

# 一种非结构化数据表征增强的术后风险预测模型\*

王亚强<sup>1,2,3†</sup>, 杨潇<sup>1,2,3</sup>, 郝学超<sup>4</sup>, 舒红平<sup>1,3</sup>, 陈果<sup>4†</sup>, 朱涛<sup>4†</sup>

<sup>1</sup>成都信息工程大学软件工程学院

<sup>2</sup>成都信息工程大学数据科学与工程研究所

<sup>3</sup>软件自动生成与智能服务四川省重点实验室

<sup>4</sup>四川大学华西医院麻醉手术中心

†通讯作者: yaqwang@cuit.edu.cn, grace\_chenguo@hotmail.com, 739501155@qq.com

## 摘要

准确的术后风险预测对临床资源规划和应急方案准备以及降低患者的术后风险和死亡率具有积极作用。术后风险预测目前主要基于术前和术中的患者基本信息、实验室检查、生命体征等结构化数据，而蕴含丰富语义信息的非结构化术前诊断的价值还有待验证。针对该问题，本文提出一种非结构化数据表征增强的术后风险预测模型，利用自注意力机制，精巧的将结构化数据与术前诊断数据进行信息加权融合。基于临床数据，将本文方法与术后风险预测常用的统计机器学习模型以及最新的深度神经网络进行对比，本文方法不仅提升了术后风险预测的性能，同时也为预测模型带来了良好的可解释性。

**关键词:** 文本数据；术后风险预测；自注意力机制；信息融合

## An Unstructured Data Representation Enhanced Model for Postoperative Risk Prediction

Yaqiang Wang<sup>1,2,3†</sup>, Xiao Yang<sup>1,2,3</sup>, Xuechao Hao<sup>4</sup>, Hongping Shu<sup>1,3</sup>, Guo Chen<sup>4†</sup>, Tao Zhu<sup>4†</sup>

<sup>1</sup>College of Software Engineering, Chengdu University of Information Technology

<sup>2</sup>Institute for Data Science and Engineering, Chengdu University of Information Technology

<sup>3</sup>Sichuan Key Laboratory of Software Automatic Generation and Intelligent Service

<sup>4</sup>Department of Anesthesiology, Sichuan University

†Corresponding author: yaqwang@cuit.edu.cn, grace\_chenguo@hotmail.com, 739501155@qq.com

## Abstract

Postoperative risk prediction has a positive effect on clinical resource planning, emergency plan preparation and reducing postoperative risk and mortality. Postoperative risk prediction is mainly based on patient's basic information, laboratory tests, vital signs and other structured data, while the value of unstructured preoperative diagnosis with rich semantic information remains to be verified. Aiming at attempting this problem, an unstructured data representation enhanced postoperative risk prediction model is proposed in this paper. The model utilizes self-attention to fuse structured data with preoperative diagnosis. Through comparing with the commonly used statistical machine learning models and the state-of-the-art deep neural networks, the proposed model has not only better prediction performance, but also better interpretability.

**Keywords:** Text Data, Postoperative Risk Prediction, Self-Attention Mechanism, Information Fusion

\*四川大学华西医院1.3.5项目(ZYJC21008)和国家重点研发计划项目(2018YFC2001800)资助

©2022 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

## 1 引言

由于术后并发症（如肺部并发症(魏娟 et al., 2021)、心血管不良(Lundberg et al., 2018)、ICU入室(Chiew et al., 2020)等）风险（简称术后风险）所导致的术后30天内死亡，已成为全球排名第三位的人群死亡原因(Li et al., 2022)。准确的术后风险预测对医生进行合理的临床资源规划、应急方案准备具有重要的辅助作用，对降低患者的术后风险和死亡率具有积极意义(Xue et al., 2021; Bonde et al., 2021)。

目前，术后风险预测主要基于术前和术中的患者基本信息（如体温、血压、体重等）、实验室检查（如氧分压、氧饱和、蛋白等）、生命体征（如手术时长、出血量等）等结构化数据，利用XGBoost、逻辑回归、随机森林、神经网络等模型实现(Lundberg et al., 2018; Xue et al., 2021)。

近年来深度神经网络在各领域的预测任务中表现优秀，受到研究者的广泛关注，也被引入到术后风险预测任务中(Bonde et al., 2021)。Fritz(2019)等人构建了一种多路径卷积神经网络，提取和融合患者基本信息、共病情况、术前实验室检查和术中生命体征等结构化数据中的特征，用于患者术后死亡风险预测。Barbieri(2020)等人利用双向门控循环单元，将结构化数据之间的时间信息以拼接的方式融入数据表征，采用注意力机制提取重要特征，用于患者术后ICU入室风险预测。现有方法的核心是如何将结构化数据中的离散型和连续型特征向量化，形成输入基于深度神经网络的术后风险预测模型的数据表征。

编号	体温	是否使用活性药物	收缩压	舒张压	术前诊断
患者1	36.5	1	156	76	1: 高血压病 (3级 很高危) 2: 肺部感染
患者2	36.4	0	113	70	直肠恶性肿瘤
患者3	36.7	0	105	66	左膝重度关节炎

