

在本章中将介绍一些常见的时间序列预测技术相关知识，主要包括三个方面：数据的形式化定义、预测模型的数学原理，以及其他与本研究密切相关的关键技术。这些理论知识将为后续的模型设计和实现提供重要支撑。

## 2.1 数据定义

### 2.1.1 EHR的定义

不规则采样的医疗时间序列（Irregularly Sampled Medical Time Series, ISMTS）广泛存在于临床实践中，比如电子健康记录（Electronic Health Record, EHR）具有显著的时间序列特征。EHR 最初是用来存储医疗信息，但随着信息技术的发展和医疗研究的深入，研究人员发现EHR中包含丰富的信息可以为多种临床信息学应用提供有效的数据支持。EHR 存储了患者每次就诊相关的数据，包括人口统计信息、当前和过去的诊断、实验室检查和结果、处方、放射图像、临床笔记等。EHR系统中包含大量不规则采样的时间序列数据，这与医疗服务的特性密切相关：同一患者由于身体健康状况的变化可能多次入院治疗，每次入院过程中会生成多个时间序列记录。例如，患者两次入院之间的间隔可能仅为数月，也可能长达数年。即使在同一住院期间，不同生理指标的测量频率也可能存在显著差异。在病情监护过程中，医疗服务提供者会根据患者的实时状况动态调整监测频率：当发现患者肾功能指标出现恶化趋势时，监护人员会增加相关生理变量的测量次数；而当病情趋于稳定或好转时，则会相应减少测量频率。

电子健康记录（EHR）由多个患者的医疗记录组成，具体可以表示为

其中为患者记录的总数。对于第个患者​, 其医疗信息包含人口统计信息, 如姓名、年龄、性别、地址等；入院记录，每位患者可能具有 ​条入院记录。患者的入院记录集合可以表示为

其中每条入院记录 由多个医学代码组成，包括静态诊断代码集合 （如疾病编码 ICD-10 和手术编码 ICD-9-CM）和动态生命体征代码集合（如血压、心率、血氧饱和度等）. 此外，每个代码都附带一个时间戳 ,用于记录该数据的具体发生时间。

### 2.1.2 ISMTS 数据定义

一个具有D维变量或者特征的不规则采样的医疗时间序列（ISMTS）可以表示为

其中，每个维度（即每个变量或特征）包含一个长度为的观测值列表。对于第个特征，其观测值序列表示为：

与之对应的时间戳序列为：

ISMTS 具有以下两类时间不规则性：

1. 同一维度内的不规则性（Intra-series Irregularity）：在同一个维度，相邻观测值之间的时间间隔可能不同，即当时：
2. 不同维度间的不规则性（Inter-series Irregularity）：对于不同的维度和(),观测时间戳序列与可能不同，即

此外，不同特征的观测次数可能不相等，即

多变量 ISMTS 数据的全局时间戳集合可表示为：

在时间点 处，对齐后的多变量 ISMTS可以表示为：

由于数据的不规则性，在对齐后的数据集中可能存在缺失值，即某些特征在特定时间点没有观测值：

### 2.1.3 数据缺失机制

ISMTS 数据通常具有非均匀的时间间隔，这种不均匀性源于三个原因：1）由于传感器损坏、数据传输失败或存储器损坏，时间序列中存在缺失数据。2）采样机本身的采样率不恒定。3）不同的时间序列通常来自具有不同采样率的不同来源。我们基于缺失数据的视角，将不规则序列视为有缺失数据，可以通过更精确的数据计算来解决问题。通常，在数据集中，不含缺失值的变量被称为完全变量(，含有缺失值的变量则称为不完全变量(。在数据缺失的处理过程中，Little 和 Rubin 提出了三种常见的数据缺失机制：

1. 完全随机缺失(Missing Completely at Random, MCAR)：缺失的数据与任何变量（包括不完全变量和完全变量）都没有关系，即数据的缺失无法通过其他变量预测或解释；
2. 随机缺失(Missing at Random, MAR)：缺失的数据仅与完全变量有关，在这种情况下，缺失的数据模式可以通过完全变量进行建模；
3. 非随机缺失(Not Missing at Random, NMAR)：缺失的数据与不完全变量本身相关，即数据的缺失是依赖于缺失值所在的变量。这种类型的缺失无法通过完全变量来解释，因此它是不可忽略的，需要特别处理。

根据数据的性质，通常认为ISMTS的缺失机制是 “随机随机”或“完全随机缺失”。

### 2.1.4 缺失值处理方法

处理缺失数据的方法主要有以下几种：

1、 删除特征：当特征数据具有高缺失率、低重要性、低覆盖率时，可以将其删除。

2、 数据填充：基于统计学原理，根据数据集其余特征值的分布情况来对一个空值进行填充，本文中主要采用了以下几种方法：

1. 最后观测值前向填充（LOCF , last observation carried forward）：对于临床数据，有的人可能第1次随访有数据，而但第2次（或第3次、第4次）就失访了，后面就没有数据了。LOCF方法就是用前一个记录值填补现在的缺失值。但是EHR数据中，通常有多个患者ID, 每个ID有多个特征，如果某个ID的第一条记录缺失，按照LOCF的定义，这条记录不存在靠前的记录，所以不能使用LOCF填补。所以在进行LOCF初步处理后还需要进行进一步填充。
2. 统计值填充：数据集中的特征可分为数值型和非数值型。如果空值是数值型的（如年龄、体重、疾病评分等），可根据该特征其他所有非空值的平均或中位数来填充；如果空值是非数值型的（如种族、疾病类型等），根据众数原理，取该特征出现次数最多的值（即出现频率最高的值）来补齐该缺失的属性值。
3. KNN（K-Nearest Neighbors，K近邻）填充方法：核心思想是通过测量缺失值样本与其他非缺失值样本之间的相似性，找到与缺失值样本最相似的K个邻居（K个最近邻），然后利用这些邻居的特征值来填充缺失值。医疗数据通常涉及多个维度（如生命体征、实验室指标、用药记录等），而这些特征可能同时存在缺失值。KNN填充方法能够同时考虑多个特征的信息，综合相关特征的信息，从而更准确地填充缺失值。

KNN填充的第一步是计算缺失值样本与所有其他非缺失值样本之间的距离，常用的欧氏距离适用于数值型数据，对于维的样本：

其中，和是不同样本的第个样本。在计算完缺失值样本与其他所有样本之间的距离后，选择个距离最近的样本作为近邻，对每个缺失值样本的距离进行排序，从最近到最远排列。根据排序结果，选择距离最小的个样本作为近邻。然后计算个近邻的加权平均进行填充，加权平均考虑了邻居之间的距离差异，权重的计算是基于距离的倒数：

最后计算的加权平均值为：

## 2.2 机器学习方法

在人工智能领域，预测模型通常可分为三类：一类是传统的基于机器学习的模型，如逻辑回归(LR)、支持向量机(SVM)；另一类是基于深度学习方法的模型，如多层感知机(MLP)、卷积神经网络(CNN)等；基于集成学习的模型，比如基于Bagging方法的随机森林(RF)、基于Boosting方法的梯度提升决策树(GBDT)等。

### 2.2.1 逻辑斯谛回归

逻辑斯谛回归（Logistic Regression，LR）是统计学中的经典分类方法，通过分析输入变量（特征）与输出结果（疾病发生与否）之间的关系，根据现有的数据对分类边界线建立回归公式，得出患者患病风险的概率，以此进行分类。逻辑斯谛回归的核心思想是将线性回归的输出映射到一个范围为0到1之间的概率值。多元线性回归方程的一般形式如下：

其中是输入特征，是偏置项，是特征的权重系数。

逻辑斯谛回归通过Sigmoid函数：

将线性组合映射到区间，表示患病的概率，应用Sigmoid函数后的逻辑斯蒂回归模型可以表达为：

根据设定的阈值（通常为0.5）来映射到类别标签，如果预测概率高于阈值，则模型预测为患病，否则预测为未患病。逻辑斯谛回归通过最大化数据的似然函数来训练模型，即找到一组最佳回归参数使得所有样本的损失之和最小。在逻辑斯蒂回归中，衡量模型预测结果与真实标签之间差异的指标通常为交叉熵损失函数：

其中为第个样本的预测为1的概率，是第个样本的真实标签。为了最小化损失函数，逻辑回归通常使用梯度下降（Gradient Descent）优化算法，即通过计算损失函数对各个参数的偏导数（梯度），然后调整参数的值，逐步降低损失。主要步骤如下：

计算损失函数对各参数的偏导：

是第个样本的第个特征值。

根据梯度和学习率更新参数：

### 2.2.2 随机森林

随机森林（Random Forest）是一种基于集成学习思想的机器学习算法，由Leo Breiman在2001年提出。它通过结合多个决策树模型进行预测，从而提高分类或回归任务的准确性和稳健性，同时减少过拟合风险。随机森林的核心思想是通过集成多个相对独立的弱学习器（即决策树），构建一个性能更优的强学习器。随机森林的构建和训练主要是利用Boosttrap方法随机从数据集中有放回地生成个训练集，每个子集的大小为, 第个样本子集可表示为：

