

DOI: 10.16781/j.CN31-2187/R.20230243

· 论 著 ·

基于脉搏波频域梅尔频率倒谱系数特征的高血压危险分层预测模型

齐晨浩¹, 杨晶东^{1*}, 邱泽浩¹, 尧明慧², 燕海霞²

1. 上海理工大学光电信息与计算机工程学院自主机器人实验室, 上海 200093

2. 上海中医药大学中医学院中医诊断学教研室, 上海 201203

[摘要] **目的** 为改进基于人工智能技术高血压时域脉搏波分类模型精度低、泛化性能差的问题, 提出一种基于融合注意力机制的频域脉搏波预测模型。**方法** 首先将时域脉搏波转换为频域梅尔频率倒谱系数特征, 增强脉搏波区分度, 采用时间卷积网络与 Transformer 结构提取脉搏波深层特征, 并将自注意力机制与选择性内核注意力进行决策融合, 提取脉搏波关联特征, 并采用 Floodings 正则化方法间接控制训练损失, 防止过拟合发生。针对上海中医药大学附属龙华医院及上海市中西医结合医院提供的 527 例临床脉诊数据, 进行 5 折交叉验证实验。此外, 采用梯度提升决策树算法统计脉搏波频域特征的贡献率排名, 分析影响模型分类精度的关键因素, 为中医临床辅助诊断提供参考价值。**结果** 本研究提出的模型分类评估指标准确度、F1 值、精确率、召回率和 AUC 值分别为 0.939 6、0.924 9、0.940 9、0.929 5 和 0.993 4。脉搏波的静态特征、一阶差分和二阶差分系数的贡献率相对均衡, 说明高血压危险程度不仅与脉搏波的静态特征相关, 也应当考虑脉搏波的动态特征。**结论** 与典型脉搏波分类模型相比, 本研究提出的模型具有较高的分类精度和泛化性能。

[关键词] 高血压; 危险分层; 梅尔频率倒谱系数; 时间卷积网络; Transformer

[引用本文] 齐晨浩, 杨晶东, 邱泽浩, 等. 基于脉搏波频域梅尔频率倒谱系数特征的高血压危险分层预测模型[J]. 海军军医大学学报, 2024, 45 (10): 1226-1240. DOI: 10.16781/j.CN31-2187/R.20230243.

Hypertension risk stratification prediction model based on frequency-domain pulse wave Mel-scale frequency cepstral coefficient features

QI Chenhao¹, YANG Jingdong^{1*}, QIU Zehao¹, YAO Minghui², YAN Haixia²

1. Autonomous Robot Lab, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

2. Department of Chinese Medicine Diagnosis, School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

[Abstract] **Objective** To propose a frequency-domain pulse wave prediction model based on fusion attention mechanism, improving the low classification accuracy and poor generalization performance of hypertension time-domain pulse wave classification based on artificial intelligence technology. **Methods** Firstly, the time-domain pulse wave was transformed into frequency-domain Mel-scale frequency cepstral coefficient features to enhance its discriminability. Then, temporal convolutional network and Transformer structures were employed to extract the deep features of pulse waves, and self-attention mechanism and selective kernel attention were combined for decision fusion to extract relevant features. Floodings regularization method was used to indirectly control the training loss and prevent overfitting. A 5-fold cross-verification experiment was conducted based on 527 clinical pulse diagnosis data provided by Longhua Hospital, Shanghai University of Traditional Chinese Medicine and Shanghai Traditional Chinese Medicine-Integrated Hospital. Additionally, the extreme gradient boosting algorithm was employed to calculate the contribution rate ranking of frequency-domain pulse wave features, and the key factors affecting the classification accuracy of the model were analyzed to provide reference for the clinical auxiliary diagnosis of traditional Chinese medicine. **Results** The evaluation metrics accuracy, F1 score, precision, recall rate and area under curve value of the model proposed in this study were 0.939 6, 0.924 9, 0.940 9, 0.929 5, and 0.993 4, respectively. The static characteristics of the pulse wave, the contribution rate of the first-order difference and the second-order difference coefficients were relatively balanced, indicating that the degree of hypertension risk was not only related to the static characteristics of the pulse wave, but

[收稿日期] 2023-04-29 **[接受日期]** 2023-09-04

[基金项目] 国家自然科学基金(81973749), 中国中医科学院科技创新工程重大攻关项目(CI2021A01503). Supported by National Natural Science Foundation of China (81973749) and Major Research Project of Scientific and Technological Innovation Program of China Academy of Chinese Medical Sciences (CI2021A01503).

[作者简介] 齐晨浩, 硕士生. E-mail: 18918498657@163.com

*通信作者(Corresponding author). Tel: 021-55271064, E-mail: eerfriend@163.com

also to the dynamic characteristics of the pulse wave. **Conclusion** The proposed model has higher classification accuracy and generalization performance compared to typical pulse wave classification models.

[**Key words**] hypertension; risk stratification; Mel-scale frequency cepstral coefficient; temporal convolutional network; Transformer

[**Citation**] QI C, YANG J, QIU Z, et al. Hypertension risk stratification prediction model based on frequency-domain pulse wave Mel-scale frequency cepstral coefficient features[J]. Acad J Naval Med Univ, 2024, 45(10): 1226-1240. DOI: 10.16781/j.CN31-2187/R.20230243.

高血压的分级及危险分层是临床上诊断高血压经常用到的标准。高血压的分级依据血压水平来划分, 主要分为 3 级。危险分层是在高血压分级的基础上, 根据患者的危险因素和病史区分, 分为低危、中危、高危、很高危。高血压是心脑血管疾病的高危因素, 现已成为我国老年人最常见的慢性病之一^[1]。目前, 我国高血压患者数量约为 2.7 亿例^[2], 但随着人口老龄化逐渐加重及高血压的诊断标准可能下调, 中国高血压患者数量接近 5 亿, 而我国高血压患者的知晓率、治疗率、控制率很低^[3]。这导致脑卒中、心肌梗死及严重的靶器官损伤等心血管疾病患者增多, 医疗负担加重, 所以高血压的防治刻不容缓。

近年来, 随着深度学习等在医学领域应用的渗透, 许多学者提出了使用人工智能技术实现高血压患者危险分层的诊断方法。崔佳嘉等^[4]通过提取梅尔频率倒谱系数 (Mel-scale frequency cepstral coefficient, MFCC), 并运用局部线性嵌入算法 (local linear embedding) 对现有的 MFCC 特征向量降维改进, 保证能保留变压器噪声信号的有效信息。Zhang 等^[5]采用 9 层卷积神经网络 (convolutional neural network, CNN) 对脉搏波进行分类, 平均准确率达到了 93.49%, 但研究的样本量较小, 缺少交叉验证支撑实验结论, 模型的泛化性没有得到进一步验证。笔者团队构建了基于 CNN-双向长短时记忆 (bi-directional long short-term memory, BiLSTM) 结构的高血压脉诊分类网络, 该模型可以有效地提取脉诊局部及全局序列特征, 其准确率可达 77.17%^[6]。马明艳等^[7]使用 CNN-BiLSTM 网络对入侵的恶意行为进行检测, 精度达到了 92%。该方法在 CNN 和 BiLSTM 基础上进行融合, 具有“并联”多通道融合特性, 但当数据时序较长时, BiLSTM 模型训练时间较长, 出现特征冗余问题。杨如民等^[8]采用蝙蝠算法和随机森林算法对脉搏波信号进行分类, 分类精度提高了 3.4%, 但

该方法需要根据经验确定特征维度, 且主观因素较多。张选等^[9]基于 GoolgLeNet 和 ResNet 框架提出 MIRNet 模型对健康和亚健康人群脉象分类, 精度可达到 87.84%。杨浩等^[10]运用 CNN 和 BiLSTM 对心房节拍分类, 精度可达到 90.21%。Ouyang 等^[11]使用一种四层多任务融合 CNN 对 2 型糖尿病患者的脉搏波进行分类, 精度最高可达 90.6%, 但当 CNN 层数较深时才能提取到脉搏波深层特征, 容易造成梯度消失问题。Chen 等^[12]采用基于基本残差块 (BasicBlocks) 的 ResNet 框架对女性脉搏波分类, 判断是否处在妊娠期, 预测准确率达到 84%, 该模型虽然在一定程度上解决了梯度消失问题, 但其深度多在 5 层附近, 无法学习到脉搏波周期内的深度时序特征。Yan 等^[13]提出了一种基于脉搏波特征融合的无创血压预测方法, 使用基于梯度提升决策树的回归算法预测患者的收缩压与舒张压, 相较于水银血压计, 这种方法虽然预测精度更高, 但对穿戴设备有一定要求。本文提出了一种通过脉搏波诊断高血压患者危险层级的神经网络融合模型, 该模型采用不同注意力机制进行决策融合, 对采集的中医脉诊数据进行分析, 从而得到患者高血压分层结果。该模型可以根据患者的脉诊数据实现对高血压层级的自主分类, 为高血压层级诊断提供了新思路。

1 基于融合模型的脉搏波分类模型

1.1 MFCC 特征提取 针对脉诊时域数据, 本研究使用了一种改进的 MFCC 提取方法来获取脉搏波的频域特征。脉搏波信号的非稳定性使得传统的 MFCC 分析方法无法获得完整的频域特征, 因此, 本研究采取了一些操作以提高分析精度。首先进行预加重操作以增加高频部分的分辨率, 然后进行分帧操作以保证观测单位内至少包含 1 个脉搏波周期^[14-15], 并选择汉明窗 (hamming) 进行加窗以增加帧左右两端的连续性。在脉搏波分解阶段, 为了更好地分离不同模态成分的固有模态函数 (intrinsic

mode function) 分量, 本研究加入了自适应白噪声, 以消除不同时间尺度成分的叠加, 提高分解精度并减轻模态混叠效应。最后, 采用脉搏波频域的 MFCC 来描述脉搏波特征, 包括静态和动态特征, 以提高脉搏波特征的区别性。

