

基于神经网络多项生命特征的脓毒症预测方法

何亚君, 潘睿

上海理工大学机械工程学院, 上海

收稿日期: 2023年3月10日; 录用日期: 2023年5月15日; 发布日期: 2023年5月22日

摘要

针对缺乏ICU诊断条件的受伤患者, 无法提前预测脓毒症感染风险的问题, 提出了一种基于多项生命特征的神经网络(深度神经网络, CNN-LSTM)建立脓毒症预测模型的方法。为了解决数据严重不平衡的问题, 引入了CLASS_WEIGHT技术。该模型使用医院ICU病患数据集进行训练和测试, 与传统医学诊断得分制(例如SIRS、SOFA等)以及早期用于预测脓毒症的机器学习方法进行了比较。在AUROC评分指标上, 该模型取得了较高的得分, 表明该方法是有有效和可行的。

关键词

脓毒症, 神经网络, CNN-LSTM, CLASS_WEIGHT, 机器学习

Sepsis Prediction Method Based on Neural Network Combined with Multiple Life Characteristics

Yajun He, Rui Pan

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 10th, 2023; accepted: May 15th, 2023; published: May 22nd, 2023

Abstract

A method has been proposed to address the problem of being unable to predict sepsis infection risk in injured patients lacking ICU diagnostic conditions. The method involves building a sepsis prediction model based on multiple vital signs using a deep neural network (CNN-LSTM). The CLASS_WEIGHT technique was introduced to address severe data imbalance. The model was trained and tested using ICU patient datasets from hospitals, and was compared with traditional

medical diagnosis scoring systems (such as SIRS and SOFA) and early machine learning methods for predicting sepsis. The model achieved a high score on the AUROC evaluation metric, indicating that the approach is effective and feasible.

Keywords

Sepsis, Deep Neural Network, CNN-LSTM, CLASS_WEIGHT, Machine Learning

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

脓毒症是创伤、烧伤、休克、感染等临床急危重患者的严重并发症之一，也是导致脓毒性休克、多器官功能障碍综合征(MODS)的重要原因之一[1]。现代医学采用 Sepsis-3、SIRS、SOFA 等评分指标对脓毒症进行诊断。此外，机器学习是医学领域中应用最广泛的人工智能技术之一。Fleuren 等[2]对机器学习在脓毒症预测方面的相关研究进行了系统评价和荟萃分析。针对庞大的数据指标，Desautels 等[3]提出了一种名为“Insight 算法”的预测方法，可以使用最少的电子健康记录数据实现对 ICU 人群脓毒症发病情况的预测，并且取得了不错的预测效果。

Seymour CW 等[4]于 2016 年介绍了 Sepsis-3 的新定义，探讨了脓毒症诊断标准和预测的困难性。然而，脓毒症的临床表现尚不清楚，这导致评分体系存在很大的隐患，从而降低了其评测的有效性。此外，脓毒症的早期症状不具备特异性，诊断标准也不一致[5]，这加大了脓毒症的诊断难度。同时，数据的复杂性和多样性也增加了预测和诊断的挑战[6]。机器学习模型的预测准确性受到数据质量、特征选择以及数据平衡性等多方面因素的影响，这也给脓毒症的预测带来了巨大的挑战。

为了解决上述研究中存在的各类问题，本研究采用有选择性的筛选数据指标的方法，最终选取患者的心率、血压、血氧和体温等数据以精简数据量和提升数据质量。结合神经网络模型和引入 CLASS_WEIGHT 技术来解决数据不平衡问题，从而提高模型的预测准确性。此方法简单易行，使患者无需住院面临繁琐的检测程序，仅需使用少量健康数据即可实现预测，并取得了不错的效果。此方法不仅降低了公共经济消耗，同时提高了患者生活质量，缩短了患者的诊断时间。

2. 基础知识

2.1. 脓毒症基础介绍

脓毒症是由感染引起的宿主免疫反应失调而导致危及生命的多器官功能障碍[7] [8]。脓毒性休克是一种严重的脓毒症，其定义为脓毒症合并严重的循环、细胞和代谢紊乱，其死亡风险较单纯脓毒症更高[7]。

脓毒症和脓毒性休克是急危重症医学中面临的重要临床问题。全球每年脓症患者数量超过 1900 万，其中有 600 万患者死亡，病死率超过 1/4。存活的患者中，约有 300 万人存在认知功能障碍[9] [10] [11]。

脓毒症的前兆症状主要表现为毒血症症状，通常包括寒战、高热。严重时，有些患者体温并不升高。此外，还会出现全身不适、软弱无力、头痛、肌肉酸痛、呼吸急促，伴有恶心、呕吐、腹痛、腹泻等症状。病人随之而来的体温升高，呼吸频率紊乱，同时人体血压降低，血氧含量低于正常水平。