为输入特征，为对应标签。对每个采样子集构建一棵决策树.在构建过程中，随机森林引入了特征随机选择的机制，每次分裂节点时，从全部个特征中随机选择 个特征进行分裂（通常 或 ）. 假设节点分裂时选择的特征为, 分裂的依据是优化基尼系数（Gini Impurity）：

其中是节点中属于类别的样本比例。或优化信息增益（Information Gain）：

在模型预测阶段，随机森林采用Bagging（Bootstrap Aggregating）策略，将棵决策树的预测结果结合。本文代码实现过程中，利用python的 scikit-learn库中 ensemble模块下的 RandomForestClassifier() 函数，使用模型默认的超参数（如n\_estimators和max\_depth）实现模型训练和预测过程。

### 2.2.3 MLP模型

多层感知机（Multilayer Perceptron, MLP）是人工神经网络（ANN）的一种经典结构, 是一种前馈神经网络，由输入层、一个或多个隐藏层以及输出层组成。输入层负责接收样本特征数据，每个输入特征连接到隐藏层中的所有神经元，通过前向传播计算输出。前向传播包括：

1. 线性变换：每层的神经元接收上一层的输出，并对其进行线性变换。对于第层，第个神经元的线性变换公式为：

是连接权重，是偏置，是上一层输出（当前层输入）。

1. 非线性激活：每个神经元的输出通过激活函数计算，在本实验中采取sigmoid激活函数：
2. 输出：输出层使用 Softmax 函数将输出转化为概率分布：

然后通过反向传播优化权重和偏置，反向传播通过链式法则计算损失对模型参数的梯度，本实验中采用交叉熵损失：

再迭代训练，直至损失收敛。

### 2.2.4 LSTM模型

|  |
| --- |
| *图示  描述已自动生成* |
| Figure 2-1．LSTM网络的单元结构 |

LSTM模型能够捕捉时间上的依赖关系和动态变化。对于每个时间步，每个LSTM单元接收当前时间步的特征输入和上一步的隐藏状态, 并输出当前时间步的隐藏状态 . LSTM单元的结构包括遗忘门（）、输入门（）和输出门（）. LSTM第一步是用来决定什么信息可以通过细胞状态(cell state)。这个决定由“forget gate”层通过sigmoid来控制，它会根据上一时刻的输出通过或部分通过。计算公式如下：

第二步是产生需要更新的新信息。这一步包含两部分，第一个是 “input gate”层通过sigmoid来决定哪些值用来更新，第二个是tanh层用来生成新的候选值相加，得到了候选值。一二步结合起来就是丢掉不需要的信息，添加新信息的过程：

最后一步是决定模型的输出，首先是通过sigmoid层来得到一个初始输出，然后使用tanh将值缩放到到间，再与sigmoid得到的输出逐对相乘，从而得到模型的输出。

## 2.3 可视化模型

本节将阐述实验中用于解释性分析的模型

### 2.3.1 attention机制

注意力机制（Attention Mechanism）能够具有选择性地关注输入信息中某些部分，从而在处理序列数据时能够更有效地捕捉到重要信息。尤其是在Transformer模型中，它对于捕捉长距离依赖关系和提高模型性能起到了关键作用。注意力机制的核心思想在处理输入数据时，通过“注意”某一特定部分的信息，从而生成更有效的输出。

注意力机制的输入数据通过线性变换生成三组向量：查询Query (Q)、键Key (K) 和 值Value (V). 和向量是用来计算注意力权重的原始向量，向量用于生成最终的注意力输出。这三个向量通常是通过线性变换（全连接层）实现的：

其中是输入序列的嵌入向量矩阵，大小为, 其中是序列长度，是模型的维度。是可学习参数矩阵。通过和的点积计算注意力分数，然后应用Softmax函数将注意力分数转化为有效的权重，最后利用注意力权重对值向量进行加权求和，得到最终的注意力输出：

其中​是键向量的维度，用于缩放以稳定梯度。

多头机制使用多个不同的注意力头并行计算，每个头使用不同的映射矩阵。单个注意力输出可表示为：

多头注意力的最终输出是由多个单头注意力头的输出拼接后经过线性变换得到的：

其中，是注意力头的数量，是可学习的线性变换矩阵。

### 2.3.2 SHAP（SHapley Additive exPlanations）

SHAP（SHapley Additive exPlanations）是一种基于博弈论的解释机器学习模型的工具，提供了对模型输出的解释性。SHAP 的核心思想是使用 Shapley 值来分配每个特征对模型预测的贡献，从而揭示特征对预测的影响。在机器学习中，模型的特征被看作“玩家”，模型的预测被看作“收益”，目标是计算每个特征对预测结果的“边际贡献”。假设有一个模型和个特征，需要衡量每个特征对预测的贡献，对于一个特定的样本, Shapley 值是通过考虑所有可能的特征子集（特征的排列顺序）并对所有可能特征排列的加权平均来计算特征的边际贡献：

其中表示不包含特征的特征子集，是模型在特征集合下的预测结果，是特征集合上加上特征后的预测结果，和是排列的权重，是所有特征的全排列数，是所有特征的全排列数。边际贡献最直观的反映了对于特征子集,如果加入特征,模型的输出变化是多少。

SHAP 基于 Shapley 值的数学原理，提供了对机器学习模型预测的公平解释。通过计算每个特征的边际贡献，Shapley 值能够准确地衡量每个特征在特定预测中的作用。

## 2.4 本章小结

本章主要介绍了关于数据的定义，缺失值处理的步骤方法、一些传统的机器学习方法以及可视化模型和工具。下一章将开始介绍算法的整体框架，包括数据预处理、模型构建、决策支持等。

# 第三章 TransLSTM预测模型设计

## 3.1 引言

随着深度学习算法的迅速发展，深度学习算法在各个领域的应用日益广泛。特别是在医疗卫生领域，深度学习为疾病风险预测提供了强大的工具，推动了疾病风险预测和个性化医疗的发展。尤其循环神经网络（RNN）作为深度学习的重要分支，能挖掘序列数据的时序信息以及潜在的语义信息特性。这一特性使其特别适合处理电子健康记录（EHR）等医疗时序数据，进而实现疾病风险的早期预测。LSTM 通过引入门控机制有效解决了传统RNN在处理长序列时面临的梯度消失问题，使网络能够保存和利用长期依赖信息。尽管LSTM在时序数据处理方面表现出色，但也存在明显的局限性，主要体现在无法充分利用多源信息以及缺乏可解释性。

针对上述挑战，我们提出了一种创新性的可解释人工智能系统——TransLSTM，该系统结合了Transformer和LSTM的优势，使用Transformer提取全局特征并结合 LSTM 模型提取时间序列特征进行预测，在保持高预测性能的同时，能够向临床医生直观展示预测结果背后的决策依据，从而为临床决策提供可信、透明的辅助支持，具有重要的临床转化价值和应用前景。

## 3.2 算法整体流程

|  |
| --- |
| 图示  描述已自动生成 |
| **图3-1 TransLSTM算法整体流程** |

根据问题背景和研究目的，并借鉴常用的数据挖掘流程，我们将TransLSTM模型设计为如图3-1所示的层次结构，主要包含三个核心部分：

第一部分为数据预处理。通过SQL查询语句，基于研究问题的相关医学知识背景提取相关临床数据。然后将动态数据与静态数据按不同的方式进行处理。根据疾病预测相关医学知识对不同特征做不同处理，包括数据离散化、变量编码、缺失值的处理等。人口统计变量和入院基本信息（如年龄、性别、既往病史等）视为静态变量，保持其在整个住院期间的不变性；而生命体征（如心率、血压、体温等）和药物信息则作为动态变量，记录其随时间的变化趋势。最后，将静态和动态特征按照预设的时间窗口整合，形成用于模型训练和预测的结构化输入数据。

第二部分为模型训练与评估。本研究使用TransLSTM模型进行疾病风险预测，该模型结合了Transformer和LSTM两种深度学习架构的优势。其中，Transformer块用于捕捉不同特征之间的依赖关系，通过自注意力机制增强模型对全局信息的理解；LSTM网络则负责处理时间序列数据，通过门控循环机制捕捉时间依赖性。这种混合架构设计使模型能够同时从全局特征关联和局部时间序列模式中提取有价值的信息，从而显著提高疾病风险预测的准确性和可靠性。

第三部分为可解释的决策支持。为增强模型的临床实用性，我们设计了多层次的可解释性分析框架：在群体层面，提供广泛的决策支持分析，帮助医疗团队理解模型在全体患者群体中的应用模式和预测依据，识别关键风险因素和普遍临床规律；在个体层面，提供定制化的决策支持分析，通过可视化展示特定患者的风险预测结果及其背后的关键影响因素，辅助医生针对个体患者制定更为精准的治疗决策和干预方案。通过这种结合全局和个体层面的解释性分析，TransLSTM模型不仅能显著提高预测的透明度，还能增强医疗专业人员对模型预测结果的信任度，促进人工智能技术在临床实践中的有效应用。