图 1. 包含患者基本信息、实验室检查和术前诊断的术前数据示例

术前数据中除结构化数据之外，还包含语义丰富的非结构化术前诊断数据。术前诊断中不仅包含医生基于医学知识，根据局部的结构化数据，对患者病情的总结信息，还包含医生以整体的结构化数据为依据，利用经验知识，对患者病况的推断信息。如图1中患者1的术前数据所示，根据结构化数据收缩压156mmHg（毫米汞柱）与舒张压76mmHg，基于医学知识，成人的收缩压和舒张压正常范围分别在90至120mmHg，因此，医生在术前诊断中总结该患者有“高血压病”，且属于“3级很高危”。此外，依据目前患者整体的结构化数据，医生根据经验知识，推断患者是“肺部感染”。更进一步地，术前诊断的整体描述，反应了当前患者的全局状态。这些语义信息能够丰富术后风险预测的特征，有助于增强预测模型的性能。

然而，术前诊断数据尚未在术后风险预测任务中被有效利用，如何充分地利用非结构化的术前诊断数据，形成有效的术后风险预测数据表征，尚有待进一步探索。因此，本文围绕非结构化的术前诊断数据如何增强术后风险预测任务展开研究，主要的贡献包括以下三个方面：

1. 与围术期医学专家合作，经过清洗、处理、转换和去隐私过程后，构建了一份包含12240个实例、面向术后风险预测任务的数据集。该数据集的结构化数据部分包含了95列离散型变量、61列连续型变量、一列非结构化的术前诊断变量，以及三列二元的术后风险标签变量，分别表示肺部并发症、心血管不良和ICU入室风险的发生情况。
2. 为充分地利用非结构化的术前诊断数据，本文提出一种非结构化数据表征增强的术后风险预测模型，利用自注意力机制，将结构化数据与局部的细粒度实体信息及全局的粗粒度文本语义加权融合，有效地将非结构化数据用于增强术后风险预测性能。
3. 本文提出的基于自注意力机制融合结构化与非结构化数据的模型结构，为术后风险预测带来了良好的可解释性。细节实验结果分析发现，利用自注意力机制获得的关系权重矩阵，可以解释和展示出非结构化数据不仅增强了重要的结构化数据的贡献度，还补充了风险预测信息。

实验结果表明，本文提出的非结构化数据表征增强的术后风险预测模型明显优于所对比的常用统计机器学习模型和最新的深度神经网络，在三种重要的术后风险预测任务上，本文

提出的模型均取得了最优的结果，F1-Score分别达到了0.669、0.558和0.608。此外，通过消融实验，进一步验证了本文提出的模型有效地加权融合了局部的细粒度实体信息和全局的粗粒度文本语义信息。利用非结构化术前诊断数据表征增强术后风险预测模型后，肺部并发症风险预测性能提升了6.878%，心血管不良风险预测性能提升了9.541%，ICU入室风险预测性能提升了7.641%。

## 2 相关工作

术后风险预测是医学信息学领域的研究热点问题，目前的研究主要集中于机器学习模型应用于术后风险预测的有效性验证及面向特定类型的术后并发症风险特征分析两个层面。Canet(2010)等人利用逻辑回归模型，确定了7个独立且具有良好鉴别能力的危险因素，构建了术后肺部并发症风险预测指标，用于评估和预测术后肺部并发症的个体风险。Hill(2019)等人采用随机森林模型，自动地发现重要的术前特征，将结构化的美国麻醉医师协会身体状况特征与术前特征结合，提升术后死亡风险的预测性能。与先前工作不同，本文提出了一种非结构化数据表征增强的术后风险预测模型，该模型基于自注意力机制，在预测中有效地融合结构化数据和非结构化语义信息，并提供可解释性。

术后风险预测目前的主要研究对象是术前和术中的结构化数据，其中包含两种类型的变量，一种是离散型变量，另一种是连续型变量。通常会将连续型变量进行离散化后，与离散型变量一起共同构建特征向量，作为术后风险预测模型的输入(Arik and Pfister, 2021)。本文实验主要基于术前患者基本信息和实验室检查等结构化数据，采用与先前工作相同的连续型变量的基本处理方法。不同的是，借鉴Fritz(2019)等人的思想，本文将离散型变量和离散化的连续型变量一并构建离散特征词典，并基于深度学习学习离散特征的嵌入表征。

术后风险预测除可利用术前和术中的结构化数据作为特征之外，通过观察发现，包含医学语义信息的非结构化术前诊断数据可用于增强术后风险预测。Zhang(2020)等人提出将英文临床文本利用Doc2Vec模型(Le et al., 2014)直接形成数据表征，然后与结构化数据合并的方式，将非结构化数据与结构化数据融合，应用于住院死亡率、住院时间长短和术后30天再入院的预测任务，该方法在英文临床数据MIMIC-III(Johnson et al., 2016)上进行了实验验证。然而，与该工作不同，本文首次探索了将中文非结构化临床文本引入术后风险预测的方法。

此外，本文通过观察发现，非结构化的术前诊断中，既包含全局的粗粒度文本语义信息，还包含局部的细粒度实体信息，它们均可作为术后风险预测提供医学语义特征(如图1所示)。为将这些信息与离散特征的嵌入表征相融合，本文首先基于目前常用的中文MedBERT<sup>1</sup>获得实体的嵌入表征，并将术前诊断视为句子后，采用词嵌入平均池化的方法将其向量化。然后利用自注意力机制(Bahdanau et al., 2014)，将离散特征的嵌入表征与实体的嵌入表征以及向量化的术前诊断进行加权融合，在充分地综合利用全局和局部的文本语义信息的基础上，还为模型带来良好的可解释性(Hao et al., 2021)。

