

# 基于时序卷积网络的早期帕金森多模态检测系统

周希武<sup>1</sup>, 杨明昭<sup>2</sup>, 胡殿雷<sup>2,3</sup>

- 淮安市第二人民医院 信息科, 江苏 淮安 223001;
- 徐州医科大学 医学信息与工程学院, 江苏 徐州 221000;
- 徐州医科大学 基础医学院, 江苏 徐州 221000

**摘要:** 帕金森病是最常见的神经退行性疾病之一, 其临床特征与其他神经退行性疾病有重叠, 且缺乏明确的病理机制, 导致早期诊断检测困难、误诊率高等问题; 为了研究有效的早期帕金森病检测方法, 深入探索帕金森病发展的时间特征规律, 并提高早期帕金森病预测、分析和诊断决策的准确性, 设计了一种基于时序卷积网络的早期帕金森病多模态检测系统, 为及时发现早期帕金森病提供辅助诊断依据; 该系统利用语音、步态和受试者自测数据, 采用多元线性池化方法进行多模态融合, 结合时间卷积网络和参数共享方式, 以提高系统的检测精度并降低过拟合风险; 实验测试结果显示, 基于时序卷积网络的早期帕金森病检测系统的准确率达到 96.22%, 在多项评估指标上优于传统的帕金森检测模型, 展现出良好的早期帕金森联合检测效果。

**关键词:** 帕金森; 时序卷积网络; 线性池化; 多模态; 过拟合

## Early Parkinson's Multimodal Detection System Based on Temporal Convolutional Networks

ZHOU Xiwu<sup>1</sup>, YANG Mingzhao<sup>2</sup>, HU Dianlei<sup>2,3</sup>

- Department of Information, Second People's Hospital of Huai'an City, Huai'an 223001, China;
- School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou 221000, China;
- School of Basic Medicine, Xuzhou Medical University, Huai'an 221000, China)

**Abstract:** Parkinson's disease is one of the most common neurodegenerative diseases, its clinical features overlap with other neurodegenerative diseases and lack a clear pathological mechanism, leading to difficulties in early diagnosis and high misdiagnosis rates; In order to study effective early detection methods for Parkinson's disease, deeply explore the temporal characteristics of Parkinson's disease development, and improve the accuracy of early Parkinson's disease prediction, analysis, and diagnostic decision-making, an early Parkinson's disease multimodal detection system based on temporal convolutional networks is designed, providing an auxiliary diagnostic basis on the timely detection of early Parkinson's disease; The system utilizes the speech, gait, and subject self-test data, adopts the multiple linear pooling method for multimodal fusion, and integrates the time convolutional networks and parameter sharing to improve the detection accuracy of the system and reduce overfitting risks; Experimental results show that the accuracy of the early Parkinson's disease detection system based on temporal convolutional networks reaches 96.22%, which is superior to traditional Parkinson's detection models in multiple evaluation indicators and demonstrates good early Parkinson's joint detection performance.

**Keywords:** parkinson's disease; temporal convolutional network; linear pooling; multimodal; overfitting

## 0 引言

帕金森病 (PD, Parkinson's disease), 又称震颤麻痹, 是一种在中老年时期常见的神经系统退行性疾病<sup>[1-2]</sup>。其主要特征包括静止性震颤、行动迟缓、肌强直和姿势平衡障碍。至今, PD 的确切病因仍然不明, 可能受到遗传、环境、年龄和氧化应激等多种因素的影响<sup>[3]</sup>。据统计数据显示, 65 岁以上人群中 PD 的患病率大约在 1%~2%, 而 85 岁以上人群则高达 3%~5%<sup>[4]</sup>。在中国, 65 岁以上人群中 PD 的患病率约为 1.7%, 已成为影响中老年健康的主要疾

病之一<sup>[5]</sup>。截至目前, 中国 PD 患者数量已超过 300 万, 位居全球首位<sup>[6]</sup>。随着老龄化社会问题的加剧, 国内 PD 患者数量将快速增长, 且病情日益严重, 导致 6 年死亡率高达 66%<sup>[7]</sup>。然而, PD 的致病机制尚不清楚, 早期症状隐匿, 且与其他神经性疾病存在相似之处, 如癔症性、紧张性、老年性震颤等, 这使得 PD 早期诊断变得更加复杂<sup>[8]</sup>。

有关 PD 早期诊断的大量研究主要包括临床症状观察、体格检查、神经影像学检查、血液检测、生物标志物检测和基因检测等手段<sup>[9]</sup>。此类 PD 诊断方式因流程复杂、早期

收稿日期: 2024-04-15; 修回日期: 2024-05-07。

基金项目: 江苏省高等学校基础科学(自然科学)研究重大项目(22KJA120002); 徐州市科技计划项目(KC21182); 徐州市科技计划项目(KC22224)。

作者简介: 周希武(1982-) 男, 大学本科, 高级工程师。

引用格式: 周希武, 杨明昭, 胡殿雷. 基于时序卷积网络的早期帕金森多模态检测系统[J]. 计算机测量与控制, 2024, 32(6): 71-77.