## 3.3 数据预处理

**图** 展示了针对 EHR 系统中不同类型数据的处理流程。原始数据主要包括三类：人口统计学数据、实验室监测数据以及临床干预数据。针对不同数据类型，采取相应的数据清理和预处理步骤，以确保数据质量，提高后续分析和预测的可靠性：

|  |
| --- |
| 图示  描述已自动生成 |
| 图3-2 EHR数据预处理流程 |

1）人口统计学数据通常包括患者的年龄、性别、入院类型等基本信息。这些数据在患者整个住院期间通常变化较小，我们主要采取经验筛选的方法，选取具有普遍临床意义且相关性较高的特征进行分析。

2）实验室监测数据来源于 ICU 监测设备和各类实验室检查，通常以时间序列形式存储，具有较高的缺失率和噪声。针对这类数据，我们采取以下处理流程：

表格透视转换，将时间序列数据按照不同的指标进行展开，让原本按时间顺序排列的纵向数据结构得以重塑，能够以更为直观的方式展现数据间的内在关联与分布规律，使数据结构更适合后续分析。

缺失值筛选特征，针对数据中某些监测指标只有极少部分患者有值的情况，对缺失率较高（>=70%）的特征进行筛选剔除，因为大量数据的缺失，会显著加剧数据的稀疏程度。通过将这些高缺失率特征从数据集中移除，能够有效降低数据的稀疏性，从而减少计算复杂度。

离群值处理，剔除因测量误差或设备故障导致的异常值，确保数据的真实性和一致性。我们的实验通过计算5%和95%的百分位数，将所有数值限制在该区间内。具体而言，将低于5%分位数的特征值替换为该特征5%分位数的值。将高于95%分位数的值替换为95%分位数的值。这种方法既保留了所有样本，又维持了数据的整体分布特性，避免了因极端值剔除导致的数据量损失，有助于模型更准确地捕捉数据中的真实模式。

时间点聚合，将数据按照一定时间间隔进行聚合，以规范时间步长，提高时间序列的可对比性。考虑医学数据的特点，我们采取一小时作为时间间隔，通过计算每小时内所有记录的平均值作为该小时的代表性值，从而形成等间隔的时间序列数据。聚合完成后，为了处理时间序列中仍然存在的空缺值，我们采用LOCF（Last Observation Carrie。时间点数据聚合后，采用 LOCF（Last Observation Carried Forward）填补方法初步处理时间空缺，即用最近一次观测到的有效值填充后续缺失时间点，使时间序列更加完整。

关于数据透视、时间点聚合记忆数据整合的具体实例如图：

|  |
| --- |
|  |

3）临床干预数据涉及患者在住院期间接受的各类治疗和干预措施，如药物使用、手术、护理记录等。对于这类数据，我们采用与处理实验室数据相似的策略：首先进行表格透视转换，将不同类型的干预措施按时间点展开；随后应用时间点聚合技术，确保所有临床干预数据在时间维度上的一致性和可用性，便于与其他类型数据的整合分析。

经过上述一系列的数据筛选、清洗、转换和填补步骤，各类数据最终被整合为结构化的数据集，符合机器学习模型的输入要求。然而，由于医疗数据的不规则性，整合后的数据集中仍然存在一定比例的缺失值。为解决这一问题，我们采用K最近邻（KNN）方法进行最终的数据填充，该方法能够基于相似样本的特征值为缺失项提供合理估计，有效保留数据间的内在关联性，为后续模型训练提供更为完整和可靠的数据基础。

## 3.4 模型预测

### 3.4.1 模型构建

**图示

描述已自动生成**

TransLSTM是Transformer和LSTM模型的结合体，其具体框架如**图**所示。特征数据首先被输入到Transformer模块，该模块包括多头注意力机制（Multi-Head Attention）、归一化层（Normalization Layer）和多层感知机（MLP）。Transformer模块提取输入序列的特征，每个患者的输入序列

其中是特征数量，每个特征 是一个包含个时间步的时间序列。首先通过线性变换：

生成查询（Query）、键（Key）、值（Value）矩阵，进一步计算注意力分数：

通过并行计算多个注意力头，每个注意力头关注输入序列的不同部分，捕捉不同特征之间的复杂关系。将每个头的输出拼接得到多头注意力机制结果输出为：

其中，是注意力头的数量，是可学习的线性变换矩阵。多头注意力的输出通过残差连接（Residual Connection）后，再进行归一化：

其中，和是输入的均值和标准差，和是可学习的参数。归一化后的特征输入到前馈网络（Feed-Forward Network, FFN）中，前馈网络由两个全连接层组成，第一层进行非线性变换：

第二层通过线性变换，恢复到原始特征维度：

其中，为权重矩阵，为偏置向量。最后前馈网络的输出与多头自注意力机制的输出进行残差连接，和层归一化处理：

患者的数据经过Transformer模块特征提取的过程可以表示为：

维度为, 然后Transformer的输出序列将作为输入传递到LSTM模型。

LSTM模型能够捕捉时间上的依赖关系和动态变化。对于每个时间步到, 每个LSTM单元接收当前时间步的特征输入 和上一步的隐藏状态 以及细胞状态，然后计算遗忘门（）：

遗忘门（）根据上一时间步的输出通过sigmoid函数来控制哪些信息可以通过细胞状态（cell state）；输入门（）通过sigmoid函数决定哪些值用来更新，并使用tanh函数生成新的候选值（）：

遗忘门（）和输入门（）结合起来丢弃不需要的信息，添加新的信息，更新当前时间步的细胞状态（）：

输出门（）决定模型的输出：

LSTM层的处理结束后，最终的隐藏状态被传递到全连接层（MLP）, 得到模型的最终输出：

### 3.4.2 模型训练与评估

在电子健康记录中，患病案例要远远多于非患病案例，尤其在ＩＣＵ中急性疾病发生风险较高，患病案例的数量显著超过非患病案例。这就使得正样本（患病案例）的数量远远多于负样本（非患病案例），这导致严重的类别不平衡问题。在模型训练过程中，占多数样本类别会对模型参数的更新产生更大影响，使得模型在预测时倾向于将样本判定为多数类，这会使得模型的训练将使模型产生偏差，不利于稀有类的识别。因此在模型训练之前我们先对数据进行不平衡处理，具体的处理方式是，根据正负样本的比例，对样本量多的类别进行多次不放回的采样，每个采样子集与少数类别组成一个相对平衡的数据集。然后对每个相对平衡的数据集分别进行训练建模，得到多个模型，通过取所有模型的平均值得到最终结果。计算方式如下：

完成类别不平衡处理后，我们按照80%、10%和10%的比例将数据随机划分为训练集、验证集和测试集。模型评估基于测试集上表现最优的模型结果，训练过程中使用Adam优化器最小化交叉熵损失函数，以提高模型对AKI预测的准确性和可靠性。

在本文的实验中，根据AKI疾病是否发生作为标签，因此是二分类任务。在分类任务评估中，混淆矩阵（Confusion Matrix）是一种基础且直观的评估工具，它通过对比真实类别和预测类别的关系，全面展示模型的分类性能。二分类任务的混淆矩阵结构如表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| 混淆矩阵 | | 真实值 | |
| Positive | Negative |
| 预测值 | Positive | TP | FN |
| Negative | FP | TN |

TP（True Positive）表示模型正确识别出的正样本数量，即真实值为positive且预测也为positive的样本；FN（False Negative）表示模型错误地将正样本预测为负样本的数量，这在统计学中被称为第二类错误（Type II Error），代表模型的漏诊情况；FP（False Positive）表示模型错误地将负样本预测为正样本的数量，这是统计学上的第一类错误（Type I Error），代表模型的误诊情况；TN（True Negative）表示模型正确识别出的负样本数量，即真实值为negative且预测也为negative的样本。

理想的分类模型应当使TP与TN的数量最大化（即提高正确预测的比例），同时将FP与FN的数量最小化（即降低错误预测的比例）。基于混淆矩阵，我们可以计算多个关键的模型评估指标，以全面评估模型在各方面的表现。为了体现模型对每一类别样本的判断能力，我们考虑使用灵敏度（Sensitivity）和特异度（Specificity）以及F1-Score 进行判断。灵敏度和特异度的计算方式如公式：

灵敏度越高，表示模型越能准确地识别出患有AKI的患者；而特异度越高，表示模型越能准确地识别出未患AKI的健康患者，减少误诊风险。

F1-Score 作为精确率（Precision）和召回率（Recall）的调和平均数，能够在两者之间取得平衡，特别适用于处理类别不平衡的数据集，F1-Score越接近1，表示模型在精确率和召回率之间达到了更好的平衡。计算公式为：