## 3 术后风险预测

### 3.1 任务定义

本文将术后风险预测定义为一项二分类任务，采用有监督学习方法解决。定义 $(\mathbf{x}, y)$ 为一个训练实例， $\mathbf{x}$ 中包含 $\mathbf{x}_{num}$ 、 $\mathbf{x}_{cat}$ 和 $\mathbf{x}_{PD}$ 三种特征数据。其中， $\mathbf{x}_{num}$ 表示表格数据中的连续型特征数据，共 $m$ 列， $\mathbf{x}_{cat}$ 表示表格数据中的离散型特征数据，共 $n$ 列， $\mathbf{x}_{PD}$ 表示非结构化术前诊断文本数据， $y$ 表示术后风险的发生情况，用1或0分别表示风险的发生或未发生。

### 3.2 表格数据的向量表示

本文提出的术后风险预测模型主要利用结构化表格数据和非结构化术前诊断文本数据对术后风险进行预测(如图2所示)。结构化表格数据由 $\mathbf{x}_{num}$ 和 $\mathbf{x}_{cat}$ 组成。本文采用分类与回归树算法(Loh, 2011)，先将连续型特征转换为离散型特征，在引入医学语义信息的同时，降低数据的复杂度。转换之后的连续型特征不仅能够表达医学语义信息，还被统一成离散型特征。转换后的连续型变量表征被定义为 $\mathbf{x}_{n2cat}$ ：