脓毒症的治疗方案通常包括使用抗生素治疗感染和进行支持性治疗。治疗过程中，医生需要密切监

测患者的生命体征和器官功能, 并及时发现并处理并发症。预防脓毒症的最佳方法是避免感染, 这可以通过维护良好的个人卫生、接种疫苗以及避免接触已知的感染源来实现。

2.2. CNN-LSTM 网络基础介绍

CNN-LSTM 是由 Shi 等[12]人在 2015 年提出的一种深度学习网络结构, 将卷积神经网络(CNN)和长短期记忆网络(LSTM)相结合。该结构利用 CNN 对输入数据进行特征提取和降维, 再利用 LSTM 对序列数据进行建模和处理, 从而能够更好地对时空信息进行建模, 并具有更好的稳定性。

CNN-LSTM 网络结构包括输入层、卷积层、池化层、LSTM 层和输出层等几个部分。在该网络中, 通常将 CNN 的输出作为 LSTM 的输入, 以充分利用 CNN 的特征提取和降维能力。同时, 为了解决梯度消失和梯度爆炸问题, 可以对 LSTM 中的门控单元和记忆单元进行限制, 从而使得模型更加稳定。本文针对脓毒症的预测模型基于 CNN-LSTM 网络进行训练和优化操作。

3. 数据与方法

本文提出的方案流程如图 1 所示。本方案主要涉及数据预处理与模型建立与评估阶段, 具体包括数据获取与预处理、特征工程建立、模型设计与训练以及模型测试评估。在数据的预处理阶段, 首先通过筛选提出大量不利于模型训练的常量数据, 保留相关生命体征指标, 并采用前后推补值的方法处理数据中的空值数据。在此基础上, 建立特征工程矩阵作为模型的输入。为提高模型性能, 引入 Dropout 和 CLASS_WEIGHT 技术。使用五折交叉验证方法训练得出最佳模型, 并在寻找最佳输出评判阈值后将模型输出结果进行可视化操作, 绘制相关混淆矩阵和 ROC 曲线, 以评估模型效果。

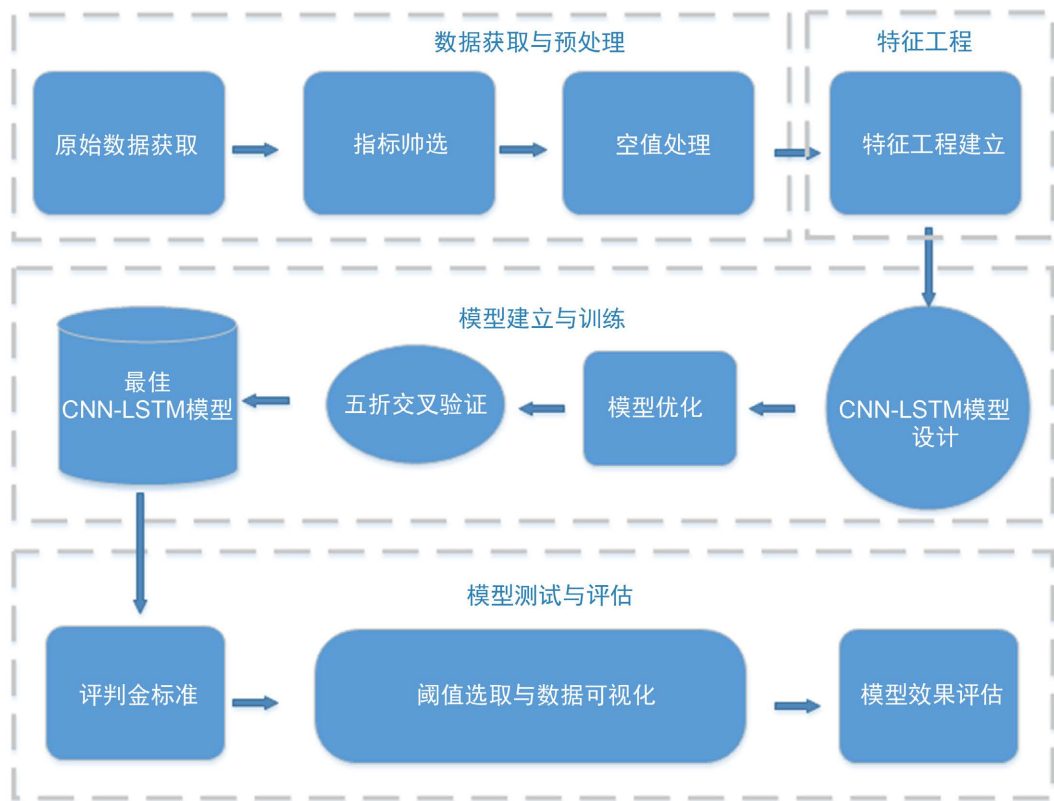


Figure 1. Scheme flowchart
图 1. 方案流程图

3.1. 数据集获取与预处理

3.1.1. 数据集获取

脓毒症预测所使用的数据来自于第 20 届全球生理测量挑战赛(PhysioNet/CinC Challenge 2019)。该数据集包含来自两个独立医院系统的 ICU 患者数据, 共计 2 万名患者。数据集被划分为训练集、测试集, 包含了生命特征和实验室检测结果等多项数据。每个患者的数据以 1 小时为时间间隔进行索引。