MFCC 特征提取的过程主要包括预加重、分帧、加窗、加噪声、经验模态分解 (empirical mode decomposition, EMD)、希尔伯特 (Hirbert) 变换、时间维度积分、Mel 三角滤波器、对数运算与离散余弦变换, 具体流程如图 1 所示。

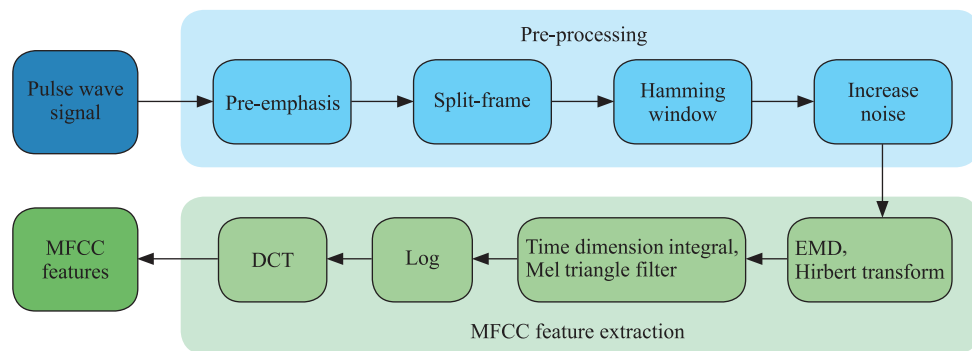


图 1 MFCC 特征提取过程

Fig 1 MFCC feature extraction process

MFCC: Mel-scale frequency cepstral coefficient; EMD: Empirical mode decomposition; DCT: Discrete cosine transform.

1.2 注意力机制融合策略 Transformer 使用的注意力机制为自注意力机制 (self-attention), 为了使模型具有较好的泛化能力和鲁棒性, 本研究考虑使用辅助注意力机制, 从不同侧面弥补单个注意力的不足。考虑选择的辅助注意力机制包括压缩和激励 (squeeze-and-excitation, SE)、选择性内核 (selective kernel, SK)、卷积块注意力模块 (convolutional block attention module, CBAM)、基于相似性的注意机制 (a simple parameter-free attention module, SimAM) 和通道注意力机制 (coordinate attention, CA) 共 5 种。在原有模型的基础上讨论了 3 种融合方法, 其模型结构如图 2 所示。融合方法 (a) 将辅助注意力机制作为辅助

模型, 对 2 个模型分别输入脉搏波频域特征, 得到的输出结果经过决策融合, 再输出最终预测结果。融合方法 (b) 将辅助注意力机制与自注意力机制并行, 2 个注意力机制都接受经过时间卷积网络 (temporal convolutional network, TCN) 处理过的特征数据, 然后再将输出结果进行决策融合, 得到最终预测。融合方法 (c) 则将辅助注意力机制融合在 TCN 模块内部。本研究在 TCN 内部的卷积层之间插入了辅助注意力机制, 以帮助提取特征之间的关联信息。经过实验, 本研究最终选择了 SK 注意力机制作为辅助注意力机制, 并采用融合方法 (a), 得到了表现最优的模型。

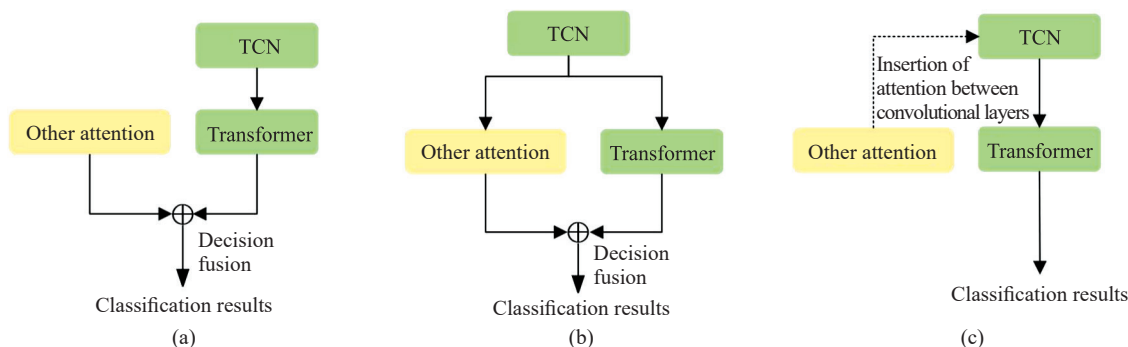


图 2 注意力机制 3 种融合方法 (a)、(b)、(c) 的结构图

Fig 2 Structural diagram of 3 fusion methods (a), (b), and (c) in attention mechanism

TCN: Temporal convolutional network.

1.3 正则化方式 在使用融合模型后, 整个网络的复杂度有所增加, 模型训练速度减慢, 更容易出

现过拟合现象, 而正则化能在训练模型的过程中防止模型出现过拟合, 为此本研究分析了常见的 3 种

正则化方式 (L1、L2 和 Max) 对模型性能的影响, 但实验结果表明这 3 种正则化方式对模型性能没有起到帮助作用, 甚至使得模型性能有所降低, 不适合处理脉搏波频域数据。于是选用了间接控制训练损失的洪泛法 (flooding) 策略^[16], Flooding 策略使训练损失在一个小常量附近浮动, 当训练损失达到合理小的值时, 该方法会阻止训练损失的进一步减少, 从而防止训练损失趋近于 0。其表达式为

$$\tilde{L}(\theta) = |L(\theta) - b| + b$$

式中 $L(\theta)$ 表示原来的损失函数, $\tilde{L}(\theta)$ 表示处理后的损失函数, b 是超参数阈值, 本实验中选择参数 b 值为 0.01。但是由于参数 b 的值较小, 其发挥作用在模型训练的后期, 本实验设想让其更早期地发挥作用, 故对其进行改进, 提出了 Floodings 策略, 即

$$\tilde{L}(\theta) = |L(\theta) - a/i - b| + a/i + b$$

式中 $L(\theta)$ 、 $\tilde{L}(\theta)$ 含义不变, a 、 b 是超参数, i 代表一轮训练中的迭代次数, 本实验选择参数 a 值为 1, 参数 b 值为 0.01。通过加入分式项, 引入额外的约束假设, 该方法就可以提前控制训练损失值的下降, 减轻过拟合现象, 使得模型更接近全局最优解。

1.4 脉搏波频域特征分类融合模型 针对原始的脉搏波时域波形数据, 本实验首先使用 Sym7(SymletsN)

小波基进行降噪处理, 然后使用平滑先验法 (smoothness priors approach, SPA) 完成对原始数据的去基线漂移处理。接着使用一种改进的 MFCC 提取方法来获取脉搏波的 36 维 MFCC 频域特征。本实验所使用的时间卷积网络 Transformer-选择性内核[temporal convolutional network Transformer-selective kernel (Floodings), TCNT-SK(Fs)] 模型结构如图 3 所示, 其中 TCN 包含有 2 个隐藏层, 隐藏通道 (hidden_channel) 数分别为 24 和 36, 为了提高准确率, TCN 中还加入了残差卷积的跳层连接及 1×1 的卷积操作; Transformer 模块只使用了编码器 (encoder) 部分的 2 层结构, 多头注意力参数 (head) 选择为 4。而 SK 模型则先通过 CNN 卷积提取特征, 升高数据维度至 18, 然后经过池化操作, 输入 SK 网络。经过改进的 MFCC 提取方法得到的 36 维频域特征包含了输入脉搏波的包络信息和频域边际谱信息, 将频域特征分别输入 TCNT 和 SK 模型, 经过各自网络模型处理后使用全连接层输出, 2 个模型的输出采用等权重的决策融合, 计算出最终的预测结果。本实验使用的 2 个不同注意力机制可以从不同角度获取特征之间的联系, 再将信息整合从而获得更好的模型性能。TCNT-SK(Fs) 模型算法步骤如表 1 所示。

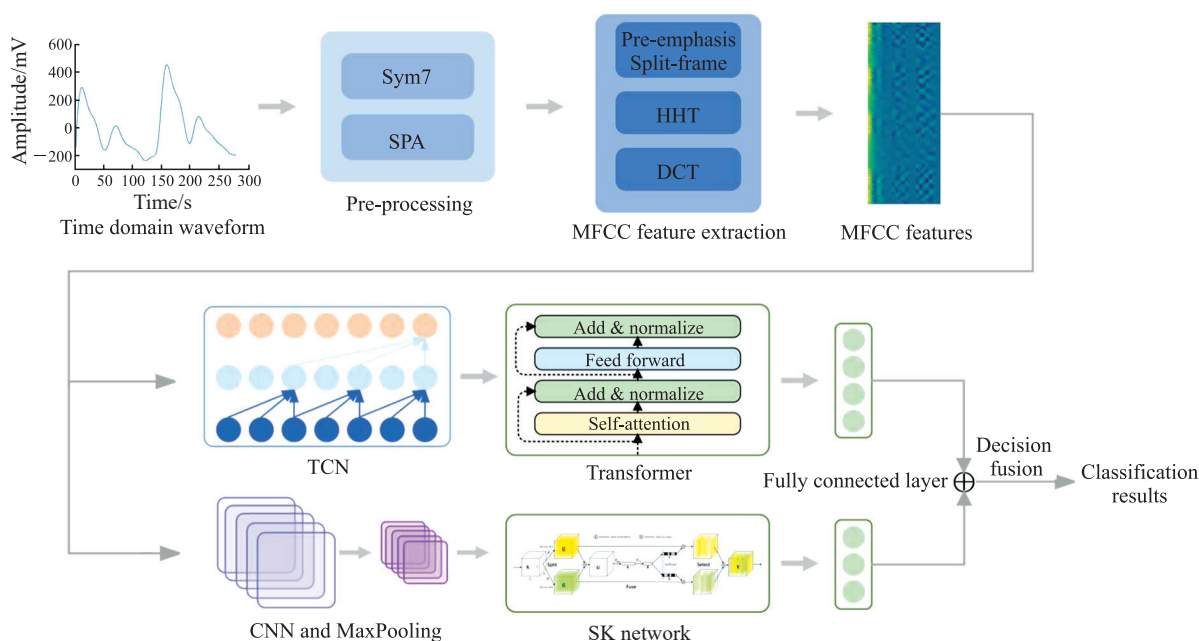


图 3 TCNT-SK(Fs) 模型结构

Fig 3 TCNT-SK(Fs) model structure

TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Floodings); SPA: Smoothness priors approach; HHT: Hilbert-Huang transform; DCT: Discrete cosine transform; MFCC: Mel-scale frequency cepstral coefficient; TCN: Temporal convolutional network; CNN: Convolutional neural network; SK: Selective kernel.