其中，精确率（Precision）的计算公式为：

此外，在分类任务中，分类模型通常需要设定一个概率阈值来将连续的预测概率转化为离散的类别标签。不同阈值的选择会直接影响模型的性能表现。为了全面评估模型在不同阈值下的表现，我们采用了接收者操作特征曲线（Receiver Operating Characteristic Curve, ROC）和曲线下面积（Area Under the ROC Curve, AUROC）作为重要评估指标。

ROC曲线通过将模型在不同阈值下的真阳性率（True Positive Rate, TPR）作为纵轴，假阳性率（False Positive Rate, FPR）作为横轴进行绘制，直观展示了模型在各种阈值设置下的性能变化。TPR和FPR的计算公式分别为：

AUROC是ROC曲线下方的面积，其数值范围在0到1之间，数值越接近1，表示模型的分类性能越优秀。AUROC的一个重要优势在于，它不受类别不平衡问题的影响，能够在样本分布不均的情况下仍然对模型做出公正的评价。AUROC的计算方式可以表示为：

其中，表示的是选定的分类阈值的数量。AUROC值为0.5表示模型的预测能力与随机猜测相当，而AUROC值超过0.7通常被认为具有良好的预测性能，超过0.9则被视为优秀的预测模型。

## 3.5 决策支持

在患者个体层面，基于Transformer模型中的注意力机制被用来捕捉序列中不同位置之间的依赖关系。多头注意力机制是基于自注意力机制（Self-Attention Mechanism）发展而来的技术，通过并行计算多组注意力，从而增强模型对不同特征子空间的关注。每组注意力称为一个头（head），每个头学习一组不同的权重，用于计算每个时间点的注意力权重。如图5所示，多头注意力机制的计算过程可以分为以下几个步骤：

首先，输入数据矩阵分别通过三个线性变换层得到查询矩阵、键矩阵和值矩阵：

然后，使用查询矩阵和键矩阵计算注意力权重：

其中，是键向量的维度，softmax函数用于归一化以得到注意力权重。注意力权重表示模型在处理当前时间步时，对序列中其他时间步输入的关注程度。权重越高，表示该时间点的信息对当前预测越重要。

最后，对于模型解释，我们可以分析注意力权重来理解模型的决策过程。如果某个特征在特定时间点上的注意力权重很高，这表示该特征在这些时间点上对模型预测具有比较显著影响。对于每个特征，将所有时间步的注意力权重进行平均，得到时间平均贡献：

其中，是序列的时间总长度。通过自注意力机制，可以直观量化每个特征对模型预测的整体贡献，增强模型解释性，帮助临床医生更好地理解模型的决策过程。

在患者群体层面，我们采用了SHAP（SHapley Additive exPlanations）方法，这是一种基于博弈论的技术，通过计算每个特征对预测结果的边际贡献，帮助理解模型的整体决策过程。SHAP值能够揭示每个特征在不同决策中的重要性，使我们能够详细了解模型是如何做出每一个预测的。对于每个特征，Shapley值计算该特征在所有可能的特征组合中的边际贡献。SHAP值基于Shapley值概念，能够公平地分配每个特征对预测结果的贡献，同时考虑特征间的交互作用。对于每个特征, Shapley值计算该特征在所有可能的特征组合中的边际贡献，并按照特定权重进行加权平均。计算公式如下：

其中是所有特征的集合，是不包含特征的任意特征子集，是只使用特征子集的模型预测输出，是特征的Shapely值。分数项表示权重，确保所有可能的特征排列被公平考虑。

通过汇总所有数据点的SHAP值，可以得到全局特征重要性排名，了解哪些特征在整体上对模型预测结果影响最大。此外，SHAP值的正负还能指示特征对预测结果的促进或抑制作用，为临床决策提供更细致的指导。

结合个体层面的注意力分析和群体层面的SHAP分析，我们构建了一个全面的决策支持系统，既能解释单个患者的预测结果，又能揭示整体人群中的风险因素模式，为临床医生提供多维度的决策参考。

|  |  |
| --- | --- |
| 图示  描述已自动生成 |  |
| **Figure 5. a：Self-Attention机制,b：注意力计算步骤** | |

## 3.6 本章小结

本章主要介绍了TransLSTM预测模型架构的设计，主要包含三个核心部分：1）数据预处理阶段，针对人口统计学数据、实验室监测数据和临床干预数据进行了系统化的清洗和标准化处理；2）模型训练与评估阶段，通过多头注意力机制和LSTM的序列建模能力实现了对医疗数据的深度分析；3）决策支持阶段，结合注意力机制和SHAP方法，在个体和群体两个层面提供了模型预测结果的可解释性分析，为临床决策提供了可信、透明的辅助支持。

# 第四章 TransLSTM模型在EHR数据集上的应用

## 4.1 引言

急性肾损伤（AKI）是指肾脏在数天至数周内迅速丧失清除血液中代谢废物的能力，是一种肾脏功能急剧衰退的病症。评估心脏衰竭患者的预后时，肾功能的状况具有至关重要的意义，越来越多的临床医生开始关注这一问题。心脏和肾脏在执行各自的生理功能时，彼此之间有着紧密的联系。心衰患者发生AKI后出现的症状和心衰症状较相似，比如都会出现少尿、水肿等症状，这使得心衰患者发生AKI的临床表现容易被忽视。在美国，每年有五百万左右的心衰患者，其中有27％至40％的患者会发生急性肾损伤。急性肾衰竭的预测不仅依赖于患者的历史医疗数据，还需要充分考虑时间因素，因为肾功能的变化可能随着时间的推移而逐步显现。通常，医护人员会通过检测血清肌酐（SCr）或尿量的变化来预警患者发生AKI。然而，血清肌酐（SCr）的升高往往滞后于肾损伤的实际发生时间，这会导致治疗延迟，并对患者产生不良后果。

急性肾衰竭的精准预测，既依赖于患者的历史医疗数据，又需充分考量时间这一关键因素。随着医疗信息数字化程度的不断提高，大量高时间分辨率的患者数据被集成到现有的电子健康记录（EHR）系统中。EHR是典型的ISMTS数据，具有不规则性。由于临床实际操作中的各种原因，诸如检测仪器故障、患者拒绝某些检测等，使得 EHR 数据集中并非每个时间点都具备完整的血清肌酐、尿量等与 AKI 密切相关的检测指标。而且，不同医院甚至同一医院不同科室对于 EHR 数据的记录规范和标准也不完全统一，进一步加剧了数据的不规则性。尽管 EHR 数据存在这些复杂的特性，但因其蕴含着丰富的患者长期健康信息，对于准确评估患者病情，尤其是像预测心衰患者发生 AKI 这类情况，有着不可替代的潜在价值。我们提出的可解释人工智能系统——TransLSTM，通过数据预处理、模型预测与评估以及决策支持，不仅可以将不规则的数据梳理规整，转化为可供模型高效处理的标准格式，实现了不规则数据处理的流程化，而且很好的平衡了模型预测的准确性与可解释性。

## 4.2 实验数据集

### 4.2.1 数据来源

本文中用到的数据主要来自MIMIC-III（Medical Information Mart for Intensive Care III）数据库和MIMIC-IV（Medical Information Mart for Intensive Care IV），是广泛用于临床医学研究的公开数据库。这两个数据库的数据均来源于美国波士顿的贝斯以色列女执事医疗中心（Beth Israel Deaconess Medical Center，BIDMC）的重症监护病房（ICU）患者，包含了大量关于重症监护病房(ICU)患者的临床数据，涵盖了患者的人口统计学特征、生命体征、实验室检查结果、用药记录以及临床诊断等多维度信息。MIMIC-III数据的收集时间是2001年6月至2012年10月，而MIMIC-IV数据集则是2008年至2019年，尽管两个数据集的时间范围有部分重叠，但它们所包含的患者群体是不同的。

从MIMIC-III和MIMIC-IV数据库中提取数据的流程主要包括以下几个步骤：首先需要注册并获取访问权限，这要求研究人员同意相关的数据使用协议，并完成必要的伦理培训课程（如CITI Program的人类研究保护课程）；其次，将数据导入到本地或云端的数据库管理系统中；最后，通过编写SQL查询语句提取研究所需的特定数据。MIMIC数据库的结构设计非常完善，包含多个相互关联的表格，每个表格存储不同类型的临床数据，如患者基本信息表（PATIENTS）、入院记录表（ADMISSIONS）、生命体征监测数据表（CHARTEVENTS）、实验室检查结果表（LABEVENTS）等，这些数据都存储在PostgreSQL数据库系统中，便于高效查询和分析。

在本研究中，我们通过SQL脚本从数据库中提取了与急性肾损伤（AKI）预测相关的多种数据，包括患者的实验室检查结果（特别是肾功能相关指标如血清肌酐、尿素氮等）、生命体征数据（如血压、心率、体温等）以及患者的临床历史记录（如既往疾病史、用药史等）。