3.1.2. 数据集指标筛选

数据指标的筛选依据来自于英国国家早期预警评分(NEWS)评分标准[13], 表 1 为英国国家早期预警评分(NEWS)评分标准。

Table 1. National early warning score

表 1. 英国国家早期预警评分(NEWS)评分标准

生理指标	3 分	2 分	1 分	0 分	1 分	2 分	3 分
呼吸频率/(次/min)	≤8		9~11	12~20		21~24	≥25
血氧饱和度/%	≤91	92~93	94~95	≥96			
吸氧		是		否			
心率/(次/min)	≤40		41~50	51~90	91~110	111~130	≥131
收缩压/mmHg	≤90	91~100	101~110	111~219			≥220
体温/°C	≤35.0		35.1~36.0	36.1~38.0	38.1~39.0	≥39.1	
意识水平				A			V, P, U

剔除了大量的常量实验室数据, 使用易于获取的生命体征数据, 包括患者的心率、血氧饱和度、体温、收缩压、舒张压和平均动脉压。数据集中的数据标签 SepsisLabel 表示患者是否患有脓毒症, 其中 0 代表患者未感染脓毒症, 1 代表患者感染脓毒症。表 2 展示了部分数据集。

Table 2. Data set

表 2. 数据集展示

HR	O2Sat	Temp	SBP	MAP	DBP	SepsisLabel
72	100	37.3	105	70	51	1
101	100	36.33	155	98	65	1
102	97	38	135	81	54	1
91	100	36.11	168	102	74	1
71	98	37.4	150	91.5	68.5	1
90	90	37.6	96.5	75.67	72	1
112	99	38.44	180	118	92	1
58	98	36.11	116	69	44	1
126	96	38	102	77	64	1
90	100	37.72	84	70	66	1

3.1.3. 数据集空值处理

经过指标筛选后脓毒症数据集中涉及大量缺失数据, 以此作为模型输入将导致模型失效, 需要对数据集空值进行进一步处理, 数据的测量时间增量均为一小时, 对于缺失值采用前推补充方法作为替代, 图 2 为数据集空值处理流程。

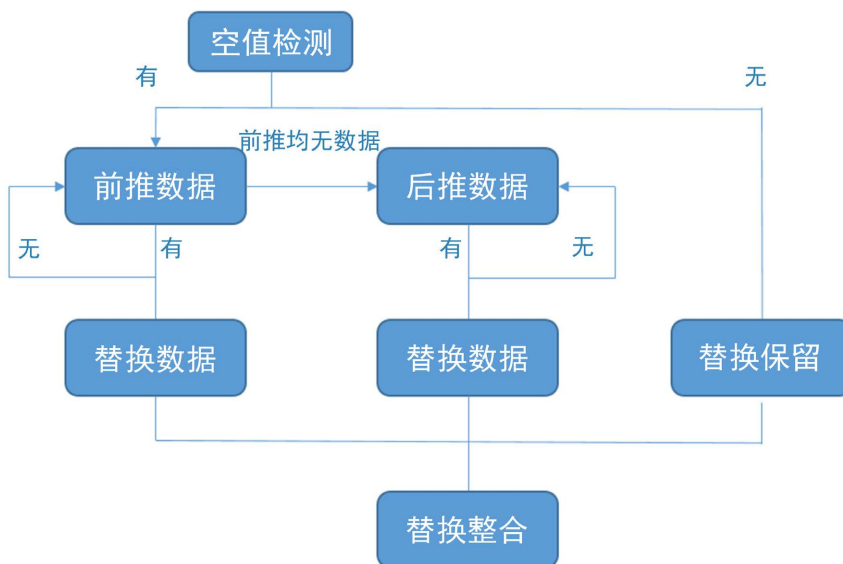


Figure 2. Dataset null processing flowchart
图 2. 数据集空值处理流程图

首先对数据集进行空值检测, 若不存在空值则将此条数据进行保留, 若存在空值则优先进行前推数据检测循环, 直至检测到前置数据为止, 然后将检测到的前置数据替换掉空值数据。若出现前置均无数据情况则启用后推数据检测循环, 直至检测到数据并进行替换操作。因六项生命特征指标不存在均为空值的现象, 所以不存在后推数据检测死循环, 所有空值数据均能得到替换。最终将整体数据进行整合, 表 3 与表 4 为空值替换前后数据展示。