表 1 TCNT-SK(Fs) 模型训练伪代码

Tab 1 TCNT-SK(Fs) model training pseudocode

Algorithm: framework for our deep learning model TCNT-SK(Fs)

Input: $X(n)$, $Y(n)$. X : target pulse data; Y : labels (0, 1, 2); n : the number of pulse data

Output: predictive labels

```

1. for  $i=1, 2, \dots, n$  do
2.      $X_i = \text{preprocessing}(X_i)$ 
3.      $features[i] = \text{MFCC}(X_i)$ 
4. end
5. iteration  $\leftarrow 0$ 
6. while iteration  $<$  number of epochs do
7.     initialize the model parameters
8.     iteration  $\leftarrow$  iteration + 1
9.     for each batch in data do
10.        TCN layer:  $X_1 = \text{TCN}(features)$ 
11.        Transformer:  $X_1 = \text{Trans}(X_1)$ 
12.        CNN layer:  $X_2 = \text{ConvReluMax}(features)$ 
13.        SK layer:  $X_2 = \text{SK}(X_2)$ 
14.        FC layer:  $dec_1 = \text{Linear}_1(X_1)$ ,  $dec_2 = \text{Linear}_2(X_2)$ 
15.         $pre = dec_1 + dec_2$ 
16.    end
17. return all pre

```

TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Flooding); MFCC: Mel-scale frequency cepstral coefficient; TCN: Temporal convolutional network; CNN: Convolutional neural network; SK: Selective kernel; FC: Fully connected.

1.5 模型参数设置 TCNT-SK(Fs) 模型的参数设置如表 2 所示。原始脉搏波信号在提取频域特征后存在 MFCC 长度不一致的问题, 因此采用数据中最大长度作为行标准, 不足时用 0 补充, 最大长度为 200, 每个维度包含 36 个特征, 将填充好的 MFCC 数据作为模型的输入。本模型主要由 TCNT 和 SK 模型的决策融合组成, TCNT 模型中的 TCN 包含 2 个隐藏层, 其隐藏通道数分别为 24 和 36; 另外为了有效地应对长历史信息这一问题, TCN 结构中使用了空洞卷积来增加感受野。空洞卷积隐含 1 个超参数膨胀尺寸 (dilation size), 表示卷积核 (kernel) 的间隔数量, 膨胀尺寸随着层数增加呈指数增长, 因此一般使用较小的卷积核就可以获得很大的感受野, 这里选择卷积核为 2。本实验中 Transformer 模型只使用了编码器 Encoder 部分中的 2 层结构, 多头注意力参数为 4, 输入通道数对应 TCN 的输出为 36 维。而 SK 模型则先通过卷积层将脉搏波频域数据提升维度至 18 维, 卷积核选择 5 可以获取更大感受野, 接着进行池化操作, 然后输入输入通道参数 (input channel) 为 18 的 SK 模型。2 个模型均通过最后的扁平层 (flatten) 和线形层 (linear) 进行输出, 再将输出结果进行决策融合得到最终的模型预测结果。

表 2 TCNT-SK(Fs) 模型超参数

Tab 2 TCNT-SK(Fs) model hyperparameters

Layer	Filter size	Output shape
Input		200×36
TCN_layer1	2×2 , stride=1	200×24
TCN_layer2	2×2 , stride=1	200×36
Transformer	Head=4, $n=2$	200×36
Flatten		$1 \times 7\ 200$
Linear		1×3
Convolutional	5×5 , stride=1	$18 \times 200 \times 36$
MaxPooling	2×2 , stride=1	$18 \times 100 \times 18$
SK	Input channel=18	$18 \times 100 \times 18$
Flatten		$1 \times 7\ 200$
Linear		1×3
Decision fusion		1×3

TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Flooding); TCN: Temporal convolutional network; SK: Selective kernel.

2 材料和方法

2.1 样本分布 本实验使用脉诊仪采集脉搏波数据, 样本采集前需要患者放松 5 min 以上, 并在 60 d 内完成采样。采样时的触点压力根据患者脉络深浅来确定, 使用的采样频率包括 200 Hz、700 Hz、1 000 Hz。本实验使用上海中医药大学附属龙华医院及中西医结合医院提供的 527 例临床脉诊数据。男性病例 223 例 (42.31%), 平均年龄为 (66.35 $\pm$ 10.37) 岁; 女性病例 304 例 (57.69%),

平均年龄为 (71.32±8.51) 岁。该数据集包含 82 例 (15.56%) 低危、303 例 (57.50%) 中危、142 例 (26.94%) 高危和极高危高血压患者。

2.2 脉诊数据预处理 模型的训练需要确保数据维度一致性, 而脉诊采样频率有 200 Hz、700 Hz、1 000 Hz 3 种, 为此本实验将 700 Hz 和 1 000 Hz 的数据进行降采样, 统一为 200 Hz。然而, 脉搏波的采集过程难免会受到一些干扰, 数据往往存在噪声和基线漂移等问题, 这可能会干扰脉搏波信号特征的提取, 因此本实验对原始脉搏波信号进行了降噪和去除高频噪声等处理。因为傅里叶变换难以分辨信号在时间轴上的突变点, 所以选择了多贝西 (Daubechies, dbN) 小波和对称多贝西 (symmetric Daubechies, SymN) 小波进行实验。dbN 小波具有较好的正则性, 该小波作为稀疏基所引入的光滑误差不容易被察觉, 使得信号重构过程比较光滑。SymN 具有近似对称性, 可以很有效地避免相位畸变, 该小波基对应的滤波器具有线性相位的特点。

本实验采用评估参数信噪比 (signal noise ratio, SNR) 和均方根误差 (root mean square error, RMSE) 评估脉搏波降噪、去除基线漂移的效果。SNR 值越大、RMSE 值越小说明去噪效果越好, 表达式为

$$SNR=10\lg\left(\sum_{n=1}^N\frac{p^2(n)}{[p'(n)-p(n)]^2}\right)$$

$$RMSE=\sqrt{\sum_{n=1}^N\frac{[p'(n)-p(n)]^2}{N}}$$

式中 $p(n)$ 为原始信号, $p'(n)$ 为处理后信号, N 为信号长度。

表 3 列出了不同小波基、不同信号长度下的评估指标 SNR 和 RMSE 值。通过对比发现, 选择 Symlets 系列中的 Sym7 小波作为小波基^[17]对脉搏波信号进行降噪处理得到的效果最好, 其 SNR 为 36.867 7, RMSE 为 14.842 5, 均是最优值。SymN 小波在连续性、支集长度、滤波器长度等方面与 dbN 小波一致, 但 SymN 小波具有更好的对称性, 即一定程度上能够减少对信号进行分析和重构时的相位失真。

常用的基线漂移去除方法包括 SPA^[18-19]、EMD^[20-21] 及低通滤波 (low-pass filter), 对降噪之后的数据使用这 3 种方法进行去基线漂移, 结果如表 4 所示。SPA 是一种信号非线性去趋势算法, 它可以保留基本的脉搏波特征。SPA 算法具有较高

的 SNR (23.888 9) 和较小的 RMSE (39.683 8), 更适合作为脉搏波信号的基线漂移去除处理。本实验的预处理操作即采用小波降噪和 SPA 分别完成对原始数据的降噪和去基线漂移处理。

表 3 不同小波基降噪效果比较

Tab 3 Comparison of different wavelet base noise reduction effects

Signal length	dbN		SymN	
	SNR	RMSE	SNR	RMSE
1	28.149 3	40.274 1	28.149 3	40.274 1
2	36.412 0	15.637 3	36.412 0	15.637 3
3	36.813 3	14.935 0	36.813 3	14.935 0
4	36.848 5	14.875 2	36.850 6	14.871 5
5	36.860 1	14.855 5	36.863 0	14.850 4
6	36.859 7	14.856 2	36.865 3	14.846 5
7	36.858 7	14.857 8	36.867 7	14.842 5
8	36.860 1	14.855 5	36.864 1	14.848 6
9	36.861 3	14.853 4	36.863 9	14.849 0

dbN: Daubechies; SymN: Symmetric Daubechies; SNR: Signal noise ratio; RMSE: Root mean square error.

表 4 不同方法去基线漂移效果比较

Tab 4 Comparison of effects of different methods for removing baseline drift

Method	SNR	RMSE
SPA	23.888 9	39.683 8
EMD	21.893 5	55.496 6
LPF	11.613 9	59.982 4

SNR: Signal noise ratio; RMSE: Root mean square error; SPA: Smoothness priors approach; EMD: Empirical mode decomposition; LPF: Low-pass filter.