症状并不典型、诊断成本相对较高,且部分偏远地区难以获取先进的 PD 诊断资源,导致 PD 临床诊断的综合诊断准确率仅为 80%<sup>[10]</sup>,而早期诊断准确率仅为 53%,严重阻碍了早期 PD 患者的诊断进程,导致病情向中晚期转化,并直接威胁患者生命安全。为探索低成本且高精度的 PD 诊断方法,大量临床研究发现,言语障碍特征和步态时空数据特征在早期 PD 患者临床表现突出,包括音量降低、音质改变、韵律障碍、重音消失、运动减少、动作缓慢、肌强直和静止性震颤等<sup>[11-12]</sup>。基于此,众多学者开展了基于语音障碍和步态数据特征的 PD 早期诊断研究。文献 [13] 设计基于多头注意力机制的卷积神经网络,使用语音数据和步态数据实现 PD 多模态远程检测,准确率达 91.3%。文献 [14] 开展了基于语音的 PD 分型识别方法研究,设计了基于图边距的特征选择算法和多标签 k 近邻分类算法结合的多标签分类方法,对震颤、冻结步态和吞咽困难三种 PD 分型的分类预测准确率达到 97%。文献 [15] 开展了基于语音诊断 PD 的算法研究,从语音数据采集分析和分类模型优化入手,实现了帕金森病的分类。文献 [16] 提出了使用机器学习和深度学习模型来分析患者语音信号,实现帕金森病的诊断分类,准确率接近 92%。文献 [17] 提出一种基于 Chirlet 变换的 PD 诊断方法,采用语音信号诊断 PD,并取得了良好的检测效果。文献 [18] 使用步态数据,借助小波变换开展了 PD 的诊断研究,提出了由多尺度特征和时域特征相结合的预测方法,准确率达 88.89%。

在 PD 的早期诊断中,语音、步态等数据均表现为时间序列信号的形式,这些数据在时间维度上呈现出一定的相关性,这一特点已在多项研究中得到证实<sup>[19-20]</sup>。通过深入剖析这些数据的时间相关性,能够更好地理解 PD 发展的时间特征规律,从而为预测、分析和决策过程提供有益的辅助信息。然而,时间序列数据通常具有数据量大、维度高以及结构复杂等特点,这使得基于时间序列数据的研究相比其他类型的研究更具挑战性<sup>[21]</sup>。目前,长短期记忆网络(LSTM, long short term memory)是时间序列预测中常用的方法,但它也存在一些明显的不足。首先,LSTM 的计算复杂度较高。其次,由于其引入了门控机制和长期记忆机制,使得模型需要更多的参数和计算量,导致模型的可解释性相对较差。此外,LSTM 还存在泛化能力不足的问题,这在一定程度上限制了其在 PD 早期诊断中的应用。针对上述问题,本文引入时序卷积网络(TCN, temporal convolutional network),设计了一种早期 PD 多模态检测系统,提高了早期 PD 的检测准确率,能够很好地辅助医生对于帕金森疾病的诊断。本文主要创新点和贡献如下:

设计明确的多模态特征提取方案,制定统一的特征数据标准规范,为多模态数据特征融合与 PD 早期诊断提供高质量多模态特征数据支持。

针对单一模态数据导致的 PD 检测准确率低和干扰性强问题,设计多元线性池化方法,融合语音、步态和患者自测多模态数据,并基于模态内和模态间的相关交互结果对合并的特征进行线性拼接,从而获取多模态特征的整体表示。

设计基于 TCN 的 PD 诊断模型,减少需要学习的参数数量,降低 PD 检测模型过拟合概率,并提升处理高维数据的效率。

## 1 系统设计

基于 TCN 的早期帕金森病检测系统整体框架如图 1 所示。

本系统由 3 部分组成,第一层为数据层,主要用于采集、传输和存储语音、步态与患者自测数据;第二层为处理层,主要处理用于检测 PD 的语音、步态和自测多模态数据中的噪声数据和干扰信息,以便于提取和融合三模态特征,并便于智能诊断模型识别数据特征;第三层为应用层,主要进行早期 PD 的多模态诊断。为保证基于 TCN 的早期 PD 多模态检测系统的可用性与有效性,为临床诊断 PD 提供辅助决策支持,本系统的设计应重点实现以下几点目标:

1) 独立性:系统各模块应能独立运行,避免因个别模块异常而导致系统整体瘫痪,造成数据外泄或 PD 智能检测服务终止等问题。

2) 鲁棒性:系统应建立完整、闭环化的威胁检测、故障定位、容灾备份和恢复机制,应能及时监测和预警外部网络威胁、内部应用程序威胁和非法操作行为等,确保 PD 多模态检测服务连续性与系统持续运行稳定性。

3) 智能化:系统应具备高度自主学习能力和精准分析分类能力、实时反馈调整和多模态数据整合能力,能够从大量的多模态检测数据中提取有效信息,并进行动态自适应的诊断模型优化。