Table 3. Datasets with null values
表 3. 含空值数据集

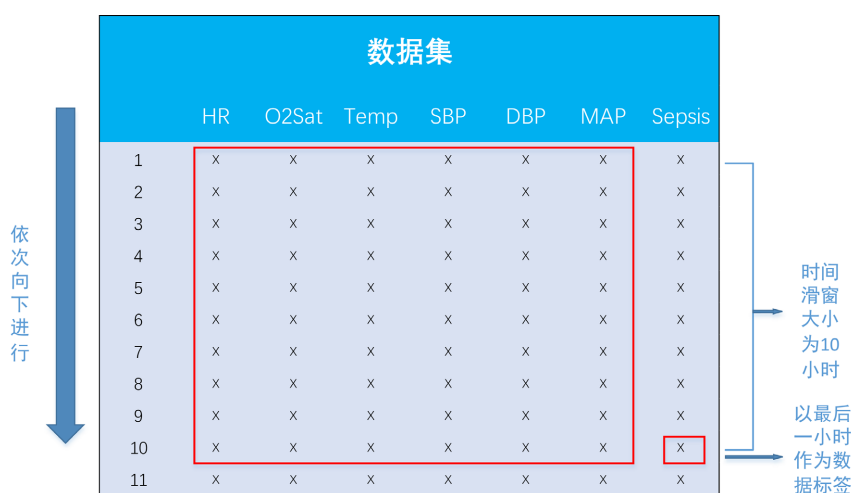
HR	O2Sat	Temp	SBP	MAP	DBP	SepsisLabel
NaN	NaN	NaN	NaN	NaN	NaN	0
74	96	NaN	136	82	NaN	0
75	95	NaN	151	NaN	NaN	0
77	94	36.56	133	85	NaN	0
77	95	NaN	127	76	NaN	0
75	94	NaN	129	87	56	0
71	NaN	36.56	116	74	54	0
85	94	NaN	116.5	67	60	0
71	96	NaN	122	76	71	0

Table 4. Datasets without null values**表 4.** 处理空值后数据集

HR	O2Sat	Temp	SBP	MAP	DBP	SepsisLabel
74	96	36.56	136	82	56	0
74	96	36.56	136	82	56	0
75	95	36.56	151	82	56	0
77	94	36.56	133	85	56	0
77	95	36.56	127	76	56	0
75	94	36.56	129	87	56	0
71	94	36.56	116	74	54	0
85	94	36.56	116.5	67	60	0
71	96	36.56	122	76	71	0

3.2. 特征工程建立

据研究显示, 人体各项生命指标与临床表现的相关性在 6~12 小时左右最为紧密。为了同时考虑模型预测的精确度和临床相关性, 我们选择 10 小时作为预测时间单位, 并基于此对数据集进行特征工程处理, 图 3 展示了特征工程的流程图。

**Figure 3.** Feature engineering flowchart**图 3.** 特征工程流程图

对数据集进行特征工程建立后作为 CNN-LSTM 神经网络的输入进行训练与测试。

3.3. 模型建立与训练

3.3.1. CNN-LSTM 模型设计

CNN-LSTM 网络可以用于解决分类问题, 特别是序列分类问题。本文中 CNN 神经网络部分用于提取序列数据的空间特征, 使用不同大小和数量的卷积核进行卷积运算, 提取不同尺寸的特征。LSTM 用于提取序列数据的时序特征, 通过多个 LSTM 层进行堆叠, 从而提取更深层次的时序信息。在训练

CNN-LSTM 网络时, 使用二值交叉熵损失函数来衡量分类误差, 并使用反向传播算法来更新网络参数。使用 Adam 优化器来加速训练过程, 并添加 Dropout 正则化技术抑制过拟合现象, 引入 CLASS_WEIGHT 参数用以解决数据比例失衡问题, 图 4 为 CNN-LSTM 网络结构示意图。

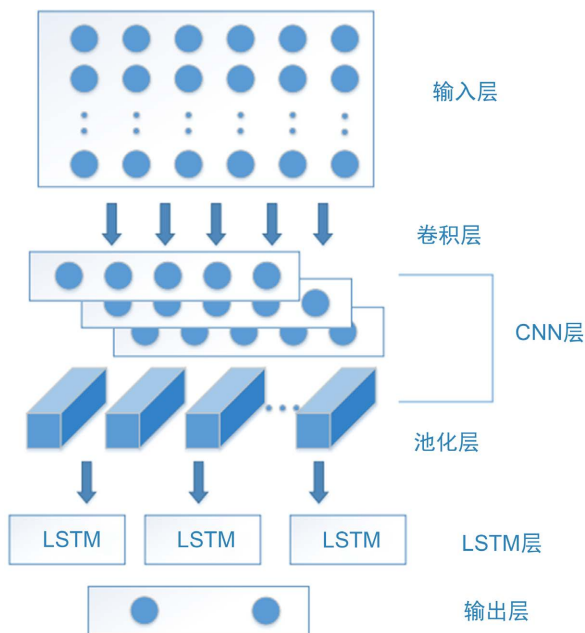


Figure 4. Schematic diagram of CNN-LSTM network structure
图 4. CNN-LSTM 网络结构示意图

3.3.2. 模型优化

1) Dropout 正则化技术

Dropout 是一种常用的正则化技术, 用于减少深度神经网络的过拟合问题。Srivastava 等[14]最早提出 Srivastava, 它的原理是在训练过程中以一定的概率随机丢弃一部分神经元, 从而强制网络去学习更加鲁棒的特征表示。在测试过程中, 所有的神经元都会被保留, 但是需要将每个神经元的输出值乘以训练时的丢失概率, 以保持输出值的期望不变。图 5 为 Dropout 示意图。