原始脉搏波样本经过小波降噪、去除基线漂移后的对比图像如图 4 所示。分析图像可知, 经过预处理操作后, 原始脉搏波信号的基线漂移和均值均有所降低, 波形噪声减少并且趋于稳定, 具有较好的周期性和可区分性。

3 实验结果和分析

3.1 数据集划分及性能评价指标 本实验在训练模型时均采用 K 折交叉验证法进行动态检验, 其中 K 取值为 5。在具体训练过程中, 首先将所有数据集划分成 5 份, 每份数据均从 3 类样本中随机抽取得到, 不重复地每次取其中 1 份做验证集, 其余 4 份做训练集进行模型训练, 之后计算该模型在验证集上的各项评价指标得分。循环上述操作 5 次得到 5 次不同的实验结果, 对 5 次实验结果进行统计计算得到平均值和方差, 并作为该模型结构的最终性能指标。为了评估不同模型的性能, 采

用了准确度、F1 值、精确率和召回率 4 个算法指标, 以及 ROC 曲线及其 AUC 值和精确率-召回率 (precision-recall, PR) 曲线及其下方围成的面积平均精度来全面评价不同模型的性能。其中准确度衡量的是所有样本分类正确的比例, 精确率代表对正例样本预测的准确程度, 召回率表示在所有正例样本中被正确分类的比例, F1 值是模型精确率和召回率的一种加权平均, 综合考虑了分类模型的精确率和召回率。由于本实验为三分类问题, 使用微平均 (micro-averaging) 方法计算各个评价指标, 微平均方法是对数据集中每个实例不分类进行统计

建立全局混淆矩阵并基于这个全局混淆矩阵计算相应指标的方法, 因此各评价指标计算公式为

$$\text{准确度} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{精确率} = \frac{TP}{TP + FP}$$

$$\text{召回率} = \frac{TP}{TP + FN}$$

$$\text{F1 值} = \frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$$

式中 TP、FP、TN、FN 分别是真阳性、假阳性、真阴性和假阴性样本数。

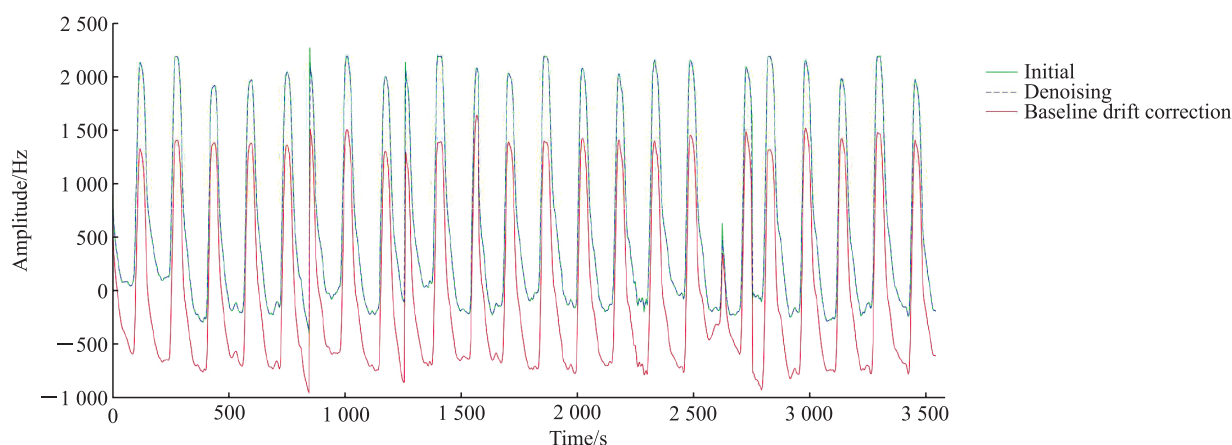


图 4 去除基线漂移、噪声后的数据波形图

Fig 4 Data waveform after removing baseline drift and noise

3.2 消融实验

3.2.1 模型选择 针对提取出频域特征的脉诊数据集, 采用 TCNT-SK(Fs) 融合模型, 并与 Transformer、CNNT 等模型相比, 进行了 5 折交叉验证。各模型的脉搏波频域特征分类结果见表 5, 分析表中数据可知, 在 Transformer 模型中加入 CNN 卷积层可以有效提高模型的性能, 其准确度提高了 2.60%, F1 值提升了 1.02%, 精确率提升了 2.16%, 召回率提升了 1.65%, 这是因为 CNN 卷积层可以对 MFCC 的维度特征进行整合, 提取出具有局部关联关系的脉诊特征, 改善了分类的效果。TCN 使用了扩张卷积, 层数越深、卷积窗口越大, 卷积窗口中的“空孔”越多, 说明这种卷积方式适合于处理与时间有关的序列。因此, TCNT 模型相较于 CNNT 模型, 将简单的 CNN 替换为了 TCN, 其在准确度上提高了 0.69%, F1 值提升了 3.68%, 精确率提升了 3.56%, 召回率提升了 2.75%。

Transformer 模型使用的注意力机制为自注意力机制, 为了进一步改善模型性能, 本实验选择了另外一种 SK 注意力机制, 从不同方面弥补单个注意力机制的不足。将脉搏波频域特征分别输入 2 个注意力模型, 结果经过决策融合后输出最终预测结果, 得到组合模型 TCNT-SK。可以看出, TCNT-SK 模型的性能相较于单个 TCNT 有所提升, 其准确度提高了 2.53%, F1 值提升了 3.35%, 精确率提升了 0.94%, 召回率提升了 2.21%。但是由于使用组合模型会使得网络更为复杂, 梯度下降过程会出现更多的不确定性, 模型有可能发生过拟合, 为此引入了 Flooding 策略, 使得训练损失在一个小常量附近浮动, 防止训练损失趋近于 0, 本实验对其进行改进并得到了鲁棒性更强、泛化性更好的 TCNT-SK(Fs) 模型。相较于 TCNT-SK 模型, TCNT-SK(Fs) 模型的准确度提高了 0.89%, F1 值提升了 1.41%, 精确率提升了 0.67%, 召回率提升了 1.90%。

表 5 不同模型的脉搏波频域特征分类结果

Tab 5 Frequency domain feature classification results of pulse wave based on different models

Model	$n=5, \bar{x} \pm s$			
	Accuracy	F1 score	Precision	Recall
Transformer	0.879 2 $\pm$ 0.023 4	0.842 5 $\pm$ 0.035 1	0.875 2 $\pm$ 0.039 1	0.854 5 $\pm$ 0.055 0
CNNT	0.902 1 $\pm$ 0.064 8	0.851 1 $\pm$ 0.143 7	0.894 1 $\pm$ 0.107 1	0.868 6 $\pm$ 0.131 4
TCNT	0.908 3 $\pm$ 0.034 5	0.882 4 $\pm$ 0.039 5	0.925 9 $\pm$ 0.025 6	0.892 5 $\pm$ 0.040 6
TCNT-SK	0.931 3 $\pm$ 0.008 3	0.912 0 $\pm$ 0.011 6	0.934 6 $\pm$ 0.008 8	0.912 2 $\pm$ 0.007 9
TCNT-SK(Fs)	0.939 6 $\pm$ 0.017 9	0.924 9 $\pm$ 0.019 4	0.940 9 $\pm$ 0.018 0	0.929 5 $\pm$ 0.020 1

CNNT: Convolutional neural network Transformer; TCNT: Temporal convolutional network Transformer; SK: Selective kernel; Fs: Floodings.

3.2.2 TCN TCN 是一种用来解决时间序列预测的算法。使用不同组合 TCNT 模型的脉搏波频域特征分类结果见表 6, 分析表中数据可知, 使用单层 TCN 时, 通道数参数选择 36 得到的分类器分类效果较好, 准确度为 0.904 2, F1 值为 0.870 3, 精确率为 0.919 1, 召回率为 0.884 2; 而继续增大通道数时分类器的分类效果反而下降, 这是由于过大的

通道数会提取到冗余信息, 从而影响后续训练, 导致模型性能下降。当使用多层 TCN 时, 可以发现对于 2 层 TCN 模型通道数 24-36 的组合性能表现优于 36-72 的组合, 这是由于通道数为 36 和 72 时提取特征信息过多, 模型复杂度增加, 导致模型性能降低。而 3 层 TCN 虽可以提取到深层特征, 但网络深度增加, 复杂度过高容易造成过拟合。

表 6 不同组合 TCNT 模型的脉搏波频域特征分类结果

Tab 6 Frequency domain feature classification results of pulse wave based on different combined TCNT models

Channel	$n=5, \bar{x} \pm s$			
	Accuracy	F1 score	Precision	Recall
24	0.889 6 $\pm$ 0.029 9	0.847 8 $\pm$ 0.034 1	0.906 9 $\pm$ 0.020 2	0.848 0 $\pm$ 0.041 3
36	0.904 2 $\pm$ 0.029 0	0.870 3 $\pm$ 0.035 7	0.919 1 $\pm$ 0.027 6	0.884 2 $\pm$ 0.046 7
72	0.868 8 $\pm$ 0.040 9	0.800 1 $\pm$ 0.111 1	0.873 5 $\pm$ 0.084 8	0.815 0 $\pm$ 0.106 4
24-36	0.908 3 $\pm$ 0.034 5	0.882 4 $\pm$ 0.039 5	0.925 9 $\pm$ 0.025 6	0.892 5 $\pm$ 0.040 6
36-72	0.868 8 $\pm$ 0.093 8	0.809 2 $\pm$ 0.149 6	0.847 7 $\pm$ 0.170 8	0.845 2 $\pm$ 0.107 3
24-36-72	0.889 6 $\pm$ 0.036 4	0.851 1 $\pm$ 0.046 6	0.904 7 $\pm$ 0.026 0	0.848 5 $\pm$ 0.052 1

TCNT: Temporal convolutional network Transformer.