$$\mathbf{x}_{n2cat} = discretized(\mathbf{x}_{num}) \quad (1)$$

<sup>1</sup><https://code.ihub.org.cn/projects/1775>

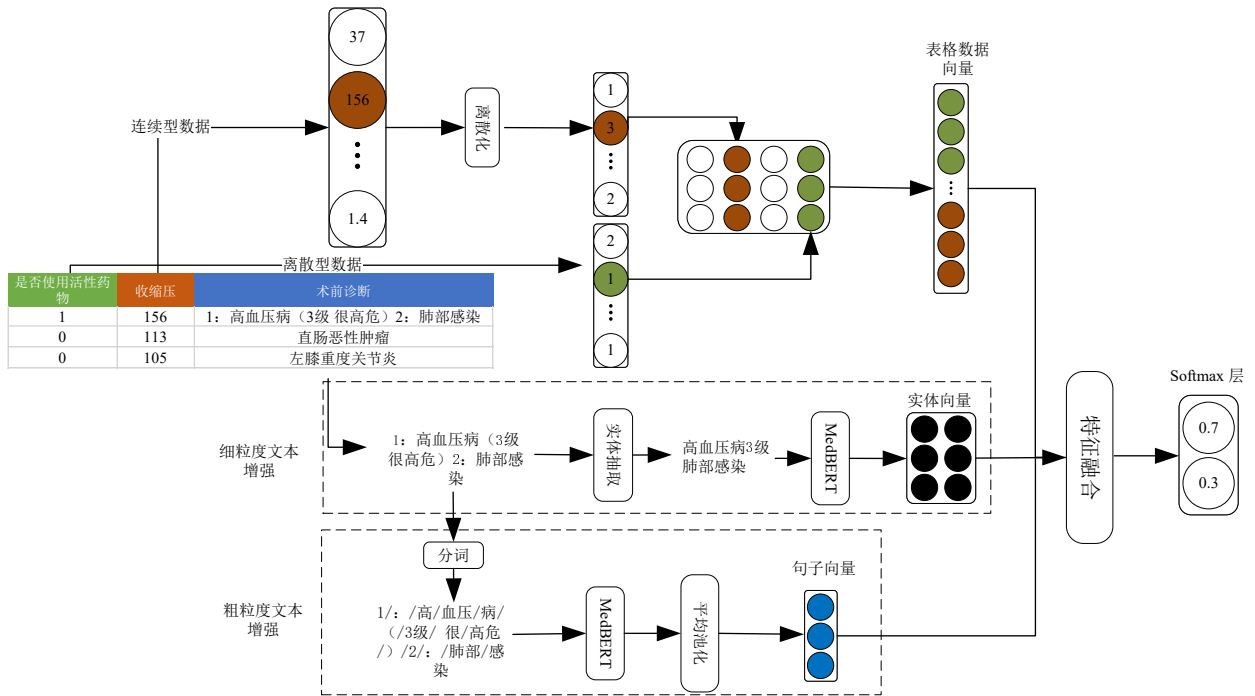


图 2. 模型结构图

处理离散型变量表征的常用方式是采用实体嵌入(Guo and Berkhahn, 2016)的方式, 即为每一个离散型变量构建一个特征词表, 词表大小为当前离散型变量的不同取值的数量。然而该方法在建模的过程中仅考虑了单一变量下的不同取值之间的语义关联性, 而不同的变量之间的相关性未被考虑其中。为引入全局不同变量之间的语义关联性, 本文改进了原始的实体嵌入方式, 让所有的离散型变量共用相同的特征词表。每一个离散型变量(包括 $x_{cat}$ 和 $x_{n2cat}$ )的不同取值, 都会赋予唯一的索引值 $x_i$ , 其中 $i \in [0, |V|]$ ,  $|V|$ 是所有离散型变量的所有不同取值的数量总和, 即共用的特征词表的词表大小。每个 $x_i$ 都将通过学习过程被映射为一个维度为 $d$ 的向量, 定义为 $e_{tabular}$ , 其中 $d$ 为超参数。通过构建全局共用的特征词表, 原始的离散型变量转换为语义向量之后, 不仅扩充了语义信息, 而且不同离散型变量之间也产生了语义关联, 相比原始的实体嵌入方式, 解决了不同离散型变量之间语义关联缺失的问题。最后, 将所有的 $e_{tabular}$ 拼接形成表格数据的向量表示 $E_{tabular}$ 。

### 3.3 文本数据的向量表示方法

术前诊断文本 $x_{PD}$ 主要包含医生总结的病人身体症状和初步推断的病情描述, 两者可统一定义为病症实体。因此, 术前诊断文本可以归纳为由多个病症实体、连接词以及标点符号构成的集合, 每个实例 $x_{PD}$ 包含 $l_{max}$ 项的病症实体,  $l_{max}$ 表示数据集中,  $x_{PD}$ 最多可饱含的病症实体数量。

术前诊断文本可以有两种向量表示方式, 一种是形如利用Doc2Vec模型(Le et al., 2014)得到的全局语义向量, 获取该类向量表示的方法我们称之为粗粒度文本的向量表示方法; 另一种是直接将病症实体对应的语义向量拼接, 形成细粒度文本的向量表示方法。后文将具体介绍两种方式获取术前诊断文本粗粒度语义信息和细粒度语义信息的方法。

#### 3.3.1 粗粒度文本的向量表示方法

为获取术前诊断文本的粗粒度语义向量, 先将文本进行分词<sup>1</sup>, 得到分词列表 $\{token_0, token_1, \dots, token_p\}$ , 其中 $p$ 表示文本分词后得到的词的数量。将分词列表输入领域微调后的预训练模型MedBERT中, 生成维度为768的动态词向量列表 $\{e_0^{768}, e_1^{768}, \dots, e_p^{768}\}$ , 768是MedBERT的词向量维度。为进一步获取句子向量, 选择采用快速且高效的平均池化方法整合词向量的语义信息。对词向量矩阵中的每一列取均值, 将词向

<sup>1</sup>本文实验中直接采用了<https://huggingface.co/hfl/chinese-macbert-base>的内置分词工具

量矩阵压缩为包含整个术前诊断语义信息的粗粒度文本的向量表示  $e_{sentence}$ 。

$$e_{sentence} = MeanPooling(\{e_0^{768}, e_1^{768}, \dots, e_p^{768}\}) \quad (2)$$

### 3.3.2 细粒度文本的向量表示方法

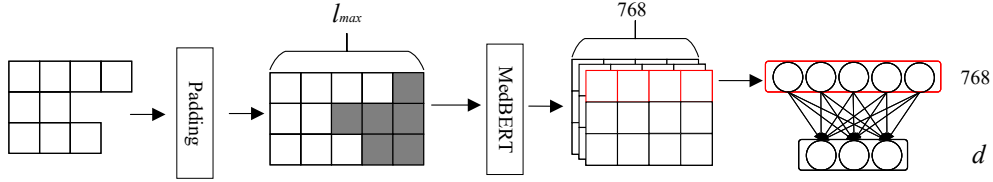


图 3. 细粒度文本增强

将术前诊断文本分词后通过MedBERT生成的词向量简单压缩为一个向量的过程中可能会造成局部语义信息的丢失，且无法明确术前诊断文本中哪些信息在术后风险预测过程中起到了关键作用。为保留术前诊断文本的局部细粒度实体语义信息，首先利用医学领域数据集基于BERT+BiLSTM+CRF模型训练得到实体抽取模型(Dai et al., 2019)，然后利用该模型抽取  $x_{PD}$  中的病症实体，形成病症实体集合  $\{w_0, \dots, w_k, \dots, w_K\}$ ，其中， $K$  表示当前  $x_{PD}$  中抽取得到的病症实体数量。

因  $x_{PD}$  中包含的实体数量可能不一致，为后续处理统一，本文将病症实体数量达到  $l_{max}$  的集合，通过补全特殊字符[*PAD*]的方式，形成数量均为  $l_{max}$  的实体集合（如图3所示）。然后，每一个实体  $w_k$  将通过MedBERT转换为蕴含医学语义的向量  $e_k^{768}$ 。为后续与表格数据的向量表示进行融合，细粒度文本语义向量将进一步通过全连接层降维，从768维降至  $d$  维，得到降维后的细粒度语义向量集合  $\{e_0^d, e_1^d, \dots, e_{l_{max}}^d\}$ 。最后，将含有全局语义信息的粗粒度文本的向量表示和含有局部语义信息的细粒度文本的向量表示组合，得到最终的文本数据的向量表示  $E_{text}$ 。