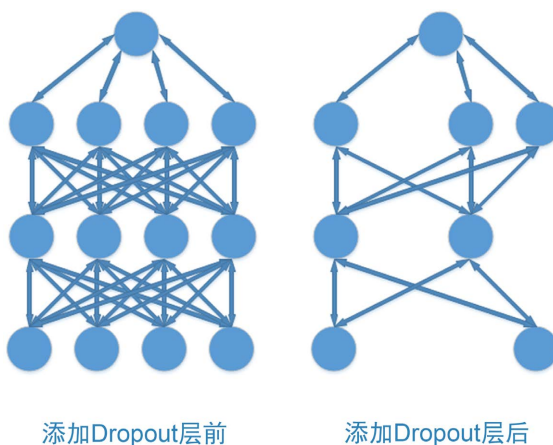


Figure 5. Dropout schematic
图 5. Dropout 示意图

CNN-LSTM 神经网络中的 Dropout 层可以在卷积层、池化层、LSTM 层和全连接层中添加。在 LSTM 层中添加 Dropout 层可以随机地丢弃一部分 LSTM 单元和输出状态, 从而减少过拟合。在全连接层中添加 Dropout 层可以随机地丢弃一部分神经元, 从而提高模型的泛化能力。

2) CLASS_WEIGHT 参数

RushiLongadge 等[15]在 2013 年介绍了类别不平衡问题的现状和挑战, 阐述了 CLASS_WEIGHT 技术能有效解决多数数据不平衡问题, 对于分类神经网络, CLASS_WEIGHT 参数可以设置为一个字典, 其中每个类别对应的权重可以手动指定。在训练过程中, 神经网络会根据这些权重进行优化, 从而更加关注少数类别的样本, 避免过度关注多数类别的样本。

本文通过引入 Dropout 正则化技术抑制了模型训练中的过拟合现象, 并引入 CLASS_WEIGHT 参数用以解决数据集中的比例失衡问题, 在提升网络性能的同时增加了模型的预测精度。

3.3.3. 五折交叉验证

为了更加有效地训练和评估模型, 我们采用了如图 6 所示的五折交叉验证方式。首先, 我们按照 4:1 的比例在经过随机打乱处理后的数据集上划分出训练集和测试集。接着, 在训练集上进行模型训练, 并使用交叉验证方法对模型进行评估。最后, 我们在测试集上进行模型性能测试, 并计算模型的准确性指标。

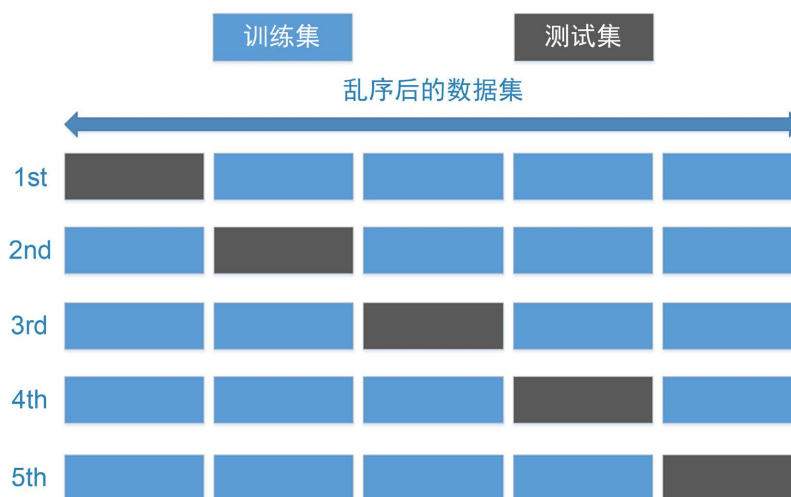


Figure 6. Five-fold cross-validation diagram

图 6. 五折交叉验证示意图

3.3.4. 最佳效果模型获取

训练模型的损失函数使用二值交叉熵(Binary Cross-Entropy), 二值交叉熵是一种常用的损失函数, 也称为对数损失函数(Log Loss), 通常用于二分类问题中。其主要思想是比较模型输出的概率与真实标签的概率之间的差异, 并将其最小化。优化器采用了可自适应地调节学习率的 Adam 优化器, 能有效缓解梯度震荡且收敛速度更快。根据经验将初始学习率设定为 $10e-4$ 。准确率(ACC)作为评价函数, 计算如下式(1):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