表 7 列出了不同卷积层数 CNNT 模型的分类结果, 分析表中数据可知, 随着卷积层的增加, 分类结果的准确度提高, 模型性能有所提升, 这是由于卷积层数较少时模型无法学习到脉搏波特征之间的深层关联信息, 导致模型的各项评估指标较低。而当卷积层增加到 4 层时, 模型性能又有所降低, 这是由于 CNN 提取到了更为抽象的深层信息, 而

丢失了脉搏波的表象信息, 降低了模型准确度。因此, 适当的卷积层数对模型的性能表现很重要。但对比 TCNT 模型与 CNNT 模型各自的最优表现发现, 合适的 TCNT 模型 (2 层、通道数为 24-36) 拥有更好的性能表现, 分类准确度为 0.908 3, F1 值为 0.882 4, 精确率为 0.925 9, 召回率为 0.892 5 (表 6)。

表 7 不同卷积层数 CNNT 模型的脉搏波频域特征分类结果

Tab 7 Frequency domain feature classification results of pulse wave based on CNNT models with different convolution layers

Layer	$n=5, \bar{x} \pm s$			
	Accuracy	F1 score	Precision	Recall
1	0.885 4 $\pm$ 0.031 6	0.839 7 $\pm$ 0.043 1	0.904 4 $\pm$ 0.024 6	0.844 2 $\pm$ 0.046 5
2	0.893 8 $\pm$ 0.017 9	0.855 4 $\pm$ 0.026 1	0.922 2 $\pm$ 0.018 6	0.878 9 $\pm$ 0.036 6
3	0.902 1 $\pm$ 0.064 8	0.851 1 $\pm$ 0.143 7	0.894 1 $\pm$ 0.107 1	0.868 6 $\pm$ 0.131 4
4	0.868 8 $\pm$ 0.085 0	0.815 3 $\pm$ 0.136 1	0.840 7 $\pm$ 0.170 1	0.831 0 $\pm$ 0.084 4

CNNT: Convolutional neural network Transformer.

3.2.3 Transformer 结构选择 Transformer 模型由编码器-解码器 (encoder-decoder) 两部分构成,

对于本实验的分类任务只选用 Encoder 部分即可完成。Transformer 原模型使用了多头注意力 (multi-

head attention) 参数为 8 的 6+6 层结构, 使用多头注意力是为了让模型能够关注多个不同的信息子空间, 提高模型的表现。多个注意力能够让模型更好地捕捉到数据中的不同特征, 有助于改善模型的性能, 并帮助网络捕捉更丰富的特征信息。这种方法与在 CNN 中使用多个卷积核的方法类似, 但它更加灵活, 可以在不同的子空间中分别使用不同的权重矩阵进行操作, 从而得到更好的性能提升。

在 TCNT 模型基础上, 本实验排列组合了编码器层数分别为 1、2、3 和多头注意力参数分别为 2、4、6 共 9 种情况下 Transformer 模型的性能表现及各项评价指标, 结果见表 8。随着模型层数和多头注意力参数的增加, 模型性能有所提高, 随后又开

始下降, 这是由于较少层数的模型提取信息能力有限, 难以挖掘脉搏波之间的深层联系, 而层数过多又会陷入信息冗余问题, 很可能会得到一些抽象层次的联系, 而忽略了特征之间的表象信息; 多头注意力参数的影响与之类似, 参数的选择需要从不同方面考虑, 因为层数与多头注意力参数的增加都会明显增加模型的复杂度, 导致计算量增加。但结合表中数据可知, Transformer 模型本身的表现良好, 层数与多头注意力参数对其影响有限, 各组合之间的差异相对较小, 本实验选择层数为 2、多头注意力参数为 4 的 TCNT 模型, 其准确度为 0.912 5, F1 值为 0.883 6, 精确率为 0.922 3, 召回率为 0.901 8, 复杂度在可接受范围内且性能表现最佳。

表 8 不同编码器层数与 head 的 TCNT 模型的脉搏波频域特征分类结果

Tab 8 Frequency domain feature classification results of TCNT model with different encoder layers and heads

$n=5, \bar{x} \pm s$					
Encoder layer	Head	Accuracy	F1 score	Precision	Recall
1	2	0.906 2 $\pm$ 0.026 4	0.875 4 $\pm$ 0.036 6	0.920 6 $\pm$ 0.020 8	0.878 0 $\pm$ 0.043 9
	4	0.908 3 $\pm$ 0.034 5	0.882 4 $\pm$ 0.039 5	0.925 9 $\pm$ 0.025 6	0.892 5 $\pm$ 0.040 6
	6	0.904 2 $\pm$ 0.034 5	0.882 7 $\pm$ 0.044 1	0.918 5 $\pm$ 0.025 9	0.884 7 $\pm$ 0.041 1
2	2	0.904 2 $\pm$ 0.020 2	0.868 7 $\pm$ 0.026 4	0.918 7 $\pm$ 0.021 8	0.877 3 $\pm$ 0.041 4
	4	0.912 5 $\pm$ 0.019 3	0.883 6 $\pm$ 0.022 9	0.922 3 $\pm$ 0.021 3	0.901 8 $\pm$ 0.029 5
	6	0.902 1 $\pm$ 0.031 3	0.866 4 $\pm$ 0.042 3	0.913 6 $\pm$ 0.029 6	0.871 9 $\pm$ 0.050 5
3	2	0.908 3 $\pm$ 0.016 7	0.883 1 $\pm$ 0.027 0	0.917 3 $\pm$ 0.012 7	0.889 9 $\pm$ 0.018 0
	4	0.900 0 $\pm$ 0.016 9	0.860 6 $\pm$ 0.023 8	0.911 1 $\pm$ 0.021 2	0.860 9 $\pm$ 0.027 8
	6	0.904 2 $\pm$ 0.025 9	0.867 2 $\pm$ 0.031 9	0.912 9 $\pm$ 0.021 1	0.858 7 $\pm$ 0.032 6

TCNT: Temporal convolutional network Transformer.

3.2.4 注意力机制 Transformer 模型使用自注意力机制, 为了改善模型性能, 在原有模型的基础上, 本研究选择包括 SE、SK、CBAM、SimAM 和 CA 共 5 种注意力机制进行实验, 讨论了 (a)、(b)、(c) 3 种不同的注意力机制融合策略, 最终得到的分类结果如表 9 所示。由于 SK 维度无法与 TCN 匹配, 故融合方式 (c) 中未进行该组合实验。实验结果表明, 不同的融合方法对不同的注意力机制效果各不相同, 总体来说, 融合方法 (a) 取得了更好的表现, 其中 TCNT+SK 的注意力机制融合方法具有很好的泛化性, 其准确度为 0.931 3, F1 值为 0.912 0, 精确率为 0.934 6, 召回率为 0.912 2, 相较于进行过参数选择的 TCNT 模型, 模型性能也有所提高, 其在准确率上提高了 2.06%, F1 值提升了 3.21%, 精确率提升了 1.33%, 召回率提升了 1.15%。这是由于多个注意力机制的使用使得模型能够从不同方面获取信息, 提取到的关联信息更丰富, 最后再将

多方面信息综合起来, 最终提高了模型性能。

单独使用上述 5 个注意力机制模型所得到的分类实验结果见表 10。SE 通过一个权重矩阵从通道的角度赋予图像不同位置不同的权重, 得到更重要的特征信息。CBAM 可以在通道和空间维度上进行注意力的计算, 保证了其能够作为即插即用的模块集成到现有的网络架构中。CA 将通道注意力分解为 2 个一维特征编码过程, 分别沿 2 个空间方向聚合特征, 这样可以将其互补地应用于输入特征图, 以增强关注对象的表示。SK 在 SE 的基础上进行了改进, 它对输入特征图中的不同大小物体自动选择对应的卷积核来提取特征, 生成通道注意力信息。SimAM 提出了优化能量函数以发掘每个神经元的重要性, 且不会引入额外的参数。由表中数据可见, 单独使用 SimAM 注意力机制得到的分类效果比单独使用 SK 更好, 但是在本实验的融合模型实验中, SK 与自注意力组合使用时模型的性能表

现最佳, 这是因为 SK 能使用不同尺寸的卷积核对输入的特征信息进行特征提取, 然后自动选择合适

的卷积核, 更适合于训练本实验使用的脉搏波频域特征信息。

表 9 不同注意力机制融合模型的脉搏波频域特征分类结果

Tab 9 Frequency domain feature classification results of pulse wave based on different attention mechanism combination models

$n=5, \bar{x} \pm s$					
Combined method	Attention	Accuracy	F1 score	Precision	Recall
(a)	TCNT+SE	0.906 2 $\pm$ 0.032 3	0.871 8 $\pm$ 0.039 5	0.921 5 $\pm$ 0.015 2	0.883 1 $\pm$ 0.022 1
	TCNT+SK	0.931 3 $\pm$ 0.008 3	0.912 0 $\pm$ 0.011 6	0.934 6 $\pm$ 0.008 8	0.912 2 $\pm$ 0.007 9
	TCNT+CBAM	0.929 2 $\pm$ 0.022 2	0.904 6 $\pm$ 0.026 0	0.936 1 $\pm$ 0.022 2	0.908 2 $\pm$ 0.040 6
	TCNT+SimAM	0.910 4 $\pm$ 0.019 3	0.880 5 $\pm$ 0.021 6	0.924 6 $\pm$ 0.012 4	0.889 5 $\pm$ 0.019 5
	TCNT+CA	0.914 6 $\pm$ 0.032 5	0.880 8 $\pm$ 0.036 7	0.928 1 $\pm$ 0.026 9	0.886 3 $\pm$ 0.035 4
(b)	TCNT+SE	0.906 2 $\pm$ 0.013 2	0.882 1 $\pm$ 0.018 7	0.914 2 $\pm$ 0.009 3	0.885 4 $\pm$ 0.024 6
	TCNT+SK	0.910 4 $\pm$ 0.034 0	0.888 5 $\pm$ 0.036 4	0.921 7 $\pm$ 0.029 9	0.904 0 $\pm$ 0.042 2
	TCNT+CBAM	0.902 1 $\pm$ 0.032 0	0.866 0 $\pm$ 0.051 1	0.913 2 $\pm$ 0.030 2	0.876 1 $\pm$ 0.047 9
	TCNT+SimAM	0.872 9 $\pm$ 0.043 9	0.823 9 $\pm$ 0.084 7	0.893 5 $\pm$ 0.027 2	0.822 0 $\pm$ 0.096 2
	TCNT+CA	0.889 6 $\pm$ 0.026 8	0.848 3 $\pm$ 0.039 0	0.896 1 $\pm$ 0.027 0	0.844 9 $\pm$ 0.054 8
(c)	TCNT+SE	0.920 8 $\pm$ 0.015 6	0.893 0 $\pm$ 0.020 7	0.940 5 $\pm$ 0.013 2	0.920 0 $\pm$ 0.019 4
	TCNT+CBAM	0.918 7 $\pm$ 0.019 1	0.885 0 $\pm$ 0.028 5	0.923 0 $\pm$ 0.020 6	0.881 5 $\pm$ 0.040 7
	TCNT+SimAM	0.918 7 $\pm$ 0.029 0	0.890 1 $\pm$ 0.034 2	0.932 3 $\pm$ 0.021 6	0.909 4 $\pm$ 0.024 5
	TCNT+CA	0.916 7 $\pm$ 0.017 4	0.884 7 $\pm$ 0.021 6	0.931 9 $\pm$ 0.012 7	0.901 0 $\pm$ 0.019 1