$$E_{text} = \{e_0^d, e_1^d, \dots, e_{l_{max}}^d, e_{sentence}\} \quad (3)$$

### 3.4 特征融合

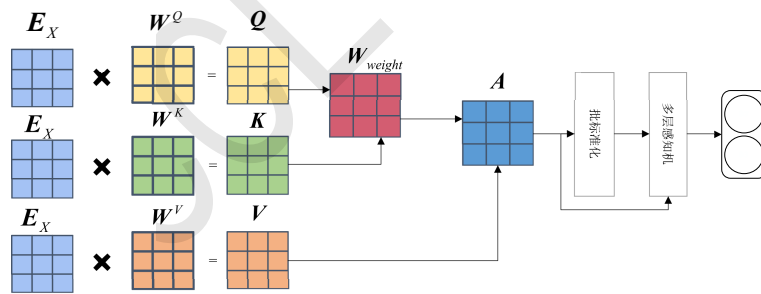


图 4. 特征融合

在特征融合层，本文选择使用Self-Attention(Bahdanau et al., 2014)将表格数据和文本数据进行特征融合（如图4所示）。首先，将表示表格数据信息的特征向量  $E_{tabular}$  与表示文本语义信息的特征向量  $E_{text}$  拼接，形成新的特征向量集合  $E_X$ 。并将  $E_X$  通过三个参数矩阵  $W^Q$ 、 $W^K$  和  $W^V$  映射为三个不同的矩阵  $Q$ 、 $K$  和  $V$ 。然后对  $Q$  和  $K^T$  执行点积并利用  $d_k$  放缩结果，以保证训练过程中梯度的稳定性， $d_k$  是指矩阵  $K$  的维度，计算方法如公式(4)  $softmax$  函数的输入所示。随后执行  $softmax$  操作进行归一化，得到不同的特征向量之间（包含文本数据向量和表格数据向量）的注意力权重  $W_{weight}$ （计算公式如公式(4)所示）。

$$\mathbf{W}_{weight} = softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \quad (4)$$

最后将  $\mathbf{W}_{weight}$  与  $\mathbf{V}$  相乘得到增强后的特征表示  $\mathbf{A}$ 。具体计算过程如下公式所示。

$$\mathbf{E}_X = \mathbf{E}_{tabular} \oplus \mathbf{E}_{text} \quad (5)$$

$$\mathbf{Q} = \mathbf{E}_X \mathbf{W}^Q, \mathbf{K} = \mathbf{E}_X \mathbf{W}^K, \mathbf{V} = \mathbf{E}_X \mathbf{W}^V \quad (6)$$

$$\mathbf{A} = Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{W}_{weight} \mathbf{V} \quad (7)$$

通过注意力机制，模型可以自动地学习到特征在推理过程中的重要性或贡献度。因此，在模型推理过程中，可以通过提取并分析注意力权重矩阵，来探究在模型预测过程中，各特征发挥作用的重要程度。

为了解决梯度消失的问题，受到(He et al., 2016; Ba et al., 2016)的启发，新的特征矩阵  $\mathbf{A}$  在输入前馈神经网络之前，还经过了残差网络和层标准化操作。接着将向量输入到带有sigmoid激活函数的前馈神经网络中，计算预测术后风险的发生概率  $P$ ，其中  $\mathbf{W}$  和  $\mathbf{b}$  均是前馈神经网络将学习得到的参数。

$$P = sigmoid(\mathbf{W}^T \mathbf{A} + \mathbf{b}) \quad (8)$$

最后，模型的损失定义为：

$$loss = -\frac{1}{M} \sum_{i=0}^M (y_i \log P_i + (1 - y_i) \log(1 - P_i)) \quad (9)$$

其中  $M$  指批量包含的实例数量。

## 4 实验

### 4.1 实验数据

本文实验采用了从医院的临床数据系统中获取的数据，其中包含患者的基本身体状况信息、实验室检查数据和术前诊断，以及病人术后发生的肺部并发症、心血管不良和ICU入室结局。该数据经过了如下基本的预处理过程：

- 删除了有关患者身份的个人身份。
- 删除了缺失率高于50%的变量。

最终得到包含12240个实例的术后风险预测任务数据集，包含有95项离散型变量和61项连续型变量以及1项术前诊断变量。该数据集包含的三种术后风险的标签分布如图5所示，阳性率分别是15.93%、6.25%和3.02%。实验将数据集按照7:1:2的比例划分得到训练集、验证集和测试集。

### 4.2 评估指标

为了评估模型的效果，本文采用精确率 (Precision)、召回率 (Recall) 和F1-Score作为主要的评估指标。计算公式如下：

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

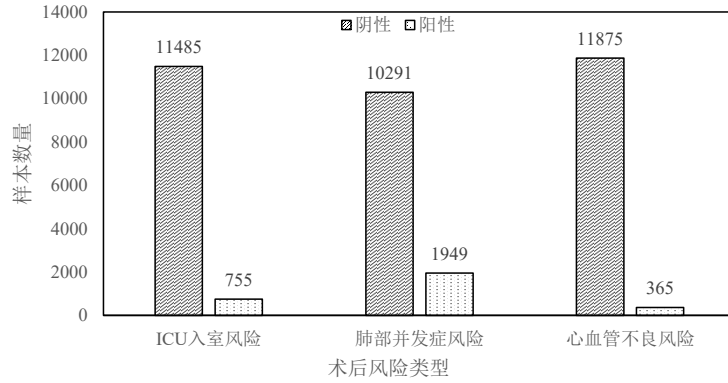


图 5. 数据集标签分布

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

其中,  $TP$ 表示在阳性实例中, 模型预测为阳性的实例数量,  $FP$ 表示在阴性实例中, 模型预测为阳性的实例数量,  $FN$ 表示在阳性实例中, 模型预测为阴性的实例数量。