在我们的研究队列中，MIMIC-III包括 61532 例 ICU 住院记录。纳排标准如下：排除了已经患有终末期肾病(End stage renal disease, ESRD)的患者，因为这些患者的肾功能严重受损，不适合用于AKI的预测研究；排除住院时长不足1天或住院时长大于30天的患者，以避免短期住院无法充分观察疾病发展或长期住院可能涉及复杂并发症的情况；排除了入院时年龄小于 18 岁或大于 89 岁的队列。为了减少数据的稀疏性，我们删除了特征缺失超过 70% 的记录。经过上述筛选后，最终纳入 39721 例 ICU 住院记录作为研究对象。为了进一步研究心力衰竭患者中急性肾衰竭的发生规律，我们根据是否患有心力衰竭，将住院记录进一步分为两个队列：心衰队列（n=12630）和非心衰队列（n=27091）。在心衰队列中，9563 例住院记录涉及 AKI 发生，3067 例未发生 AKI；在非心衰队列中，16343 例住院记录涉及 AKI 发生，10748 例未发生。AKI的定义根据 KDIGO（改善全球肾脏病预后组织，Kidney Disease: Improving Global Outcomes）指南确定，患者符合以下三种情况之一即可确诊为AKI：

1. 在 48 小时内，若血清肌酐（serum creatinine，Scr）升高达到≥26.5 μmol/L或0.3 mg/dl；
2. 患者7天内的Scr 升高超过基础值的 1.5 倍及以上；
3. 尿量＜0.5ml/（kg・h），或者持续时间大于等于6小时。

我们应用同样的筛选标准处理MIMIC-IV数据库，最终获得了51,462例心衰患者队列. 完整的数据筛选流程如下图所示：

|  |
| --- |
| 图示  描述已自动生成 |
|  |

按照上述纳排过程以及队列划分，我们将MIMIC-III数据集分成心衰患者队列与其他患者队列，在MIMIC-IV数据集中获取心衰患者队列，分别命名为MIMICIII-HF、MIMICIII-NO-HF、MIMICIV-HF数据集，我们的模型将在三个数据集上进行测试。由于心衰患者发生 AKI 的风险较高且症状易混淆，对这一特定群体单独研究，有助于更深入了解心衰与 AKI 的关联，明确心衰患者发生 AKI 的独特模式和影响因素。同时，利用非心衰队列可作为对照，通过在非心衰队列上进行模型预测，能对比分析出不同疾病状态下（心衰与否）AKI 预测的差异。由于MIMICIII与MIMICIV患者数据不重叠，所以能涵盖更广泛的心衰和非心衰患者情况，使模型训练数据更具多样性。这样训练出的模型能够更好地适应真实世界中复杂多变的患者数据，提高模型在不同场景下对 AKI 预测的可靠性 。

### 4.2.2数据统计分析

为了了解三个数据集的特征分布情况，分别计算了这三个数据集的连续特征项的统计特征，包括平均值、标准差、下四分位数和上四分位数，因为数据集之间特征存在差异，我们选择共同特征进行统计进行横向比较。四分位数是将数据按大小排序后平均划分为四份的分割点值，下四分位数（Q1）代表排序后位于25%位置的数值，上四分位数（Q3）代表位于75%位置的数值，这两个指标能够有效反映数据的分布范围和集中趋势，不受极端值影响。下面三个表分别展示了MIMICIII-HF、MIMICIII-NO-HF和MIMICIV-HF数据集中连续特征项的基线统计信息：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表 MIMICIII-HF连续特征项基线信息统计 | | | | |
| 特征项 | 平均值 | 标准差 | 下四分位数 | 上四分位数 |
| Anion Gap | 13.937 | 2.905 | 12.000 | 16.000 |
| Bicarbonate | 25.892 | 2.812 | 24.000 | 28.000 |
| Chloride | 103.303 | 4.719 | 100.000 | 106.000 |
| Creatinine | 1.160 | 1.689 | 0.800 | 1.200 |
| Heart Rate | 85.231 | 17.599 | 72.000 | 96.000 |
| MCH | 29.839 | 1.333 | 28.900 | 30.900 |
| MCHC | 33.275 | 1.079 | 32.500 | 34.200 |
| MCV | 89.774 | 4.317 | 86.000 | 93.000 |
| Magnesium | 2.055 | 1.239 | 1.800 | 2.200 |
| PTT | 34.268 | 16.818 | 26.700 | 34.100 |
| Phosphate | 3.536 | 0.515 | 3.100 | 4.000 |
| Platelet Count | 240.892 | 83.238 | 181.000 | 290.000 |
| Potassium | 4.162 | 1.132 | 3.800 | 4.500 |
| RDW | 14.263 | 0.845 | 13.600 | 15.000 |
| Respiratory Rate | 23.116 | 1962.183 | 16.000 | 23.000 |
| Sodium | 138.908 | 3.681 | 136.000 | 141.000 |
| Urea Nitrogen | 14.985 | 3.887 | 12.000 | 18.000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表 MIMICIII-NO-HF连续特征项基线信息统计 | | | | |
| 特征项 | 平均值 | 标准差 | 下四分位数 | 上四分位数 |
| Anion Gap | 13.334 | 2.819 | 11.000 | 15.000 |
| Bicarbonate | 25.353 | 2.531 | 23.000 | 27.000 |
| Chloride | 104.493 | 4.390 | 102.000 | 107.000 |
| Creatinine | 0.915 | 1.654 | 0.600 | 1.000 |
| Heart Rate | 85.982 | 18.736 | 73.000 | 97.000 |
| MCH | 30.052 | 1.295 | 29.200 | 31.100 |
| MCHC | 33.601 | 1.015 | 32.900 | 34.400 |
| MCV | 89.370 | 4.187 | 86.000 | 92.000 |
| Magnesium | 1.991 | 0.641 | 1.800 | 2.200 |
| PTT | 31.705 | 13.555 | 25.600 | 32.800 |
| Phosphate | 3.437 | 0.503 | 3.000 | 3.800 |
| Platelet Count | 237.830 | 92.177 | 176.000 | 285.000 |
| Potassium | 4.065 | 0.543 | 3.700 | 4.400 |
| RDW | 13.850 | 0.903 | 13.200 | 14.600 |
| Respiratory Rate | 18.816 | 5.665 | 15.000 | 22.000 |
| Sodium | 138.872 | 3.570 | 137.000 | 141.000 |
| Urea Nitrogen | 13.293 | 4.055 | 10.000 | 17.000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表 MIMICIV-HF连续特征项基线信息统计 | | | | |
| 特征项 | 平均值 | 标准差 | 下四分位数 | 上四分位数 |
| Anion Gap | 13.641 | 2.735 | 12.000 | 16.000 |
| Bicarbonate | 25.396 | 2.680 | 23.000 | 27.000 |
| Chloride | 102.741 | 3.466 | 100.000 | 106.000 |
| Creatinine | 0.809 | 0.213 | 0.600 | 1.000 |
| Heart Rate | 85.885 | 115.471 | 72.000 | 97.000 |
| MCH | 29.790 | 1.436 | 28.800 | 30.900 |
| MCHC | 33.260 | 1.028 | 32.500 | 34.000 |
| MCV | 90.489 | 4.285 | 87.000 | 94.000 |
| Magnesium | 22.018 | 4472.393 | 1.800 | 2.200 |
| PTT | 112.848 | 8959.209 | 27.100 | 32.100 |
| Phosphate | 3.482 | 0.510 | 3.100 | 3.900 |
| Platelet Count | 386.097 | 12271.788 | 178.000 | 273.000 |
| Potassium | 4.116 | 0.443 | 3.800 | 4.400 |
| RDW | 13.863 | 0.970 | 13.100 | 14.600 |
| Respiratory Rate | 19.929 | 884.070 | 16.000 | 22.000 |
| Sodium | 138.894 | 3.147 | 137.000 | 141.000 |
| Urea Nitrogen | 13.761 | 4.082 | 11.000 | 17.000 |

通过对三个数据集共同连续特征项的基线信息分析，我们发现三个数据集共同的连续特征项的基线信息存在差异。比如对于肌酐值（Creatinine, mg/dL），MIMICIII-HF群体的平均肌酐值为1.160，明显高于MIMICIII-NO-HF群体（0.915）和MIMICIV-HF群体（0.809），差异超过0.2 mg/dL。氯化物（Chloride, mmol/L）水平也有明显差异，MIMICIII-NO-HF群体的平均值为104.493 mmol/L，高于MIMICIII-HF群体（103.303 mmol/L）和MIMICIV-HF群体（102.741 mmol/L）。这种差异可能与不同患者群体的基础疾病、治疗方案有关。