TCNT: Temporal convolutional network Transformer; SE: Squeeze-and-excitation; SK: Selective kernel; CBAM: Convolutional block attention module; SimAM: A simple parameter-free attention module; CA: Coordinate attention.

表 10 单独注意力机制模型的脉搏波频域特征分类结果

Tab 10 Frequency domain feature classification results of pulse wave based on attention mechanism models alone

$n=5, \bar{x} \pm s$				
Attention	Accuracy	F1 score	Precision	Recall
SE	0.697 9 $\pm$ 0.162 3	0.489 1 $\pm$ 0.304 2	0.558 1 $\pm$ 0.291 3	0.554 8 $\pm$ 0.271 3
CBAM	0.797 9 $\pm$ 0.075 9	0.691 4 $\pm$ 0.128 1	0.793 7 $\pm$ 0.092 7	0.697 3 $\pm$ 0.118 8
CA	0.856 2 $\pm$ 0.106 5	0.786 5 $\pm$ 0.174 3	0.896 1 $\pm$ 0.062 0	0.800 0 $\pm$ 0.177 2
SK	0.904 2 $\pm$ 0.024 1	0.875 5 $\pm$ 0.032 4	0.907 6 $\pm$ 0.026 0	0.874 8 $\pm$ 0.040 1
SimAM	0.914 6 $\pm$ 0.021 2	0.880 3 $\pm$ 0.024 5	0.922 9 $\pm$ 0.016 9	0.890 6 $\pm$ 0.011 0

SE: Squeeze-and-excitation; CBAM: Convolutional block attention module; CA: Coordinate attention; SK: Selective kernel; SimAM: A simple parameter-free attention module.

3.2.5 基于 Floodings 的正则化 在 TCNT-SK 模型使用融合策略 (a) 的基础上, 选择不同正则化方法时融合模型的分类结果见表 11, 模型性能的 5 折交叉验证结果见图 5, 分析数据可知常见的 3 种正则化手段没能改善模型性能, 因为 L1、L2 是通过添加惩罚项来限制模型参数的取值范围, 从而减少过拟合的风险。然而, 如果过度限制模型参数的取值范围可能会导致模型无法很好地适应训练数据, 从而影响模型的准确度。Max 是一种基于约束的正则化方法, 它通过限制参数向量的范数来控制模型的复杂度, 因此可能会阻止模型使用某些参数组合来拟合数据, 导致模型性能降低。而 Floodings 策略通过引入额外的约束假设, 间接控制训练损失值

的下降, 能够提高模型的鲁棒性和泛化性。本实验结果显示, 引入 Floodings 策略的融合模型 TCNT-SK(Fs) 的准确度为 0.939 6, F1 值为 0.924 9, 精确率为 0.940 9, 召回率为 0.929 5, 相较于未加入正则化的 TCNT-SK 模型, 其在准确度上提高了 0.89%, F1 值提升了 1.41%, 精确率提升了 0.67%, 召回率提升了 1.90%。

不使用正则化、使用 Flooding 策略和使用 Floodings 策略 3 种情况下模型训练过程的损失值下降过程见图 6, 可以看出直观正则化手段对模型性能的影响。Floodings 策略通过设置损失值的下限, 阻止了损失值的过快下降, 模型训练过程更加平滑, 最终得到的解也更接近全局最优解。

表 11 选择不同正则化方法的 TCNT-SK 模型的脉搏波频域特征分类结果

Tab 11 Frequency domain feature classification results of TCNT-SK model with different regularization methods

Regularization	$n=5, \bar{x} \pm s$			
	Accuracy	F1 score	Precision	Recall
L1	0.916 7±0.023 8	0.879 8±0.040 4	0.925 2±0.025 4	0.884 6±0.059 4
L2	0.920 8±0.019 3	0.893 8±0.024 1	0.923 8±0.020 6	0.895 1±0.030 9
Max	0.918 7±0.034 5	0.896 3±0.045 4	0.922 7±0.033 3	0.901 6±0.048 1
Flooding	0.933 3±0.021 4	0.915 3±0.024 7	0.938 3±0.019 7	0.919 2±0.022 6
Floodings	0.939 6±0.017 9	0.924 9±0.019 4	0.940 9±0.018 0	0.929 5±0.020 1

TCNT: Temporal convolutional network Transformer; SK: Selective kernel.

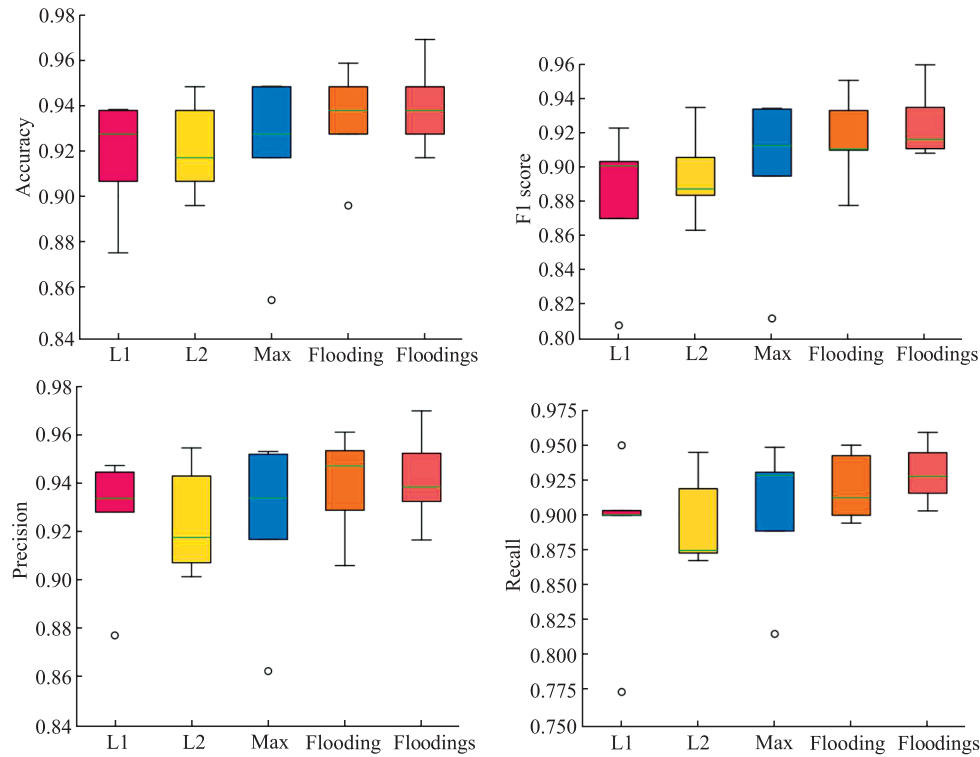


图 5 选择不同正则化方法的 TCNT-SK 模型性能的 5 折交叉验证结果箱线图

Fig 5 Boxplot of 5-fold cross-verification results of TCNT-SK model with different regularization methods

$n=5$. TCNT: Temporal convolutional network Transformer; SK: Selective kernel.

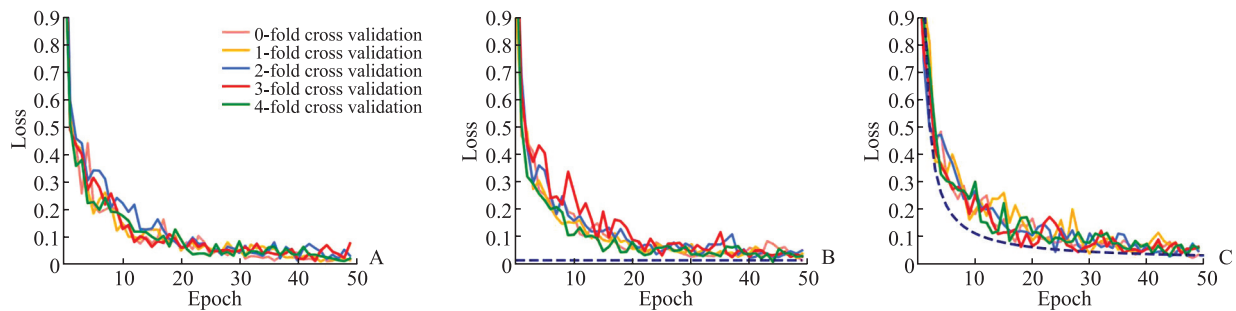


图 6 模型训练过程中损失值的下降

Fig 6 Decrease of loss values during model training

A: Without using regularization; B: Flooding; C: Floodings.