### 4.3 参数设置

模型训练采用了Adam优化器, 初始学习率设置为 $3e-4$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , 批量大小设置为128, 训练轮次设置为100, 超参数 $d$ 设置为32, dropout参数设置为0.5。在以上超参数设置条件下, 本文提出的术后风险预测模型达到了收敛。

为验证本文提出的模型在术后风险预测任务上的有效性, 选择了两种常见的统计机器学习模型LR和XGBoost以及两种用于表格数据的最新的深度神经网络Wide&Deep(Cheng et al., 2016)和Tabtransformer(Huang et al., 2020)作为对比模型。LR和XGBoost采用scikit-learn框架(Pedregosa et al., 2011)实现, Wide&Deep和Tabtransformer分别采用开源的代码库实现<sup>1</sup>。

### 4.4 实验分析

首先在三项术后风险预测任务上对比了模型的预测性能, 实验结果如表1所示。

Model	肺部并发症风险			ICU 入室风险			心血管不良风险		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
LR	71.508	31.068	43.316	72.840	36.875	48.963	45.833	13.415	20.755
XGBoost	69.965	48.293	57.143	74.118	39.375	51.429	41.176	8.537	14.141
WideDeep	73.214	54.089	62.215	74.603	37.600	50.000	52.778	29.788	37.255
Tabtransformer	68.563	60.422	64.236	75.385	39.200	51.579	55.556	30.303	39.216
Our	68.378	65.723	<b>66.909</b>	65.088	57.664	<b>60.833</b>	77.395	44.260	<b>55.888</b>

表 1. 实验整体结果

从表1中可以观察得到, Wide&Deep和Tabtransformer在三项术后风险的预测任务上均优于LR和XGBoost, 特别是在阳性率较低的心血管不良风险预测任务上, Wide&Deep和Tabtransformer的表现远优于LR和XGBoost。该结果说明, 深度神经网络在术后风险预测任务上的性能优于统计机器学习模型, 这与(Bonde et al., 2021; Fritz et al., 2019)报告的结果保持一致。

此外, 从表1还可以看出, 通过引入了术前诊断文本数据表征, 本文提出的模型在肺部并发症、心血管不良和ICU入室三个风险预测任务上均取得了最优的效果, F1-Score分别达到了66.909%、55.888%和60.833%。该结果证明, 本文提出的文本数据表征增强的术后风险预测模型是有效的。

进一步观察表1中的结果发现, 相比于其它模型, 本文提出的模型是在保持了良好的精确率的条件下, 召回率得到了大幅提升, 从而提升F1-Score结果。该结果说明, 当模型预测引入

<sup>1</sup>[https://github.com/jrzaaurin/pytorch-widedeep/tree/pytorch\\_widedeep](https://github.com/jrzaaurin/pytorch-widedeep/tree/pytorch_widedeep)

术前诊断数据表征后，进一步丰富了特征的语义信息，对阳性实例的预测带来了额外的语义信息补充，从而帮助模型将之前无法判断的阳性实例准确地预测为阳性，进而提高了模型的召回率。

### 4.5 消融实验

为进一步验证文本数据表征对模型预测效果增强的作用，并探究文本的粗粒度文本语义信息和细粒度文本语义信息对预测任务的影响，本文还设计了不加入文本以及分别加入粗粒度和细粒度的文本语义信息的对比消融实验，结果如表2所示，其中-E是指去除细粒度文本的语义向量，-S是指去除粗粒度文本的语义向量，-E-S是指去除所有的文本数据。

Model	肺部并发症风险			ICU 入室风险			心血管不良风险		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Our-E-S	60.842	59.444	60.031	53.878	52.598	53.192	52.854	42.687	46.347
Our-E	62.547	62.394	62.754	56.365	53.670	54.913	53.551	42.663	46.905
Our-S	68.089	66.010	66.883	61.129	58.152	59.570	79.697	43.029	55.577
Our	68.378	65.723	<b>66.909</b>	65.088	57.664	<b>60.833</b>	77.395	44.260	<b>55.888</b>

表 2. 消融实验结果

观察表2发现，在加入文本数据后，本文提出的模型在肺部并发症风险预测任务上的性能提高了6.878%，在ICU入室风险预测中提高了7.641%，在心血管不良风险预测中提高了9.541%，并且无论是单独加入粗粒度文本的语义向量还是细粒度文本的语义向量，模型的预测性能均得到明显改善。该结果说明，非结构化术前诊断中的信息对术后风险预测具有积极的作用，为术后风险预测提供了额外的决策信息，有效地增强了模型的预测能力。

此外，观察表2还可以发现，阳性率越低的术后风险，通过引入非结构化术前诊断数据表征后，模型的预测性能提升越高。该结果说明，对于阳性实例更少的术后风险，模型需要更多的特征才能更准确地预测阳性病例，术前诊断的引入能够为模型带来更丰富的特征，从而使得本文提出的模型在阳性率越低的术后风险预测中表现得越出色。