进一步对三个数据集的静态特征进行统计分析。图4.至图4.分别展示了MIMIC数据库中年龄、性别以及住院时长在Non-AKI组和AKI组的分布情况。在三个数据集中，发生AKI的男性患者人数都明显多于女性，这种性别差异可能与男女之间的激素水平、肾脏解剖结构的不同有关；年龄分布分析显示，相比于Non \_AKI组，AKI组年龄偏向于高年龄段，平均年龄达到70岁，显著高于 Non \_AKI 组的平均年龄60岁。随着年龄增长，肾脏会发生结构和功能的生理性衰退，导致肾脏储备能力下降，使得老年人群对各种肾损伤因素（如药物毒性、感染等）的敏感性显著增加。住院时长分析结果表明，AKI组平均住院时间明显长于Non \_AKI 组，这表明AKI患者的病情较为复杂，AKI也常与其他器官功能障碍相互影响，形成恶性循环，需要更长时间的治疗和更为精细的护理干预。

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |

为了进一步验证三个数据集的共同特征是否存在显著差异，我们采用了曼-惠特尼U检验(Mann-Whitney U Test)这一非参数统计方法进行严格验证。该检验方法适用于比较两个独立样本的分布差异，不要求数据服从正态分布。我我们提出以下统计假设：

零假设（H₀）：三个数据集的特征分布相同，即不存在统计学意义上的差异；

备择假设（H₁）：至少有一个数据集的特征分布与其他数据集存在显著差异。

我们对MIMICIII-HF、MIMICIII-NO-HF和MIMICIV-HF三个数据集中的所有共同特征进行了两两比较，检验结果汇总如表4.3所示：

|  |  |
| --- | --- |
| 表 数据集间特征分布的曼-惠特尼U检验结果 | |
| **特征项** | P值 |
| Anion Gap | <0.001 |
| Bicarbonate | <0.001 |
| Chloride | <0.001 |
| Creatinine | <0.001 |
| Heart Rate | <0.001 |
| MCH | <0.001 |
| MCHC | <0.001 |
| MCV | <0.001 |
| Magnesium | <0.001 |
| PTT | <0.001 |
| Phosphate | <0.001 |
| Platelet Count | <0.001 |
| Potassium | <0.001 |
| RDW | <0.001 |
| Respiratory Rate | <0.001 |
| Sodium | 0.087 |
| Urea Nitrogen | <0.001 |
| age | <0.001 |
| gender | <0.001 |
| los | <0.001 |
|  |  |

检验结果显示，除钠离子(Sodium)水平(P=0.087)外，其余19项特征的P值均小于0.001，说明MIMICIII-HF、MIMICIII-NO-HF、MIMICIV-HF三份数据集包含的绝大多数特征项的分布存在显著性差异。阴离子间隙（Anion Gap）、碳酸氢盐（Bicarbonate）、氯化物（Chloride）等电解质指标，肌酐（Creatinine）等肾功能标志物，在三个数据集之间均呈现出显著差异。

进一步对数据集中AKI的发作时间以及数据集的缺失率情况进行分析。图 是AKI发生时间的统计情况，横轴表示进入ICU的时长，纵轴表示发生AKI 的累计人数比率。可以发现进入ICU后24小时是AKI发生的高峰期，大量患者在进入ICU后迅速发生AKI，72小时内（图中黄色虚线标注）几乎包括了所有 AKI 发生患者。其中48小时内，三个数据集累计发生AKI 的比率达分别达到85.13%，84.93%和91.81%. 对于临床工作而言，医护人员需要特别关注 ICU 患者在入住后的早期阶段，尤其是 48 小时内的肾脏功能变化，加强对血清肌酐、尿量等肾功能指标的监测频率，以便能够及时发现 AKI 的发生并尽早采取干预措施，降低 AKI 对患者预后的不良影响。

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 图 AKI发生时间分布 | | |

图 展示了数据集中特征的缺失情况。数据缺失率随时间线性下降，患者刚入ICU 时数据记录较少，随着时间推进，每个患者的记录更加完整，因此更多生理指标和实验数据被采集，缺失率下降。而MIMICIII-HF、MIMICIII-NO-HF两个数据集的缺失情况十分相似，MIMICIV-HF数据中特征的缺失率随着时间的发展保持在40%以上，体现出更大的不规则性。

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 图 数据的缺失率 | | |

### 4.2.3 时间窗口选择

急性肾损伤(AKI)作为一种发展迅速的临床症状，其病理变化往往在短时间内急剧恶化，这使得单一时间点的预测模型难以全面捕捉疾病的动态演变过程。基于这一临床特性，我们设计了多时间点预测框架，以提高模型对AKI早期预警的准确性和及时性。根据AKI的病理生理特征及临床进展规律，结合前文数据分析中发现的AKI发生时间分布特点，我们精心选择了五个关键时间点进行预测：提前24小时、12小时、6小时、3小时以及发病时刻(0小时)。这种梯度式的时间设计使我们能够全面评估模型在不同预警提前量下的性能表现，为临床干预提供时间梯度参考。如图所示，我们采用颜色深浅变化直观地表示疾病随时间推移的进展过程，颜色由浅至深形象地反映了AKI病情的持续恶化趋势。

考虑到AKI的突发性特点以及前文分析中发现的数据在不同时间段的缺失模式，为了充分验证模型对不规则医疗时间序列(ISMTS)的适应能力，我们在所有五个预测时间点均采用了24小时的数据采集窗口。这一窗口大小的选择基于两方面考量：首先，它能够包含足够的时序信息以捕捉AKI的早期征兆；其次，这一时间跨度与AKI的病理生理发展机制高度契合，能够覆盖肾功能从初始损伤到临床表现的关键转变期。通过这种设计，我们既保证了模型获取充分的时序特征，又避免了引入过多无关信息导致的噪声干扰。

|  |
| --- |
|  |
| 图 时态特征数据时间窗 |

## 4.3 实验结果与分析

在本实验中，我们设计了一系列对比试验来评估TransLSTM模型在急性肾损伤(AKI)预测任务中的表现，以验证深度学习模型在时间序列建模上的优势，尤其是结合 LSTM 和 Transformer 机制后的改进效果。为确保实验结果的可靠性与可比性，我们在相同数据集上系统性地设置了以下对比实验框架：

1. 神经网络架构消融实验，旨在验证LSTM与Transformer结合的有效性，我们设置三个对比模型：① LSTM基线模型：通过门控机制捕捉局部时间依赖；② 标准Transformer模型：保留编码器层结构和多头自注意力机制，建立全域时间节点间关联；③ TransLSTM采用级联架构——首先通过Transformer编码层计算跨时间步的动态注意力权重，然后LSTM层提取细粒度时序特征矩阵。
2. 与传统机器学习对比试验，我们选取三种经典模型：① 逻辑回归：作为线性模型的代表，提供基准性能参考，评估简单线性关系的建模能力；② 随机森林：通过集成学习和特征重要性排序执行非线性建模；③ 多层感知机：作为深度学习与传统方法的桥梁，在移除时序信息的情况下，通过多层全连接网络拟合特征间的复杂交互关系。

为确保评估的公平性和结果的可靠性，我们的研究采用统一的方法划分训练集、验证集和测试集，重点比较不同方法预测能力上的差异。

### 4.3.1 神经网络架构消融实验结果

通过设计三个模神经网络架构的对比试验来测试LSTM、Transformer和TransLSTM三种模型的表现，实验采用 AUROC、Sensitivity、Specificity和F1-score作为评价指标。如**表** 所示，在MIMICIII-HF数据集中的所有预测时间点，TransLSTM的AUROC、Sensitivity和F1-score均优于其他模型，充分证明了我们的模型在AKI预测任务中的优越性。在 24h 提前预警的任务中，TransLSTM的AUROC达到 0.964，Sensitivity达到0.967，均优于基线模型，这一结果表明TransLSTM可以更早地检测到AKI发生的风险。Transformer 在3h 预警任务中，Specificity为 0.996，而 Sensitivity 仅 0.438，说明 Transformer 容易错过大量正例，可能导致漏诊。相比之下，TransLSTM 不像 Transformer 那样过度关注 Specificity，导致 Sensitivity 过低。TransLSTM在 Sensitivity 和 Specificity 之间取得了更好的平衡，避免了 Transformer的缺点，同时提升了识别能力。TransLSTM 结合了 Transformer 的长距离依赖建模能力与 LSTM 的时序信息处理能力，因此能够提供更稳定的预测效果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表 模型的消融实验-MIMICIII-HF | | | | | |
|  | **Model** | **AUROC** | **Sensitivity** | **Specificity** | **F1-Score** |
|  | LSTM | 0.961 | 0.960 | 0.816 | 0.949 |
| 24 hour early | Transformer | 0.963 | 0.906 | 0.893 | 0.933 |
|  | TransLSTM | **0.964** | **0.967** | 0.824 | **0.954** |
|  | LSTM | 0.930 | 0.816 | 0.897 | 0.881 |
| 12 hour early | Transformer | 0.937 | 0.812 | 0.917 | 0.883 |
|  | TransLSTM | **0.940** | **0.953** | 0.640 | **0.918** |
|  | LSTM | 0.908 | 0.742 | 0.948 | 0.843 |
| 6 hour early | Transformer | 0.922 | 0.793 | 0.930 | 0.873 |
|  | TransLSTM | **0.938** | **0.933** | 0.702 | **0.917** |
|  | LSTM | 0.901 | 0.685 | 0.961 | 0.807 |
| 3 hour early | Transformer | 0.859 | 0.438 | 0.996 | 0.609 |
|  | TransLSTM | **0.936** | **0.899** | 0.754 | **0.906** |
|  | LSTM | 0.860 | 0.836 | 0.733 | 0.868 |
| 0 hour early | Transformer | 0.885 | 0.743 | 0.862 | 0.830 |
|  | TransLSTM | **0.949** | **0.931** | 0.758 | **0.925** |