3.3 不同模型对比实验结果 为了验证融合模型 TCNT-SK(Fs) 的分类效果, 本实验对比了 TCNT-SK(Fs) 模型与长短时记忆 (long short-term memory, LSTM)、CNN-LSTM、双边残差网络 (bi-residual network, BRNet)、多尺度残差网

络 (multi-scale residual network, MIRNet) 和密集连接卷积网络 (dense convolutional network, DenseNet) 5 种典型时间序列分类模型及随机森林 (random forest) 和梯度提升决策树 (extreme gradient boosting, XGBT) 2 种机器学习模型的分

类评价指标, 结果见表 12; 使用 5 折交叉验证得到的每一折对应的损失曲线、ROC 曲线和 PR 曲线见图 7。机器学习模型随机森林和 XGBT 难以学习到脉搏波频域特征的深层次关联, 模型精度偏低。典型时间序列分类模型 LSTM 模型的准确率为 0.900 0, F1 值为 0.878 9, 精确率为 0.911 7, 召回率为 0.903 3, 具有较好的泛化性; 而 CNN-LSTM 增加了 CNN 卷积层后表现更好, 这是因为卷积层可以提取具有局部关联关系的脉诊特征, 提高了模型分类速度和准确度; BRNet 虽然采用了 ResNet 基本架构, 但准确度比 TCNT-SK(Fs) 低了 4.64%, 这是因为脉搏波的频域特征是深层特征, BRNet 模型的深度较浅, 无法学习到所有的脉搏波特征信息; MIRNet 模型在 ResNet 模块的基础上加入了 Inception 模块, 模

型深度较深, 准确度比 TCNT-SK(Fs) 低了 4.40%, 但模型结构复杂, 训练所需的资源与时间成本较高; DenseNet 的思路与 ResNet 一致, 但是它建立的是前面所有层与后面层的密集连接, 通过特征在通道维度 Channel 上的连接来实现特征重用, 使得 DenseNet 在参数和计算成本更少的情形下实现比 ResNet 更优的性能, 但相较于 TCNT-SK(Fs) 模型其准确度低了 2.66%, F1 值低了 5.43%。与 LSTM、BRNet、MIRNet、DenseNet 模型相比, 本实验的 TCNT-SK(Fs) 模型的 AUC 值分别增加了 4.81%、5.33%、2.43%、5.78%, 因此, 对于高血压患者危险分层数据, TCNT-SK(Fs) 模型可以提取关联性较强的深层特征, 具有较高的分类精度和较好的泛化性能。

表 12 不同模型对高血压患者危险分层的评价指标

Tab 12 Evaluation indicators of different models for risk stratification of hypertension patients

<i>n=5, $\bar{x} \pm s$</i>					
Model	Accuracy	F1 score	Precision	Recall	AUC
RF	0.544 5 $\pm$ 0.025 1	0.434 7 $\pm$ 0.021 8	0.394 5 $\pm$ 0.032 3	0.328 4 $\pm$ 0.020 6	0.691 7 $\pm$ 0.016 1
XGBT	0.565 4 $\pm$ 0.008 0	0.428 0 $\pm$ 0.006 3	0.430 9 $\pm$ 0.089 1	0.332 8 $\pm$ 0.008 8	0.709 5 $\pm$ 0.010 6
LSTM	0.900 0 $\pm$ 0.032 0	0.878 9 $\pm$ 0.039 5	0.911 7 $\pm$ 0.028 6	0.903 3 $\pm$ 0.039 9	0.947 8 $\pm$ 0.041 9
CNN-LSTM	0.918 7 $\pm$ 0.034 5	0.901 2 $\pm$ 0.035 8	0.927 4 $\pm$ 0.029 5	0.904 1 $\pm$ 0.029 3	0.918 1 $\pm$ 0.038 0
BRNet	0.897 9 $\pm$ 0.036 3	0.865 9 $\pm$ 0.044 8	0.910 9 $\pm$ 0.023 5	0.879 3 $\pm$ 0.032 6	0.943 1 $\pm$ 0.048 4
MIRNet	0.900 0 $\pm$ 0.029 5	0.872 1 $\pm$ 0.029 9	0.932 1 $\pm$ 0.014 4	0.904 8 $\pm$ 0.021 3	0.969 8 $\pm$ 0.035 0
DenseNet	0.914 6 $\pm$ 0.043 4	0.874 7 $\pm$ 0.106 9	0.932 7 $\pm$ 0.023 2	0.892 4 $\pm$ 0.097 8	0.939 1 $\pm$ 0.056 4
TCNT-SK(Fs)	0.939 6 $\pm$ 0.017 9	0.924 9 $\pm$ 0.019 4	0.940 9 $\pm$ 0.018 0	0.929 5 $\pm$ 0.020 1	0.993 4 $\pm$ 0.003 3

RF: Random forest; XGBT: Extreme gradient boosting; LSTM: Long short-term memory; CNN: Convolutional neural network; BRNet: Bi-residual network; MIRNet: Multi-scale residual network; DenseNet: Dense convolutional network; TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Floodings); AUC: Area under curve.

4 讨论

4.1 模型复杂度 不同模型训练完整的 5 折交叉验证的模型参数量 (parameter)、计算量 (floating point operations) 和运行时间 (running time) 见表 13。机器学习模型随机森林和 XGBT 模型简单, 运行较快, 但性能表现一般。LSTM 模型运行时间较短, 参数量适中, 在增加 CNN 卷积层后, 由于 CNN 具有特征提取和降维作用, 减少了 LSTM 的参数量和运行时间, 改善了模型性能。BRNet 网络相对简单, 运行时间较短。而 MIRNet 和 DenseNet 网络深度较深, 模型复杂度高, 参数量、计算量和运行时间都大大增加, 模型综合性能一般。TCNT-SK(Fs) 模型在 5 折交叉验证中的模型复杂度下, 获得了较好的性能表现, 是权衡资源与性能后的更优

选择。

4.2 特征贡献率 特征贡献率是衡量每个输入特征对模型预测结果影响的指标。贡献率越大, 特征越重要, 对分类结果的影响也就越大。本实验采用 XGBT 算法分析 36 个脉搏波频域 MFCC 特征, 计算出各特征的贡献率, 并绘制可视化图 (图 8)。其中二阶差分的 S3 和 S12 两个特征贡献率均超过了 4%, 但总体来看, 脉搏波的静态特征、一阶差分和二阶差分系数的贡献率相对均衡, 各个特征对分类结果都具有重要的影响, 说明高血压危险程度不仅与脉搏波的静态特征相关, 也应当考虑脉搏波的动态特征。因此, 应积极寻找脉搏波静态与动态特性的物理意义, 并在临床实践中加以应用, 将有助于更准确地诊断和治疗。