训练阶段的 Batch Size 设置为 128, 共迭代 100 个训练周期, 最终依据准确率评判型最佳模型。

3.4. 模型测试与评估

3.4.1. 脓毒症评判金标准

Raith EP 等[16]在 2016 年提出了脓毒症评判的金标准依据, 包括了 SOFA、qSOFA 和 SIRS 等评分系统。Fawcett [17]介绍了 ROC 曲线和 AUROC 指标。AUROC (Area Under the Receiver Operating Characteristic Curve)是一种用于评估二分类模型预测能力的指标。本文中模型是典型的二分类模型, 因此寻至模型判定最优阈值后绘制混淆矩阵将数据效果可视化。图 7 为混淆矩阵示意图。

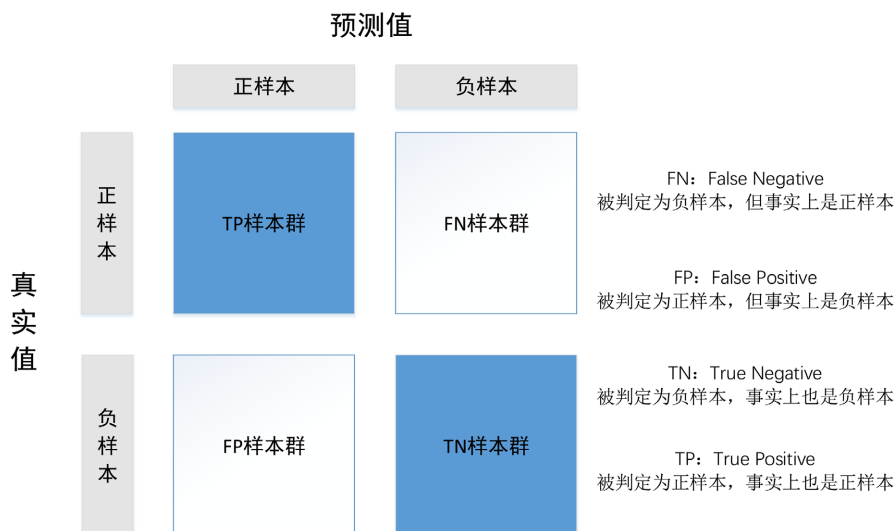


Figure 7. Diagram of the confusion matrix
图 7. 混淆矩阵示意图

在医学领域中, AUROC 经常被用来评估诊断试验的准确性和可靠性。同时在不损失模型的预测准确度的前提下应提尽量召回率(Recall), 以此降低危及患者风险。提升召回率并结合 AUROC 曲线寻求最佳解决方案。

AUROC 的取值范围是 0 到 1 之间, 其中 0.5 表示随机猜测, 1 表示完美预测。通常情况下, AUROC 值大于 0.7 被认为是有用的, 而大于 0.9 则被认为是非常准确的。

ROC 曲线反映了不同阈值下真正例率(True Positive Rate, TPR)与假正例率(False Positive Rate, FPR)之间的关系。TPR 指的是分类器正确识别出阳性样本的比例, 而 FPR 指的是分类器错误地将阴性样本分类为阳性的比例, 计算如下:

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{FP + FN} \quad (3)$$

AUROC 的计算方法是先将模型的预测结果按照从高到低的顺序排序, 然后逐个调整分类器的决策阈值, 并计算相应的 TPR 和 FPR。最终, 将 TPR 与 FPR 绘制在坐标系中, 得到 ROC 曲线。AUROC 的值则等于 ROC 曲线下的面积。

召回率(Recall, 同上文 TPR)指的是在所有实际为正样本的样本中, 分类器正确识别出阳性样本的比例。召回率可以用来评估分类器对于正样本的识别能力, 高召回率意味着分类器能够较好地识别出正样本, 但也可能导致误判负样本的数量增加。

3.4.2. 阈值选取与数据可视化

根据对 CNN-LSTM 模型的训练结果,我们寻找了最优的模型输出阈值,以尽可能提高模型的 AUROC 值。最终得出的最优阈值为 0.454。同时,我们绘制了 ROC 曲线,并生成了预测结果的混淆矩阵,具体如下图 8 至图 9 所示。