在MIMICIII-NO-HF数据集上，TransLSTM同样展现出优异的性能，尤其在AUROC指标上，在所有五个预测时间点均取得最高值。值得注意的是，在24小时预警任务中，TransLSTM与LSTM在Sensitivity上并列达到0.925的高水平，但TransLSTM的AUROC和F1-score仍然较高，分别为0.961和0.924。这表明TransLSTM在不同患者群体中都能保持稳定的预测性能。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表 模型的消融实验-数据集MIMICIII-NO-HF | | | | | |
|  | **Model** | **AUROC** | **Sensitivity** | **Specificity** | **F1-Score** |
|  | LSTM | 0.958 | **0.925** | 0.876 | 0.921 |
| 24 hour early | Transformer | 0.942 | 0.911 | 0.826 | 0.898 |
|  | TransLSTM | **0.961** | **0.925** | 0.885 | **0.924** |
|  | LSTM | 0.919 | 0.830 | 0.846 | 0.859 |
| 12 hour early | Transformer | 0.890 | 0.812 | 0.816 | 0.832 |
|  | TransLSTM | **0.925** | 0.819 | 0.865 | 0.858 |
|  | LSTM | 0.906 | 0.833 | 0.821 | 0.849 |
| 6 hour early | Transformer | 0.858 | 0.822 | 0.730 | 0.821 |
|  | TransLSTM | **0.912** | 0.799 | 0.854 | **0.842** |
|  | LSTM | **0.909** | 0.823 | 0.854 | 0.857 |
| 3 hour early | Transformer | 0.860 | 0.770 | 0.787 | 0.805 |
|  | TransLSTM | **0.904** | 0.805 | 0.867 | 0.850 |
|  | LSTM | 0.904 | 0.803 | 0.855 | 0.845 |
| 0 hour early | Transformer | 0.876 | 0.771 | 0.828 | 0.817 |
|  | TransLSTM | **0.907** | 0.840 | 0.825 | 0.859 |

在MIMICIV-HF数据集上，TransLSTM在大多数时间点的四项评价指标上表现优秀。也证明该模型架构在不同数据采集环境下的适应性和泛化能力。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表 模型的消融实验-数据集MIMICIV-HF | | | | | |
|  | **Model** | **AUROC** | **Sensitivity** | **Specificity** | **F1-Score** |
|  | LSTM | 0.966 | 0.883 | 0.936 | 0.923 |
| 24 hour early | Transformer | 0.950 | 0.858 | 0.916 | 0.906 |
|  | TransLSTM | **0.968** | **0.912** | **0.915** | **0.937** |
|  | LSTM | 0.948 | 0.855 | 0.914 | 0.904 |
| 12 hour early | Transformer | 0.915 | 0.832 | 0.862 | 0.880 |
|  | TransLSTM | **0.948** | **0.869** | **0.902** | **0.909** |
|  | LSTM | **0.936** | 0.869 | 0.856 | 0.900 |
| 6 hour early | Transformer | 0.898 | 0.814 | 0.828 | 0.863 |
|  | TransLSTM | **0.935** | **0.829** | **0.904** | **0.887** |
|  | LSTM | **0.937** | **0.849** | **0.880** | **0.893** |
| 3 hour early | Transformer | 0.901 | 0.823 | 0.822 | 0.867 |
|  | TransLSTM | **0.937** | **0.871** | **0.850** | **0.900** |
|  | LSTM | 0.941 | 0.865 | 0.862 | 0.899 |
| 0 hour early | Transformer | 0.895 | 0.755 | 0.877 | 0.835 |
|  | TransLSTM | **0.945** | **0.862** | **0.889** | **0.903** |

### 4.3.2 与传统机器学习对比试验结果

我们将TransLSTM与几种传统的代表性机器学习模型（如随机森林、MLP、线性模型）进行了对比，为了保留数据的完整性和对比实验的公平性，我们将每个时间点的相同输入变量的值表示为不同的变量来“扁平化”（flattening）数据，即通过将不同时间点的相同输入变量展开为多个特征列（如）,保留每个时间点的变量信息，从而使传统模型能够处理患者的时间序列数据。

在MIMICIII-HF数据集上，TransLSTM在所有预测时间点（24小时至0小时）的AUROC指标均达到最高值，尤其在24小时预警任务中，其AUROC高达0.964，显著优于MLP（0.857）和线性回归（0.883），略高于随机森林（0.963）。在敏感度方面，TransLSTM在24小时、12小时和0小时预警任务中分别达到0.967、0.953和0.931。同时我们也发现，随机森林在特异度指标上多个时间点表现突出，但TransLSTM在F1-Score这一综合指标上较高，显示了其在平衡敏感度和特异度方面的优越性。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 表 不同模型的性能对比MIMICIII-HF | | | | |
|  | Model | AUROC | Sensitivity | Specificity | F1-Score |
| 24 hour early | MLP | 0.939 | 0.921 | 0.810 | 0.927 |
| RF | 0.963 | 0.878 | 0.933 | 0.924 |
| LR | 0.883 | 0.876 | 0.697 | 0.884 |
| TransLSTM | **0.964** | **0.967** | 0.824 | **0.954** |
| 12 hour early | MLP | 0.891 | 0.830 | 0.817 | 0.876 |
| RF | 0.932 | 0.813 | 0.924 | 0.884 |
| LR | 0.812 | 0.948 | 0.223 | 0.854 |
| TransLSTM | **0.940** | **0.953** | 0.640 | **0.918** |
| 6 hour early | MLP | 0.882 | 0.817 | 0.798 | 0.866 |
| RF | 0.925 | 0.813 | 0.899 | 0.880 |
| LR | 0.767 | 0.943 | 0.205 | 0.849 |
| TransLSTM | **0.938** | 0.933 | 0.702 | **0.917** |
| 3 hour early | MLP | 0.887 | 0.802 | 0.801 | 0.857 |
| RF | 0.920 | 0.813 | 0.899 | 0.879 |
| LR | 0.836 | 0.963 | 0.275 | 0.869 |
| TransLSTM | **0.936** | 0.899 | 0.754 | **0.906** |
| 0 hour early | MLP | 0.904 | 0.813 | 0.841 | 0.870 |
| RF | 0.916 | 0.797 | 0.869 | 0.865 |
| LR | 0.859 | 0.953 | 0.333 | 0.872 |
| TransLSTM | **0.949** | 0.931 | 0.758 | **0.925** |

在MIMICIII-NO-HF数据集上，TransLSTM同样展现出了稳定的性能优势。在AUROC指标上，TransLSTM在24小时、12小时和6小时预警任务中均取得最高值，分别为0.961、0.925和0.912。特别是在24小时预警任务中，TransLSTM的F1-Score达到0.924，明显高于其他模型。值得关注的是，线性回归模型在该数据集上表现出明显的局限性，尤其在特异度方面，多个时间点接近于0，表明其无法有效区分阴性样本。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 表 不同模型的性能对比-MIMICIII-NO-HF | | | | |
|  | Model | AUROC | Sensitivity | Specificity | F1-Score |
| 24 hour early | MLP | 0.944 | 0.904 | 0.862 | 0.906 |
| RF | 0.947 | 0.856 | 0.937 | 0.902 |
| LR | 0.832 | 0.998 | 0.008 | 0.750 |
| TransLSTM | **0.961** | 0.925 | 0.885 | **0.924** |
| 12 hour early | MLP | 0.904 | 0.888 | 0.728 | 0.858 |
| RF | 0.920 | 0.796 | 0.910 | 0.858 |
| LR | 0.817 | 0.269 | 0.982 | 0.419 |
| TransLSTM | **0.925** | 0.819 | 0.865 | 0.858 |
| 6 hour early | MLP | 0.885 | 0.754 | 0.864 | 0.817 |
| RF | 0.907 | 0.776 | 0.886 | 0.838 |
| LR | 0.825 | 0.997 | 0.029 | 0.754 |
| TransLSTM | **0.912** | 0.799 | 0.854 | **0.842** |
| 3 hour early | MLP | 0.899 | 0.847 | 0.779 | 0.849 |
| RF | **0.908** | 0.795 | 0.871 | 0.845 |
| LR | 0.837 | 1.00 | 0.00 | 0.749 |
| TransLSTM | **0.904** | 0.805 | **0.867** | **0.850** |
| 0 hour early | MLP | 0.896 | 0.803 | 0.822 | 0.835 |
| RF | **0.911** | 0.799 | 0.858 | 0.844 |
| LR | 0.856 | 0.974 | 0.229 | 0.782 |
| TransLSTM | **0.908** | 0.840 | 0.825 | **0.859** |