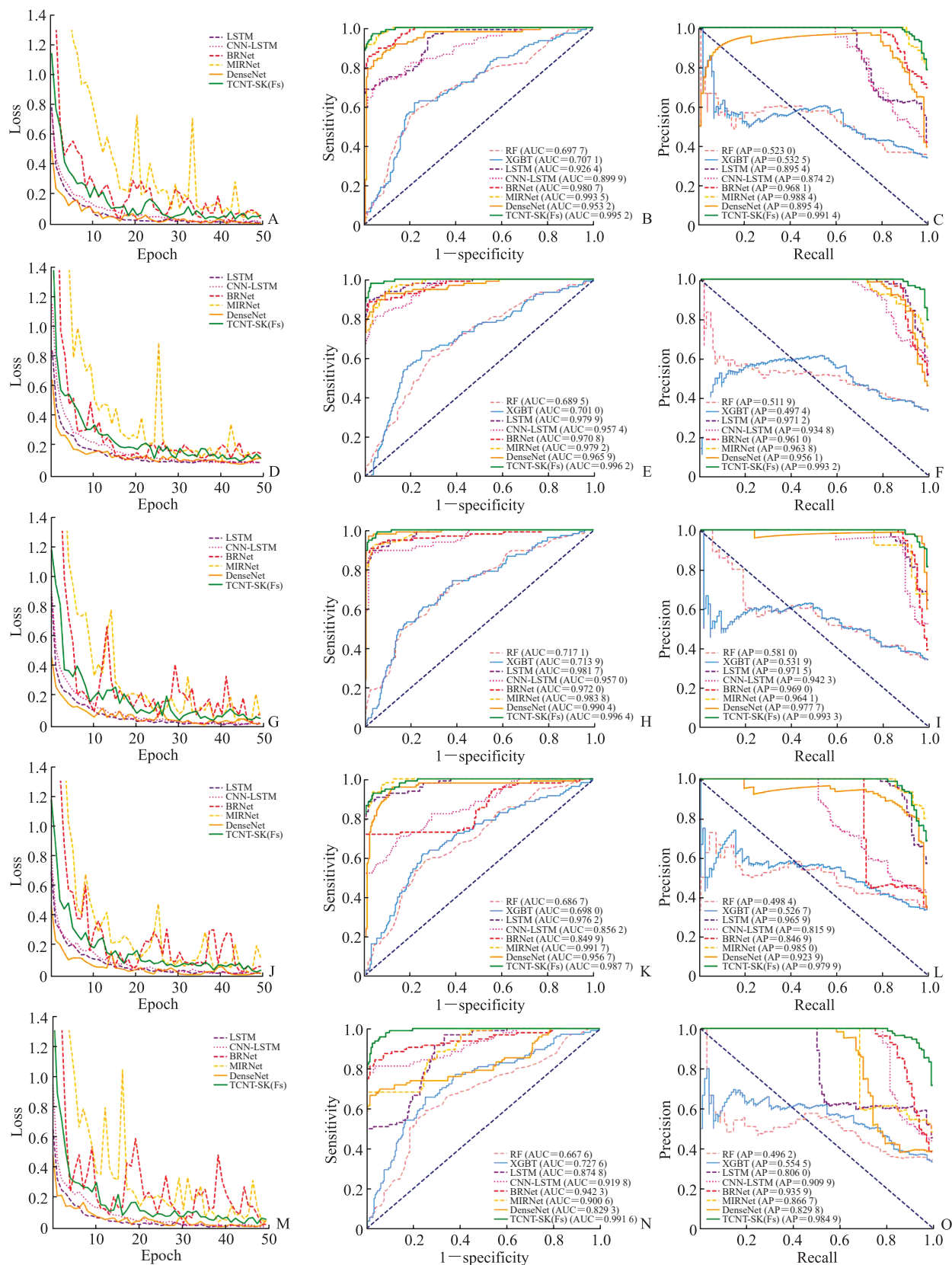


图7 不同模型5折交叉验证的损失(A、D、G、J、M)、ROC(B、E、H、K、N)和PR(C、F、I、L、O)曲线
Fig 7 Loss (A, D, G, J, M), ROC (B, E, H, K, N) and PR (C, F, I, L, O) curves of 5-fold cross-verification of different models
 A-C: 0-fold cross-validation; D-F: 1-fold cross-validation; G-I: 2-fold cross-validation; J-L: 3-fold cross-validation; M-O: 4-fold cross-validation. ROC: Receiver operating characteristic; PR: Precision-recall; LSTM: Long short-term memory; CNN: Convolutional neural network; BRNet: Bi-residual network; MIRNet: Multi-scale residual network; DenseNet: Dense convolutional network; TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Flooding); RF: Random forest; XGBoost: Extreme gradient boosting; AUC: Area under curve; AP: Average precision.

表 13 不同模型的参数量、FLOP 和运行时间

Tab 13 Parameters, FLOP, and running time of different models

Model	Parameter/byte	FLOP ($\times 10^9$)	Running time/s
RF			6.6
XGBT			35.9
LSTM	97 107	0.072	55.6
CNN-LSTM	60 423	0.077	41.4
BRNet	86 473	0.008	45.0
MIRNet	8 088 418	29.222	4 821.2
DenseNet	9 841 859	12.590	1 198.6
TCNT-SK(Fs)	526 414	0.542	346.6

RF: Random forest; XGBT: Extreme gradient boosting; LSTM: Long short-term memory; CNN: Convolutional neural network; BRNet: Bi-residual network; MIRNet: Multi-scale residual network; DenseNet: Dense convolutional network; TCNT-SK(Fs): Temporal convolutional network Transformer-selective kernel (Flooding); FLOP: Floating point operation.

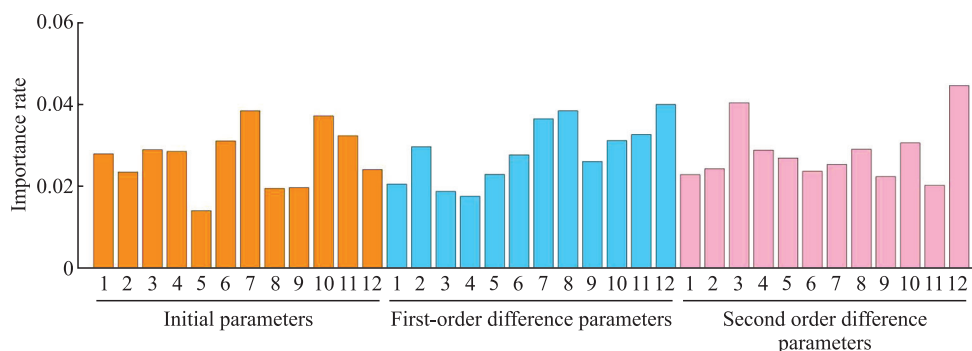


图 8 基于 XGBT 算法的 36 个脉搏波频域 MFCC 特征贡献率

Fig 8 Contribution rates of 36 MFCC features in the frequency domain of pulse waves based on XGBT algorithm

XGBT: Extreme gradient boosting; MFCC: Mel-scale frequency cepstral coefficient.

[参考文献]

[1] 潘禹安, 宋佳男, 张鹭鹭, 等. 我国老年高血压患者抑郁状况及其影响因素分析——基于第 8 轮 CLHLS 数据[J]. 海军军医大学学报, 2023, 44(10): 1173-1178. DOI: 10.16781/j.cn31-2187/R.20230398.

PAN Y, SONG J, ZHANG L, et al. Depression and its influencing factors among elderly hypertensive patients in China: based on the 8th round CLHLS data[J]. Acad J Naval Med Uni, 2023, 44(10): 1173-1178. DOI: 10.16781/j.cn31-2187/R.20230398.

[2] 国家卫生健康委员会. 中国卫生健康统计年鉴 (2021)[M]. 北京: 中国协和医科大学出版社, 2021: 3.

[3] WANG Z, CHEN Z, ZHANG L, et al. Status of Hypertension in China: Results From the China Hypertension Survey, 2012—2015[J]. Circulation, 2018, 137(22): 2344-2356. DOI: 10.1161/CIRCULATIONAHA.117.032380.

[4] 崔佳嘉, 马宏忠. 基于改进 MFCC 和 3D-CNN 的变压器铁心松动故障声纹识别模型[J]. 电机与控制

5 结论

本研究提出了一种使用深度学习融合模型对中医脉波信号进行分析, 从而实现对高血压患者进行危险分层的方法。该模型采用不同注意力机制决策融合方法, 可以有效地提取脉诊数据频域特征的深层次关联, 具有较高的分类性能。通过实验验证, 与典型分类模型相比, 该模型具有较高的分类精度和泛化性能, 准确率为 0.939 6, F1 值为 0.924 9, 精确率为 0.940 9, 召回率为 0.929 5, AUC 值为 0.993 4。该模型为中医脉诊对高血压患者进行风险分层研究提供了一种有效方法, 为高血压层级诊断提供了新思路。

学报, 2022, 26(12): 150-160. DOI: 10.15938/j.emc.2022.12.015.

[5] ZHANG S R, SUN Q F. Human pulse recognition based on convolutional neural networks[C]//2016 International Symposium on Computer, Consumer and Control (IS3C). Xi'an, China. IEEE, 2016: 366-369. DOI: 10.1109/IS3C.2016.101.

[6] 杨晶东, 陈磊, 蔡书琛, 等. 混合深度学习机制下的 H 型高血压脉诊预测[J]. 数据采集与处理, 2022, 37(4): 883-893. DOI: 10.16337/j.1004-9037.2022.04.016.

[7] 马明艳, 陈伟, 吴礼发. 基于 CNN_BiLSTM 网络的入侵检测方法[J]. 计算机工程与应用, 2022, 58(10): 116-124. DOI: 10.3778/j.issn.1002-8331.2102-0062.

[8] 杨如民, 雷倩, 余成波. 基于蝙蝠算法-随机森林分类算法和脉搏波的精神疲劳检测算法[J]. 科学技术与工程, 2022, 22(4): 1495-1501. DOI: 10.3969/j.issn.1671-1815.2022.04.025.

[9] 张选, 胡晓娟. 基于 GoogLeNet 和 ResNet 的深度融合神经网络在脉搏波识别中的应用[J]. 计算机系统应用, 2019, 28(10): 15-26. DOI: 10.15888/j.cnki.

- csa.007110.
- [10] 杨浩, 黄茂林, 蔡志鹏, 等. 融合 CNN 和 BiLSTM 的心律失常心拍分类模型[J]. 中国生物医学工程学报, 2020, 39(6): 719-726. DOI: 10.3969/j.issn.0258-8021.2020.06.009.
- [11] OUYANG C, GAN Z, ZHEN J, et al. Inter-patient classification with encoded peripheral pulse series and multi-task fusion CNN: application in type 2 diabetes[J]. IEEE J Biomed Health Inform, 2021, 25(8): 3130-3140. DOI: 10.1109/JBHI.2021.3061114.
- [12] CHEN J, HUANG H, HAO W, et al. A machine learning method correlating pulse pressure wave data with pregnancy[J]. Int J Numer Method Biomed Eng, 2020, 36(1): e3272. DOI: 10.1002/cnm.3272.
- [13] YAN J, CAI X, ZHU G, et al. A non-invasive blood pressure prediction method based on pulse wave feature fusion[J]. Biomed Signal Process Contr, 2022, 74: 103523. DOI: 10.1016/j.bspc.2022.103523.
- [14] 苏志刚, 吕江波, 郝敬堂. 基于平滑先验法去除脉搏波基线漂移[J]. 中国医学物理学杂志, 2018, 35(10): 1197-1202. DOI: 10.3969/j.issn.1005-202X.2018.010.017.
- [15] 张伟. 基于卷积神经网络的轴承故障诊断算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [16] ISHIDA T, YAMANE I, SAKAI T, et al. Do we need zero training loss after achieving zero training error? [EB/OL]. 2020: arXiv: 2002.08709. <http://arxiv.org/abs/2002.08709>.
- [17] 张冬雨. 面向脉诊的脉搏信号与血流信号分类研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
- [18] LIU T T, FRASER-SMITH A C. Detection of transients in 1/f noise with the undecimated discrete wavelet transform[J]. IEEE Trans Signal Process, 2000, 48(5): 1458-1462. DOI: 10.1109/78.839991.
- [19] MIN L, CUI Q, JIN Z, et al. Inhomogeneous image segmentation based on local constant and global smoothness priors[J]. Digit Signal Process, 2021, 111: 102989. DOI: 10.1016/j.dsp.2021.102989.
- [20] NGUYEN H P, BARALDI P, ZIO E. Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants[J]. Appl Energy, 2021, 283: 116346. DOI: 10.1016/j.apenergy.2020.116346.
- [21] 吴全玉, 张文强, 潘玲佼, 等. 一种结合自适应噪声完备经验模态分解和盲反卷积去除脑电中眼电伪迹的新方法[J]. 数据采集与处理, 2020, 35(4): 720-729. DOI: 10.16337/j.1004-9037.2020.04.013.
- [本文编辑] 杨亚红