从表2结果还能够看出，相比于全局的粗粒度文本语义向量的缺失，模型对于局部的细粒度文本语义向量的缺失更加敏感，该结果说明，在术后风险预测的过程中引入围术期医学领域知识，对模型的预测性能提升具有重要的作用。这也进一步说明了本文提出的非结构化数据表征增强的术后风险预测模型的有效性和应用价值。

更进一步地，从表2还可以看出，当模型同时引入粗粒度文本的向量表示和细粒度文本的向量表示时，模型的预测性能达到最优。该结果说明，在非结构化数据表征增强术后风险预测模型时，既需要引入粗粒度文本向量携带的全局语义信息，又需要引入细粒度文本向量携带的局部语义信息。

### 4.6 细节分析

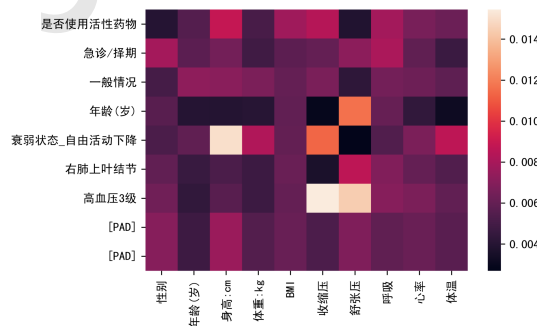


图 6. 心血管不良实例的注意力权重热力图

本文提出的模型通过自注意力机制为术后风险预测模型带来了可解释性，为验证和说明该效果，本文选取了一个发生了术后心血管不良的病人实例，该实例的术前诊断是“右肺上叶结节，高血压3级”。本文提出的模型准确地预测得到该病人实例将发生术后心血管不良风险。通



过提取模型中的注意力权重矩阵  $W_{weight}$ ，画出热力图，见图6。在图6中，纵轴的中“右肺上叶结节”和“高血压3级”是指术前诊断中的实体病症，“PAD”指补全的字符，其余的特征均是表格数据的部分特征，横轴也是部分表格数据的特征。

从图6中可以看到，在术前诊断描述中，“高血压3级”显著地与表格数据中的收缩压和舒张压变量具有强关联。该强关联预示着模型通过训练，学习到了数据集中包含的医学领域知识关联信息，该关联信息保存在了  $W_{weight}$  中，在术后风险预测中起到了重要的预示作用。另一个方面，该结果还说明，利用自注意力机制为术后风险预测模型还带来了可解释性。总体地，实验结果验证了本文提出的模型在增强术后风险预测性能方面的鲁棒性。

通过对比引入非结构化数据表征前后， $W_{weight}$  中包含的权重值按列求和后得到每个变量在术后风险预测中的权重比率排序，进一步观察在风险预测中，起重要作用的变量与术后风险结局是否存在医学语义相关性，对比结果如图7所示。

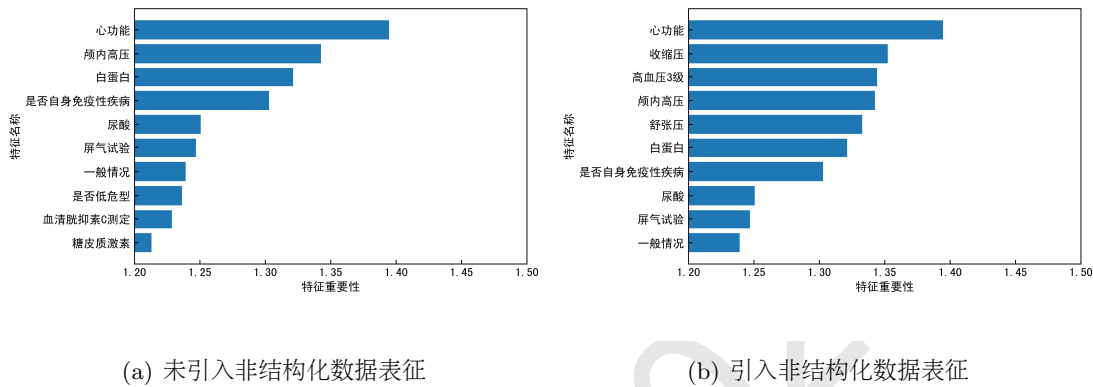


图 7. 变量在模型术后风险预测中的重要性排序

在图7中，权重比率越高，说明变量在预测中具有更高的重要性。从结果可以看出，在引入非结构化数据表征后，与术后心血管不良风险强相关的医学变量收缩压与舒张压的权重比率排序更加靠前。该结果进一步证实了本文提出的模型在提升术后风险预测性能的同时，还学习到了医学领域知识，并且具有更好的可解释性。