## 2 系统软件设计

该系统以帕金森患者的语音和步态客观数据为主,融合帕金森病自测主观文本数据,来实现更加全面、精准、

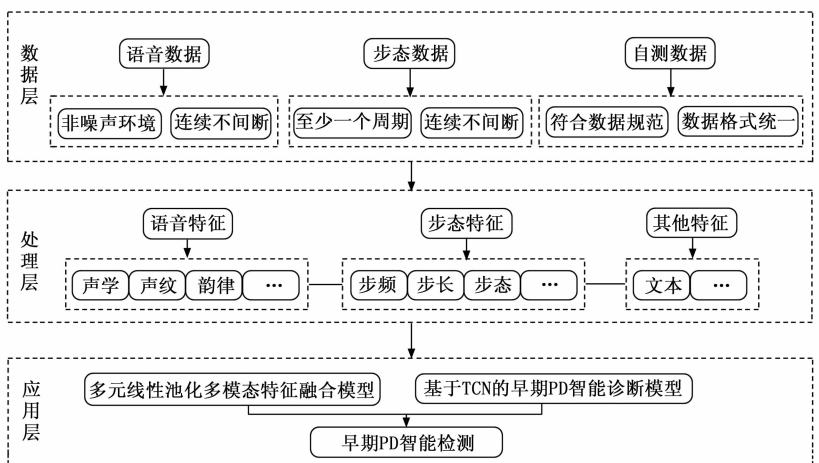


图 1 基于 TCN 的早期 PD 检测系统框架

科学的 PD 早期检测。为提升早期 PD 筛查检测准确率, 降低单模态数据检测导致的检测准确率低等问题, 设计基于多元线性池化的多模态特征融合方法, 融合语音、步态和自主测试三模态数据特征。其次引入 TCN, 设计基于 TCN 的早期 PD 智能检测算法, 并实现低成本、高效率和高精度的 PD 早期检测。

为解决早期 PD 的时序数据诊断存在的多模态融合、低精度预测等问题, 本文设计了一种基于 TCN 的早期 PD 多模态检测模型, 如图 2 所示。首先考虑语音、步态和受试者自测数据, 采用多元线性池化方法进行多模态特征融合。其次引入 TCN 进行 PD 分类训练, 避免常见 LSTM 方法存在的高过拟合风险。在实际应用中, 本地客户端接收并下载 TCN 模型, 通过本地数据集进行模型训练。

## 2.1 系统设计的关键技术及算法

### 2.1.1 线性池化

线性池化主要用于帕金森数据多模态特征融合, 以双线性池化<sup>[22]</sup>为例, 通过两个结构相同的网络对同一样本提取出特征  $x, y$  进行双线性融合, 得到两个特征融合后的向量, 并将其用于分类。一个双线性模型通常包含 4 个部分,

特征提取函数  $f_A$  和  $f_B$ , 池化函数  $P$  和分类函数  $C$ ,  $Bilinear = (f_A, f_B, P, C)$ 。双线性池化模型定义为:

对于图像  $I$  在位置  $L$  的两个特征  $f_A(L, I) \in \mathbb{R}^{T \times D}$  和  $f_B(L, I) \in \mathbb{R}^{T \times E}$ , 采用一个双线性操作将  $f_A(L, I)$  和  $f_B(L, I)$  得到的两组特征在每一空间位置  $L$  按照矩阵的外积进行组合, 得到双线性特征  $B_j$ , 见公式 (1):

$$B_j(L, I, f_A, f_B) = f_A^T(L, I) f_B(L, I) \quad (1)$$

其中:  $L$  涵盖位置和尺度。池化函数  $P$  通过整合操作, 将所有位置的双线性特征进行融合, 得到原始图像的全局特征, 并将其转化为双线性向量  $\Phi(I)$ , 计算公式如 (2) 所示:

$$\Phi(I) = \sum_L B_j(L, I, f_A, f_B) \quad (2)$$

接着对双线性特征向量  $x = \text{vec}(\Phi(I))$  进行平方根变换  $y = \text{sign}(x) \sqrt{|x|}$ , 并增加  $L_2$  标准化处理, 最终得到融合后的特征  $z$ , 如公式 (3), 可用于细粒度的分类:

$$z = \frac{y}{\|y\|_2} \quad (3)$$

### 2.1.2 时序卷积网络

TCN<sup>[23]</sup>的提出是为了卷积神经网络 (CNN, convolu-

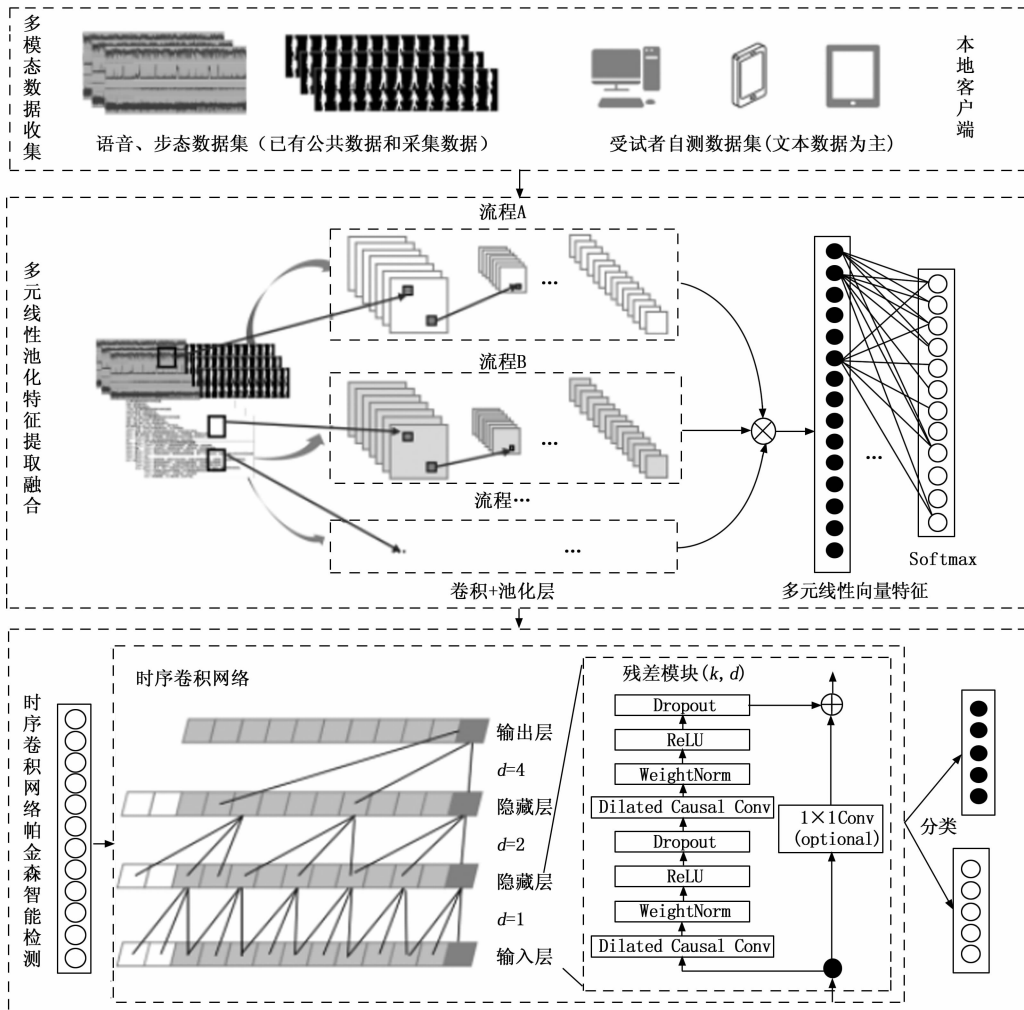


图 2 基于 TCN 的早期 PD 检测流程

tional neural network) 具备时序特性。与一般 CNN 不同, TCN 的卷积为因果关系, 任意时刻的状态均考虑之前的全部状态。假设  $Q$  表示一维卷积核的数量,  $k$  表示每个卷积核的大小,  $T_i$  表示单位序列长度,  $G$  表示输入序列特征维度。经过一维卷积得到  $T_i \times Q$  时序特征序列,  $ReLU$  作为激活函数。

1) 因果卷积 (Casual Convolution): 因果卷积适合处理具有时间依赖性的语音、文本和视频等数据。因果卷积模型是单向结构, 无法看到未来的数据, 对于上一层  $t$  时刻的值, 只依赖于下一层  $t$  时刻及其之前的值, 需要根据  $x_1, x_2, \dots, x_{t-1}$  和  $y_1, y_2, \dots, y_{t-1}$  去预测  $y_t$  的值, 使得  $y_t$  接近于实际值, 如式 (4) 所示:

$$p(x) = \prod_{i=1}^T p(x_i | x_1, \dots, x_{i-1}) \quad (4)$$

其中:  $x_t$  表示  $t$  时刻的信息。因果卷积网络结构的每一层输出都由前一层对应位置的输入及其前一个位置的输入共同得到。当输出层与输入层间有较多隐藏层时, 输出对应的所有输入就越多, 且输入和输出离得越远。此时需要考虑更早之前的输入变量参与运算。这会增加因果卷积网络的层数, 会造成梯度消失、训练复杂以及拟合效果差的问题。为解决此问题, 出现了膨胀卷积。

2) 膨胀卷积 (Dilated Convolution): 作为因果卷积的一种扩展, 通过向卷积核中引入扩张率来使卷积核的感受野在时间维度更加广泛。膨胀卷积允许卷积时的输入存在间隔采样, 采样率受  $d$  控制 (见图 2 时序卷积网络部分), 最下层  $d=1$  表示输入时每个点都采样, 中间层  $d=2$  表示每 2 个点采样一个输入。一般情况, 层级越高, 使用的  $d$  越大。因此, 膨胀卷积使得有效窗口的大小随着层数呈指数型增长。普通卷积感受野计算公式如 (5) 所示:

$$l_k = l_{k-1} + [(g_{k-1}) * \prod_{i=1}^{k-1} s_i] \quad (5)$$

其中:  $l_k$  表示第  $k$  层的感受野大小,  $g_k$  表示第  $k$  层的卷积核大小,  $s_i$  表示第  $i$  层的步长。

3) 残差模块 (Residual Block): 是训练深层网络的有效方法, 使网络可以以跨层的方式传递信息, 每个残差模块包含两层卷积和非线性映射, 在每层中加入 WeightNorm 和 Dropout 来正则化网络。在 TCN 中应用了残差模块。