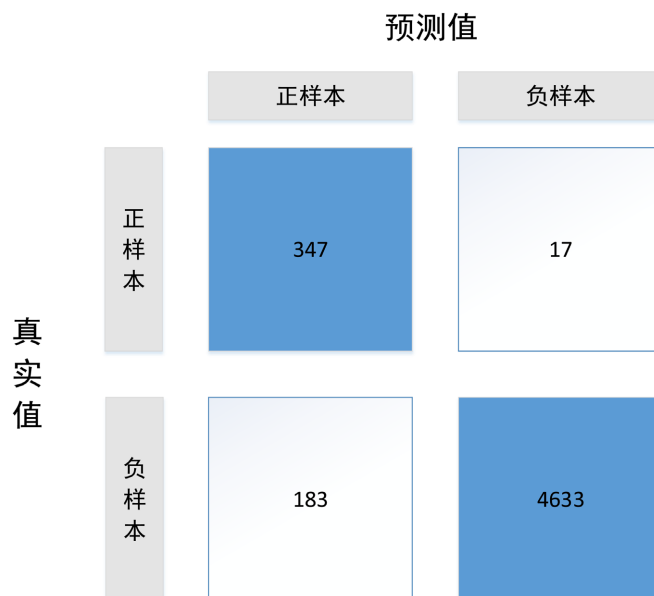


Figure 8. Confusion matrix for prediction

图 8. 预测混淆矩阵

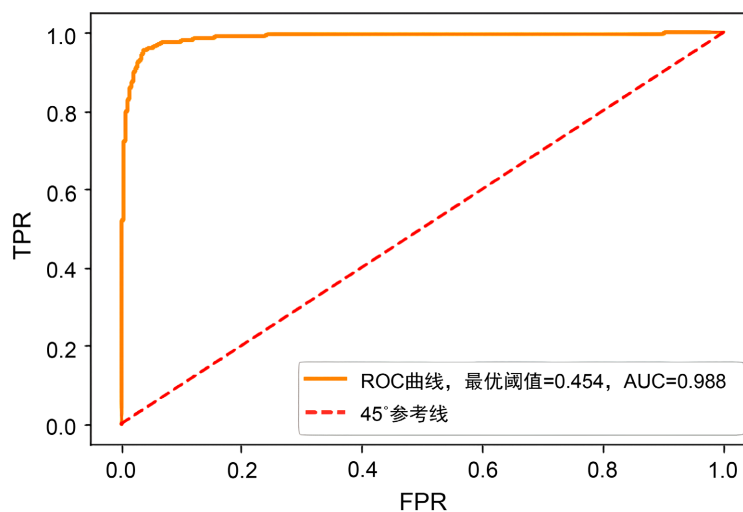


Figure 9. Predicted ROC curve

图 9. 预测 ROC 曲线

该模型在 AUROC 评价指标上表现出色,达到了 0.98 的高分值,召回率也达到了 0.95,充分展现了其出色的预测能力。这一结果为更好地预测脓毒症的爆发提供了有力支持。

3.4.3. 模型效果评估

本研究旨在病发前 10 小时对患者进行预测,并准确鉴别脓毒症患者。为此,我们对大量的指标数据

进行筛选和优化处理, 以确保数据集的质量, 并提高模型的计算能力。为验证所提出的模型的有效性和先进性, 我们进行了与公开文献[3] [18]-[23]中现有模型及评价指标的比较, 其在评估指标 AUROC 上取得了最优的结果。具体来说, 我们对模型表现进行了详细比较, 并将比较结果列于表 5 中:

Table 5. Comparison of related studies

表 5. 相关研究对比

作者	模型	数据来源	预测提前量	AUROC
Desautels	InSight	MIMIC-III	4 h	ML = 0.88 SIRS = 0.61 SOFA = 0.73 QSOFA = 0.77 MEWS = 0.8
Hornig	SVM	MIMIC-III	4 h	ML = 0.86
@Mao	InSight	University of California, San Francisco MIMIC-III	4 h	ML = 0.92 NEWS = 0.76 SOFA = 0.63 SIRS = 0.75
Nemati	APeX	Emory University Hospitals MIMIC-III	4 h	ML = 0.85 SOFA = 0.87
Calvert	InSight	MIMIC-III	3 h	ML = 0.83
Kam	Deep neural network	MIMIC-III	3 h	ML = 0.92
Faisa	Logistic regression	York hospital Northern Lincolnshire Google hospital	4 h	ML = 0.78
Yajun He	Deep neural network	PhysioNet/CinC Challenge 2019	10 h	ML = 0.98

ML = 机器学习, SIRS = 全身炎症反应综合症, SOFA = 序贯器官功能衰竭评分, QSOFA = 快速序贯器官衰竭评分, MEWS = 改良版早期预警评分。

因此, 通关相关研究结果对比可以得出结论, 本文所提出的模型表现出了优秀的性能和可靠性。

4. 结论

本研究采用神经网络算法结合人体生命特征数据, 对是否患有脓毒症进行预测, 并对模型进行了改进, 引入了 CLASS_WEIGHT 与 Dropout 技术, 使得模型提升了对小类样本的识别准确度与泛化能力, 对于患者脓毒症的预测更为准确, 并在以往相关研究的基础之上提前了预测时间点, 相比之下评分标准上达到了更高的分数。对于脓毒症的预测研究, 相比于传统医学得分诊断制与机器学习方法, 引入神经网络算法得到了更好的效果, 同时减轻了繁琐医疗诊断带来的公共资源消耗, 提升了患者的生活质量, 这证明了此方法的与有效性与先进性。

本文研究的方法中仍存在一些需改进之处, 未来计划采集更广泛的患者数据, 以及纳入更多生命特征指标, 改进模型层数与结构等方法提高模型的泛化能力与预测准确度。随着神经网络方法渗入医学领域, 使得医学得到了飞快发展, 也预示神经网络在医学诊断领域将具有更广阔的发展前景。

参考文献

- [1] 姚咏明, 盛志勇, 林洪远, 柴家科. 脓毒症定义及诊断的新认识[J]. 中国危重病急救医学, 2004(6): 321-324.