从图7中还可以看到，在引入非结构化数据表征后，临床医生根据经验总结和/或推断得到的额外的重要医学语义信息，也在风险预测中起到了更重要的作用，例如临床医生基于收缩压和舒张压总结并记录下的“高血压3级”疾病。一方面，该结果进一步证明了，本文提出的模型学习到了医学领域知识，并对术后风险预测起到了积极的作用。另一方面，该结果也说明，本文模型的直觉观察是正确的，术前诊断中包含了大量的医学领域信息，这些信息既包含表格数据中已有的信息，还包含了大量可以用于丰富原始表格数据的额外的医学领域知识信息，这些信息会对模型的预测性能提升起到积极的作用。更进一步地，该结果也说明，本文提出模型在提升了术后风险预测性能的同时，还具有良好的模型可解释性。

## 5 结束语

术后风险预测在临床医学中具有重要意义，基于表格数据构建统计机器学习模型和深度神经网络，实现术后风险预测是常见的方式。非结构化术前诊断数据中蕴含了大量额外的医学领域知识，可为术后风险预测提供丰富的信息，然而它们尚未被有效利用。针对这个问题，本文提出了一种新的模型，将非结构化数据表征增强术后风险预测，并在模型中引入自注意力机制，在有效融合表格数据和非结构化数据的同时，为模型带来良好的可解释性。实验结果表明，本文提出的非结构化数据表征增强的术后风险预测模型的性能显著高于其他比较的基线模型。通过消融实验，验证了在术后风险预测中引入非结构化术前诊断数据的重要性，证明了本文提出的模型的有效性。此外，通过对模型的注意力权重的细节分析，发现通过自注意力机制，将表格数据与非结构化的术前诊断数据融合到术后风险预测中，为模型带来了良好的模型可解释性。

## 参考文献

- 魏娟, 邓惠民, 吕欣. 2021. 术后肺部并发症围手术期风险因素及防治策略. *同济大学学报(医学版)*, 42(06):736-743.
- Alexander Bonde, Kartik M. Varadarajan, Nicholas Bonde, Prof Anders Troelsen, Orhun K. Muratoglu, Henrik Malchau, Anthony D. Yang, Prof Hasan Alam, Martin Sillesen. 2021. Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. *The Lancet Digital Health*, 3(8):e471-e485.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Bing Xue, Dingwen Li, Chenyang Lu, Christopher R. King, Troy Wildes, Michael S. Avidan, Thomas Kannampallil, Joanna Abraham. 2021. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Network Open*, 4(3):e212240-e212240.
- Bradley A. Fritz, Zhicheng Cui, Muhan Zhang, Yujie He, Yixin Chen, Alex Kronzer, Arbi Ben Abdallah, Christopher R. King, Michael S. Avidan. 2019. Deep-learning model for predicting 30-day postoperative mortality. *British Journal of Anaesthesia*, 123(5):688-695.
- Brian L. Hill, Robert Brown, Eilon Gabel, Nadav Rakocz, Christine Lee, Maxime Cannesson, Pierre Baldi, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, Eran Halperin. 2019. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. *British Journal of Anaesthesia*, 123(6):877-886.
- Cheng Guo, Felix Berkhahn. 2016. Entity embeddings of categorical variables. *arXiv preprint arXiv:1604.06737*.
- Calvin J. Chiew, Nan Liu, Ting Hway Wong, Yilin E. Sim, Hairil R. Abdullah. 2020. Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission. *Annals of Surgery*, 272(6):1133-1139.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825-2830.
- Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah. 2016. Wide & deep learning for recommender systems. *In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp.7-10.
- Yaru Hao, Li Dong, Furu Wei, Ke Xu. 2021. Self-Attention Attribution: Interpreting Information Interactions Inside Transformer. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12963-12971.
- Jaume Canet, Lluís Gallart, Carmen Gomar, Guillem Paluzie, Jordi Vallès, Jordi Castillo, Sergi Sabaté, Valentín Mazo, Zahara Briones, Joaquín Sanchis. 2010. Prediction of postoperative pulmonary complications in a population-based surgical cohort. *Anesthesiology*, 113(6):1338-1350.
- Jimmy Lei Ba, Jamie Ryan Kiros, Geoffrey E. Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. 2016. Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778.
- Wei-Yin Loh. 2011. Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1):14-23.
- Peiyi Li, Yunmei Luo, Xuexin Yu, Elizabeth Mason, Zhi Zeng, Jin Wen, Weimin Li, Mohammad S. Jalali. 2022. Readiness of Healthcare Providers for e-Hospitals: A Cross-sectional Analysis in China before COVID-19. *BMJ Open*, 12(2):e054169.
- Sercan O. Arik, Tomas Pfister. 2021. Tabnet: Attentive interpretable tabular learning. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):6679-6687.

- Sebastiano Barbieri, James Kemp, Oscar Perez-Concha, Sradha Kotwal, Martin Gallagher, Angus Ritchie, Louisa Jorm. 2020. Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-risk. *Scientific Reports*, 10(1):1111.
- Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10):749-760.
- Xin Huang, Ashish Khetan, Milan Cvitkovic, Zohar Karnin. 2020. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*.
- Zhenjin Dai, Xutao Wang, Pin Ni, Yuming Li, Gangmin Li, and Xuming Bai. 2019. Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records. *In the 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp.1-5.
- Dongdong Zhang, Changchang Yin, Jucheng Zeng, Xiaohui Yuan, and Ping Zhang. 2020. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Medical Informatics and Decision Making*, 20(1):280.
- Quoc V. Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. *In Proceedings of the 31st International Conference on Machine Learning*, pp.1188-1196.
- Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035.