### 2.2 多模态特征提取

1) PD 语音数据特征提取: 使用 Python 工具包 librosa 从完整的音频数据中获取声纹、韵律、声学等语音信号特征信息, 在预加重、分帧和加窗处理后, 使用 STFT 和 Mel 滤波器获得频谱图。此外, 对所有的语音时长分布进行分析, 并通过补齐或裁剪将时间维度设置为 1 600 左右。

2) PD 步态数据特征提取: 采用每个时间步的四元数值确定的 3D 旋转矩阵来投影时间坐标系中的数据点。对于每个数据点, 提取了连续 2 500 个具有最大加速度幅值的时间步长。对于不足 2 500 个时间步长的数据点, 在末尾用 0 加速度填充。此外, 实施数据增强以防止过拟合, 通过应用一系列保留原始数据核心组件的转换生成一组新的数据

点。对于每个数据点, 应用 5 次转换, 生成 5 倍以上的数据点。最后, 采用窗口化短时傅里叶变换捕获频率空间中数据的动态特征。

3) 受试者自测数据特征提取: 数据通过患者在 mPower 应用程序进行敲击测试收集。包括游戏分数、游戏次数、失败次数、开始时间戳、结束时间戳、每个已完成游戏的信息、点击样本以及完成游戏每个步骤的时间间隔和用药时间点等多种数据。这些数据最终整理成 JSON 格式的文件。本研究将对每个阶段的分数值与总体分数进行量化和归一化操作。

### 2.3 基于多元线性池化的特征融合方法设计

多模式特征融合在方法上主要分为特征层融合和决策层融合两种方式。目前, 大部分研究侧重于特征层融合, 旨在捕获并保留更多的交互特征。在这些方法中, 双线性汇聚方法占据主导地位。然而, 这些方法在处理过程中通常忽视了不同模态之间的相关性。因此, 本研究提出了一种多元线性汇聚融合方法 (MPF, multilinear pooling fusion), 旨在同时探索模态内部和模态之间的相关性。首先, 整合每个模态的特征; 随后, 基于模态内部和模态间的交互结果, 对整合后的特征进行线性拼接, 从而实现全面的多模态特征表示, 如图 3 所示。

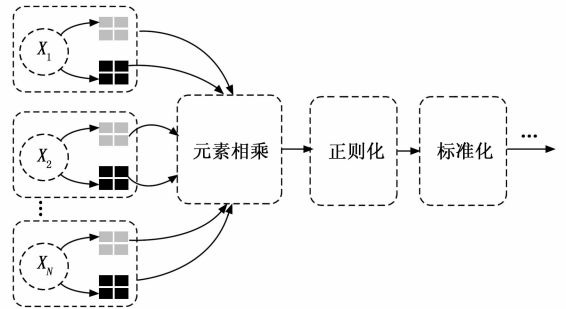


图 3 多元线性池化方法

定义  $\{X_1, X_2, \dots, X_N\}$  ( $x_i \in Z^d$ ), 表示为多模态特征, 其中  $x_i \in Z^d$  表示  $x_i$  是其中一个模态。对语音、步态和自测数据间的参数化交互进行编码, 即集成特征:

$$F = (((((T \times 1x) \times 2x) \times 3x) \times 4y) \times 5y) \times 6y) \quad (6)$$

$F$  表示融合的特征,  $T [T \in Z^d \times d_x \times d_y \times d_f]$  为需要学习的张量。 $\times_i$  表示张量和矩阵的第  $i$  个模态的乘积结果操作。 $F$  的计算过程如公式 (7) 所示:

$$F_i = x^T T[:, :, i] y \quad (7)$$

$T[:, :, i] y$  表示最后一个维度的切片操作。由于参数规模大, 会出现过拟合问题, 导致计算成本过高。因此, 本文使用低秩约束解决过拟合问题。集合  $N$  个矩阵  $[ZP_1, ZP_2, \dots, ZP_N]$  或  $[ZQ_1, ZQ_2, \dots, ZQ_N]$ , 其与  $F$  的映射积为:

$$F = \sum_{n=1}^N (ZP_n^T x) \circ (ZQ_n^T y) \quad (8)$$

从列的角度展平即得到最终的矩阵, 如公式 (9) 所示:

$$F = \text{Sumpooling}[(ZP_1^T x) \circ (ZQ_1^T y)]$$

$$(ZP_2^T x) \circ (ZQ_2^T x) \circ (ZP_3^T x) \circ (ZQ_3^T x), N] \quad (9)$$

*Sum pooling* 表示一维非重叠窗口在  $x$  向量上的运行和池化的结果。公式 (9) 表示三模态的特征向量融合后的结果。

## 2.4 基于 TCN 的早期 PD 诊断模型设计

在 PD 诊断中, 使用的语音、步态等数据均为时间序列信号数据, 具有体量大、维度高、结构复杂等特点。目前, 常用的时间序列预测方法为 LSTM, 但该方法存在计算复杂度高以及可解释性差等问题。因此, 本文引入 TCN。TCN 的输出长度与输入长度相同, 主要步骤包括扩张因果卷积和平行残差连接。