在MIMICIV-HF数据集上，TransLSTM的优势更加突出。在五个预测时间点上，TransLSTM在AUROC、敏感度、特异度和F1-Score四项指标上几乎全部领先。特别是在24小时预警任务中，TransLSTM的AUROC达到0.968，敏感度和特异度分别为0.912和0.915，F1-Score高达0.937，超越其他对比模型。这一结果充分证明了TransLSTM在不同数据集上的泛化能力和稳定性。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 表 不同模型的性能对比-MIMCIV-HF | | | | |
|  | Model | AUROC | Sensitivity | Specificity | F1-Score |
| 24 hour early | MLP | 0.957 | 0.887 | 0.912 | 0.922 |
| RF | 0.935 | 0.840 | 0.887 | 0.890 |
| LR | 0.804 | 0.806 | 0.627 | 0.820 |
| TransLSTM | **0.968** | **0.912** | **0.915** | **0.937** |
| 12 hour early | MLP | 0.921 | 0.815 | 0.876 | 0.873 |
| RF | 0.928 | 0.840 | 0.851 | 0.883 |
| LR | 0.819 | 0.659 | 0.813 | 0.758 |
| TransLSTM | **0.948** | **0.869** | **0.902** | **0.909** |
| 6 hour early | MLP | 0.912 | 0.878 | 0.777 | 0.890 |
| RF | 0.927 | 0.831 | 0.854 | 0.878 |
| LR | 0.824 | 0.612 | 0.861 | 0.733 |
| TransLSTM | **0.935** | **0.829** | **0.904** | **0.887** |
| 3 hour early | MLP | 0.919 | 0.798 | 0.878 | 0.863 |
| RF | 0.928 | 0.842 | 0.847 | 0.883 |
| LR | 0.837 | 0.866 | 0.610 | 0.853 |
| TransLSTM | **0.937** | **0.871** | **0.850** | **0.900** |
| 0 hour early | MLP | 0.928 | 0.814 | 0.884 | 0.874 |
| RF | 0.934 | 0.850 | 0.862 | 0.891 |
| LR | 0.849 | 0.714 | 0.837 | 0.801 |
| TransLSTM | **0.945** | **0.862** | **0.889** | **0.903** |

## 4.4 外部验证

MIMIC-III数据的收集时间是2001年6月至2012年10月，而MIMIC-IV数据集则是2008年至2019年，且它们所包含的患者群体是不同的。我们使用MIMICIV-HF数据集中2013 年至 2019 年的数据作为时间验证集，使用MIMICIII-NO-HF作为外部验证集。为了方便验证，我们提取三个数据集共同的特征。具体试验方案如下：

1. 在0h的预测事件点，以MIMICIII-HF数据集作为源数据集进行学习，并以MIMICIV-HF数据集为目标数据集进行时间验证。
2. 在0h的预测事件点，以MIMICIV-HF数据集作为源数据集进行学习，并以MIMICIII-NO-HF数据集为目标数据集进行外部验证。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MIMICIII-HFMIMICIV-HF | | MIMICIV-HFMIMICIII-NO-HF | |
|  | Sorce\_AUC  (MIMICIII-HF) | Target\_AUC  (MIMICIV-HF) | Sorce\_AUC  (MIMICI**V**-HF) | Target\_AUC  (MIMICIII-NO-HF) |
| MLP | 0.928 | 0.857 | 0.947 | 0.914 |
| RF | **0.946** | 0.817 | 0.954 | 0.916 |
| LR | 0.696 | 0.614 | 0.731 | 0.650 |
| LSTM | 0.921 | 0.848 | 0.965 | **0.935** |
| Transformer | 0.911 | **0.865** | 0.959 | 0.924 |
| TransLSTM | **0.942** | **0.893** | **0.973** | **0.942** |

## 4.5 决策支持可视化

|  |
| --- |
| 电脑萤幕画面  描述已自动生成 |

在患者个体层面上，我们通过注意力机制对模型的关注程度进行了可视化分析。具体而言，模型在计算当前输出时会对输入序列中各位置的特征信息进行加权关注，这种关注程度可以通过注意力权重（即注意力激活值）来量化，并生成反映患者个体特征的注意力图谱。为了更好地理解模型关注的特征及其临床意义，我们从0-hour预测任务中选取了两类具有代表性的患者进行可视化分析（如图所示）。其中，ICUid A代表本次入院中未发生急性肾损伤（AKI）的患者，而ICUid B则是在本次住院过程中出现AKI的患者。每个记录从上至下依次展示了以下内容：条形图报告每个特征在时间维度上的平均贡献；注意力激活值热图；输入特征实际值的热图。

从图中可以发现，对于患者A，肌酐值（creatinine）在某个时间段的注意力激活值显著升高（接近白色），但在其他时间段肌酐值保持在正常水平。而对于患者B，模型最为关注的特征是uo\_rt\_6hr(6小时内尿量,单位ml/kg/h), 根据图中显示，在靠近AKI发生的一段时间内，患者B的uo\_rt\_6hr一直处于较低水平。这种差异化的关注模式反映了两类患者的临床特征变化趋势。对于患者A而言，尽管其肌酐值在某段时间内出现上升趋势，但随后得到有效控制，因此未发展为AKI；而患者B的情况则截然不同。虽然患者B的uo\_rt\_6hr值在某些时间点有所改善，但整体未能维持在一个稳定水平，最终导致AKI的发生。

uo\_rt\_6hr和肌酐值都是国际共识定义中被广泛认可的AKI发生判断标准的关键特征。其中，uo\_rt\_6hr的持续性异常通常被认为是一个更为敏感的指标，能够更早地反映肾脏功能的损伤。通过对这些特征的关注和分析，不仅能帮助临床医生更好地理解模型预测结果背后的风险因素，还能为精准的临床干预决策提供重要参考。这种基于注意力机制的特征可视化分析方法，不仅揭示了模型对患者个体特征的关注程度，也进一步验证了特征在AKI预测中的临床意义。这种研究思路为探索深度学习模型的可解释性提供了新的视角，同时也为临床实践中的个体化风险管理提供了更有价值的参考信息。

|  |
| --- |
|  |

在群体层面上，为了全面了解所有样本中各特征对模型输出的一般性影响，我们采用了SHAP（SHapley Additive exPlanations）方法来估计每个特征的全局重要性。SHAP是一种广泛应用于机器学习模型解释的前沿方法，它能够量化每个特征对模型输出的贡献程度，从而揭示模型决策背后的关键因素。

图中展示了0-hour组前20个特征的全局平均贡献值，结果显示，胆红素（Bilirubin）对模型输出的影响最大，其次是白蛋白（Albumin）、氯化物（Chloride）、钠离子浓度（Sodium）和葡萄糖（Glucose）。这些特征的排序反映了它们在AKI预测中的重要性。SHAP摘要图展示了每个特征对模型输出影响的整个分布。颜色表示特征值的变化如何影响模型的输出：红色代表高特征值与模型预测输出正相关，蓝色则表示低特征值与模型预测输出负相关。点距基线SHAP值（0）的距离越远，表示该特征对模型输出的影响越显著。具体来看，胆红素（Bilirubin）、白蛋白（Albumin）、氯化物（Chloride）和钠离子浓度（Sodium）的SHAP值分布总体上集中在中心。进一步分析影响方向，高胆红素（Bilirubin）显著增加了AKI的发生风险，而高氯化物（Chloride）和高白蛋白（Albumin）则与AKI的发生风险呈负相关。

这些发现表明，胆红素（Bilirubin）可能是预测AKI的重要生物标志物，其升高与AKI的发生具有一定的相关性。同时，氯化物和白蛋白的水平对模型输出也具有显著影响，但它们的高值反而与较低的AKI风险相关。这反映了高白蛋白水平和高氯化物水平在一定程度上对肾脏功能的保护作用或机体的代偿机制。通过全局性分析，我们不仅能够识别出对AKI预测具有重要影响的关键特征，还能够深入理解这些特征的生物学意义及其对模型预测结果的贡献作用。

## 4.6 本章小结

在本章，我们

# 总结与展望

# 参考文献

1. Sun C, Hong S, Song M, et al. A review of deep learning methods for irregularly sampled medical time series data[J]. arXiv preprint arXiv:2010.12493, 2020.
2. Johnson A E W, Pollard T J, Shen L, et al. MIMIC-III, a freely accessible critical care database[J]. Scientific data, 2016, 3(1): 1-9.
3. Che Z, Purushotham S, Cho K, et al. Recurrent neural networks for multivariate time series with missing values[J]. Scientific reports, 2018, 8(1): 6085.

# 致谢