- [2] Fleuren, L.M., Klausch, T.L.T., *et al.* (2020) Machine Learning for the Prediction of Sepsis: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy. *Intensive Care Medicine*, **46**, 383-400. <https://doi.org/10.1007/s00134-019-05872-y>
- [3] Desautels, T., Calvert, J., *et al.* (2016) Prediction of Sepsis in the Intensive Care Unit with Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics*, **4**, e28. <https://doi.org/10.2196/medinform.5909>
- [4] Seymour, C.W., Liu, V.X., Iwashyna, T.J., *et al.* (2016) Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, **315**, 762-774. <https://doi.org/10.1001/jama.2016.0288>
- [5] Vincent, J.L. (2008) Clinical Sepsis and Septic Shock—Definition, Diagnosis and Management Principles. *Langenbeck's Archives of Surgery*, **393**, 817-824. <https://doi.org/10.1007/s00423-008-0343-1>
- [6] Johnson, A.E., Pollard, T.J., Shen, L., *et al.* (2016) MIMIC-III, a Freely Accessible Critical Care Database. *Scientific Data*, **3**, Article ID: 160035. <https://doi.org/10.1038/sdata.2016.35>
- [7] Singer, M., Deutschman, C.S., Seymour, C.W., *et al.* (2016) The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, **315**, 801-810. <https://doi.org/10.1001/jama.2016.0287>
- [8] Shankar-Hari, M., Phillips, G.S., Levy, M.L., *et al.* (2016) Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, **315**, 775-787. <https://doi.org/10.1001/jama.2016.0289>
- [9] Perner, A., Cecconi, M., Cronhjort, M., *et al.* (2018) Expert Statement for the Management of Hypovolemic Sepsis. *Intensive Care Medicine*, **44**, 791-798. <https://doi.org/10.1007/s00134-018-5177-x>
- [10] Prescott, H.C. and Angus, D.C. (2018) Post Sepsis Morbidity. *JAMA*, **319**, 91. <https://doi.org/10.1001/jama.2017.19809>
- [11] Reinhart, K., Daniels, R., Kissoon, N., *et al.* (2017) Recognizing Sepsis as a Global Health Priority—A WHO Resolution. *The New England Journal of Medicine*, **377**, 414-417. <https://doi.org/10.1056/NEJMp1707170>
- [12] Shi, X.J., Chen, Z.R., *et al.* (2015) A Machine Learning Approach for Precipitation Nowcasting.
- [13] Royal College of Physicians (2012) National Early Warning Score (NEWS): Standardising the Assessment of Acute Illness Severity in the NHS. Report of a Working Party, RCP, London.
- [14] Nitish, S., Geoffrey, E., *et al.* (2014) A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, **15**, 1929-1958.
- [15] Longadge, R. and Dongre, S. (2013) Class Imbalance Problem in Data Mining Review.
- [16] Raith, E.P., Udy, A.A., Bailey, M., *et al.* (2017) Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality among Adults with Suspected Infection Admitted to the Intensive Care Unit. *JAMA*, **317**, 290-300. <https://doi.org/10.1001/jama.2016.20328>
- [17] Fawcett, T. (2006) An Introduction to ROC Analysis. *Pattern Recognition Letters*, **27**, 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [18] Kam, H.J. and Kim, H.Y. (2017) Learning Representations for the Early Detection of Sepsis with Deep Neural Networks. *Computers in Biology and Medicine*, **89**, 248-255. <https://doi.org/10.1016/j.compbiomed.2017.08.015>
- [19] Calvert, J.S., Price, D.A., *et al.* (2016) A Computational Approach to Early Sepsis Detection. *Computers in Biology and Medicine*, **74**, 69-73. <https://doi.org/10.1016/j.compbiomed.2016.05.003>
- [20] Faisal, M., Scally, A., Richardson, D., *et al.* (2018) Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results. *Critical Care Medicine*, **46**, 612-618. <https://doi.org/10.1097/CCM.0000000000002967>
- [21] Horng, S., Sontag, D.A., Halpern, Y., *et al.* (2017) Creating an Automated Trigger for Sepsis Clinical Decision Support at Emergency Department Triage Using Machine Learning. *PLOS ONE*, **12**, e0174708. <https://doi.org/10.1371/journal.pone.0174708>
- [22] Mao, Q., Jay, M., Hoffman, J.L., *et al.* (2018) Multicentre Validation of a Sepsis Prediction Algorithm Using Only Vital Sign Data in the Emergency Department, General Ward and ICU. *BMJ Open*, **8**, e017833. <https://doi.org/10.1136/bmjopen-2017-017833>
- [23] Nemat, S., Holder, A., Razmi, F., *et al.* (2018) An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical Care Medicine*, **46**, 547-553. <https://doi.org/10.1097/CCM.0000000000002936>