扩张因果卷积的概念是在传统卷积操作中引入洞 (Dilation), 通过间隔采样跳过一些输入, 以增加感受野的大小。这样可以利用更小的卷积核和更少的层来建模长时间序列依赖性, 如图 4 所示。

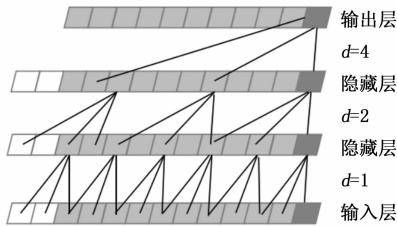


图 4 TCN 扩张因果卷积过程

相比于一般的 CNN 卷积, 由于扩张卷积严格遵循时间序列等顺序特性, 其表现更出色。为了从多模态数据中提取和挖掘更丰富的特征, 使用扩张因果卷积来推断时间序列规律。时间序列  $x_i \in R^N$ ,  $f$  是时间序列上的扩张卷积函数, 可以定义元素  $v$  如公式 (10) 所示:

$$f(v) = \sum_{i=0}^{k-1} f(i) \cdot x_{v-d \cdot i} \quad (10)$$

$v-d \cdot i$  是下一个训练层的数据输入。 $d$  和  $k$  分别表示膨胀因子和卷积核的大小。膨胀因子表示使用卷积核时的固定步长, 即等于卷积核两个相邻部分之间的距离。膨胀因子  $d$  的数值决定了考虑每个卷积核时跳过多少个值。通常情况下, 在引入扩张因果卷积时, 膨胀因子  $d$  会随着网络层数  $i$  呈指数增长, 例如  $d=2 \wedge i, i=(0, 1, 2, \dots, N)$ 。这不仅能确保在较浅网络时卷积核覆盖所有输入的有效历史信息, 而且在更深层次时能够考虑非常长的有效历史信息。提出的 FTCN-SA 用于对时间序列情绪数据进行分类, 采用一维卷积, 膨胀因子分别设置为 1、2、4、8、16。

常用的解决梯度爆炸问题的方法是归一化和正则化, 然而这些方法并不适用于深度网络。因此本文应用两个非对称残差模块的方法。第一个残差模块包含三个单元, 输出是这两个通道的结果之和。第二个残差模块比第一个残差模块多一个单元。每个残差模块使用两种不同的卷积操作得到相同的输入, 然后对结果求和。

因此第  $i$  个滤波器的输入数据  $x$  被表示为:

$$\text{ReLU}(x) = \max(0, x) \quad (11)$$

$$f_{1i} = \text{ReLU}(w_1 * x + c) \quad (12)$$

$$f_{2i} = \text{ReLU}(w_2 * x + c) \quad (13)$$

$$f_i = f_{1i} + f_{2i} \quad (14)$$

$f_i$  表示通道的结果,  $*$  是 TCN 中的卷积运算。输出被定义为:

$$x_{k+1} = x_k + f(x_k) \quad (15)$$

$x_k$  表示输入数据在  $k$  层,  $x_{k+1}$  表示输出在  $k+1$  层。全连接层的输出  $o$  如公式 (16) 所示, 其中  $h \times w$  表示空间维度。

$$o = \frac{1}{h \times w} \sum_{p=1}^h \sum_{q=1}^w x_n(p, q) \quad (16)$$

## 3 实验结果与分析

为验证本文提出的早期 PD 智能检测系统的性能, 本文从系统检测的多维精度 (Accuracy, Precision, Recall, Specificity,  $F_1$ ), 及与其他检测方法的对比实验来验证了本系统的有效性。

### 3.1 PD 数据集处理

实验采用 mPower 的帕金森语音和步态数据, 可在 Synapse 研究门户网站获取, 数据包括 4 052 个人, 524 个为帕金森患者, 3 528 个为健康人。其中帕金森语音数据集包含 65 022 个 10 秒语音样本。步态数据集包含原始加速度计和陀螺仪数据。

系统测试实验首先需要将每个个体的语音和步态样本整理到相应患者编号的文件夹中, 然后进行多模态特征融合。基于 TCN 的早期 PD 诊断模型输入数据包括语音、步态和患者自测的特征数据。

### 3.2 实验环境设置

本次实验在浪潮服务器中进行, 硬件环境为 24 核 CPU, GeForce RTX3080, 10 GB 内存, 环境为 Linux。训练、验证和测试的数据集划分比例为 8:1:1, 便于在模型训练过程对模型进行优化和实验的比较分析。模型的训练参数设置如下, 训练轮次为 500 轮, batch-size 设置为 32, 学习率为  $1 \times 10^{-4}$ , 优化器为 Adam, momentum 设置为 0.5, 同时使用 dropout 层降低模型过拟合的风险。

### 3.3 系统性能的具体评估指标

系统测试主要采用混淆矩阵生成的准确率、回归率、精确率、 $F_1$  分数、GFLOPs 作为评价指标。其中 GFLOPs 是浮点运算次数, 用于描述模型的执行效率, 类似于算法中的时间复杂度。准确率为样本正确分类的概率, 即分类正确的样本占全部样本的比例:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

精确率为预测为正类的样本实际也是正样本的比例:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

召回率和敏感度为实际为正类的样本被预测为正类的比例:

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN} \quad (19)$$

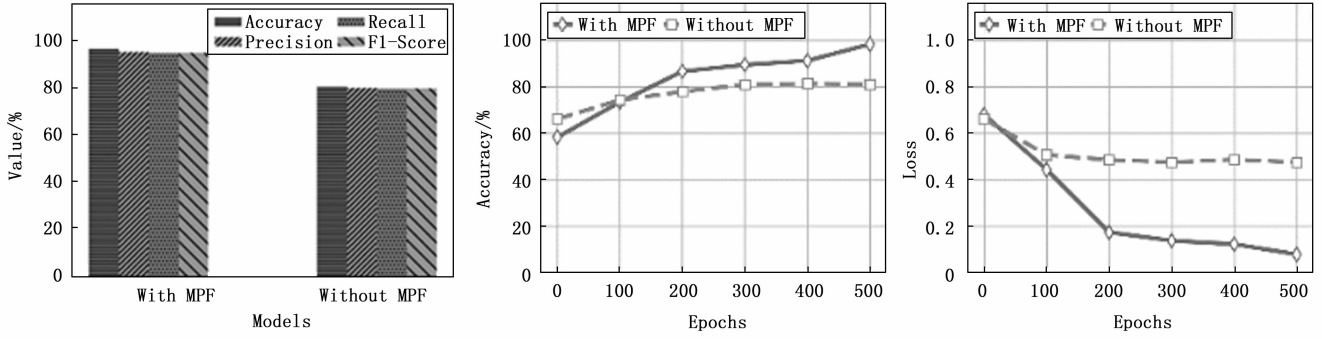


图 5 有无 MPF 的消融实验结果

特异度为实际为负类的样本被预测为负类的比例:

$$Specificity = \frac{TN}{TN + FP} \quad (20)$$

F<sub>1</sub> 值为精确率和召回率的调和平均数, 其值越高代表模型的分类效果越好:

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (21)$$

### 3.4 实验分析

#### 3.4.1 本系统有无多元线性池化融合方法 (MPF) 的对比消融实验

为了验证系统提出的多元线性池化方法 (MPF) 在进行多模态特征融合时的性能优势, 首先进行有无多元线性池化融合方法的消融实验, 实验结果如图 5 所示。

根据实验结果显示, 添加多元线性池化融合方法的特征融合模型的测试精度优于没有采用多元线性池化融合方法的模型。随着 epoch 迭代次数的增加, 具有多元线性池化融合方法的特征融合模型的精度迅速提高并趋于稳定。同时, 损失值迅速降低并趋于平稳。当迭代次数达到 500 次时, 精度值接近 1, 损失值低于 0.1。这表明多元线性池化方法在进行多模态特征融合时能够有效提升模型性能和训练效率。

#### 3.4.2 本系统多元线性池化融合方法 (MPF) 与其他特征融合方法的对比实验

为验证本系统的多元线性池化方法与其他特征融合方法相比是否具有优势, 进行了对比实验分析。选择常用的几个融合方法为本实验的对比算法, 包括平均值融合 M2、加权求和融合 M3 和最大值融合方法 M4, 实验结果如图 6 所示。实验结果表明, 本文提出的多元线性池化方法能够有效地融合多模态数据, 具体来说, 在准确率、召回率、精确度和 F<sub>1</sub> 等多项评估指标中, 多元线性池化融合方法明显高于其他 3 种融合方法。此外, 多元线性池化 Recall 值比其他方法略高 5%。其中, 加权求和融合方法以简单高效的低维形式融合多模态信息。

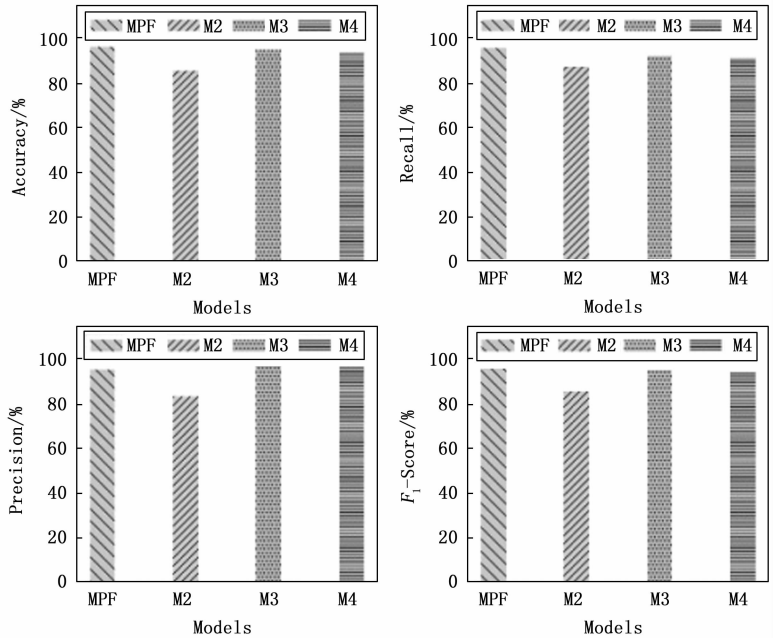


图 6 MPF 与其他多模态特征融合方法的对比实验 (M2 指平均值融合方法, M3 是加权求和融合方法, M4 指最大值融合方法)

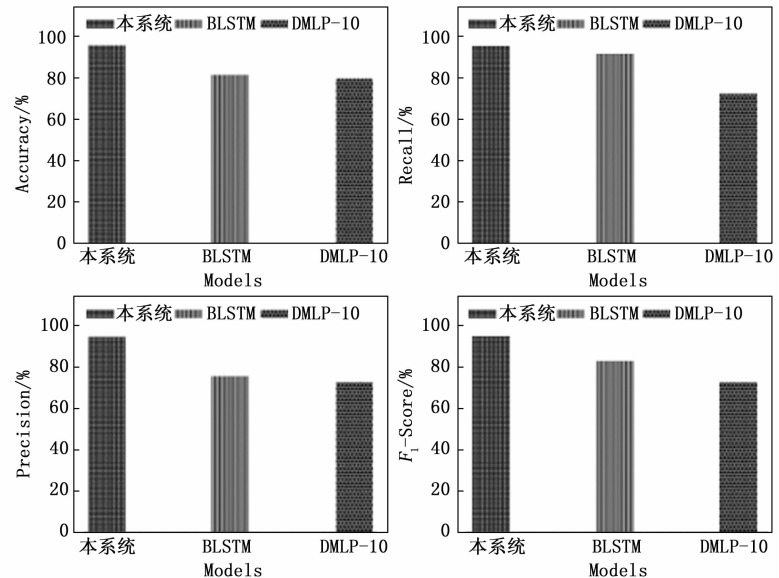


图 7 本系统与其他帕金森病检测方法 (BLSTM 和 DMLP-10) 的检测结果对比

### 3.4.3 本系统 PD 诊断结果与其他诊断方法的对比实验

为验证本系统检测性能, 与当前常用的帕金森诊断方法 (BLSTM 和 DMLP-10) 进行对比, 结果如图 7 所示。实验结果表明, 本系统模型在准确率、精确率、召回率和  $F_1$  分数等方面均优于递归 BLSTM 和 DMLP-10 两种方法。尤其是在准确率和精确度评价指标上, 相较于 BLSTM 和 DMLP 两种方法, 本系统模型精度高出近 15%。对比分析显示, 本文提出的多模态帕金森联合诊断系统能够以高精度对帕金森早期症状进行诊断。

## 4 结束语

本文提出一种基于时间卷积网络的帕金森联合诊断模型, 解决了早期帕金森病诊断精度低的问题。该方法结合语音、步态和患者自测数据, 并利用多元线性池化方法进行多模态特征融合, 有效考虑了模态间和模态内的特征相关性, 提高了多模态特征的融合效率。通过时间卷积网络进行多模态特征识别, 最终实现了早期帕金森病的诊断。消融实验和对比实验表明, 该基于 TCN 的早期 PD 智能检测模型的准确率为 96.2%, 召回率为 95.9%, 精确度为 95.2%,  $F_1$  值为 95.5%, 优于递归 BLSTM 和 DMLP-10 等多模态帕金森病检测方法。仿真实验结果表明, 该设计的基于 TCN 的早期 PD 检测系统能够高效、准确、智能地诊断 PD 疾病。

在后续研究中, 将从以下两个方面进一步提升模型的性能: 一方面是考虑 PD 检测模型性能, 加速模型训练收敛速度以提高模型精度; 另一方面是考虑早期 PD 智能诊断的数据隐私安全问题, 将融入隐私保护算法, 设计基于隐私保护算法的早期 PD 诊断安全防护机制, 综合保护敏感信息的安全。

### 参考文献:

- [1] 王 坚, 邹剑军, 常 颖, 等. 中国帕金森疾病蓝皮书 [J]. 中国临床神经科学, 2024, 32 (s1): 1-41.
- [2] BIOEM R, OKUN S, KLEIN C. Parkinson's disease [J]. *Lancet*, 2021, 397 (10291): 2284-2303.
- [3] 甘 雪. 病理性帕金森病猴模型构建过程中的病理学动态检测研究 [D]. 昆明: 昆明理工大学, 2022.
- [4] ZIRRA A, RAO C, BESTWICK J, et al. Gender differences in the prevalence of Parkinson's disease [J]. *Movement Disorders Clinical Practice*, 2022, 10 (1): 86-93.
- [5] 陈 敏, 宋清扬, 李 敏, 等. 老年帕金森患者心理健康危险因素 logistic 回归分析 [J]. *中华保健医学杂志*, 2022, 24 (6): 478-481.
- [6] 贾 杰. 基于功能障碍解决的老年帕金森病全周期康复管理 [J]. *中国医刊*, 2023, 58 (2): 117-119.
- [7] 王 崧. 帕金森病患者生存分析——一项十年随访研究 [D]. 大连: 大连医科大学, 2022.
- [8] 陈宁宁. 基于深度卷积神经网络的帕金森病分类算法研究 [D]. 曲阜: 曲阜师范大学, 2023.
- [9] 季 薇, 杨茗淇, 李 云, 等. 基于掩蔽自监督语音特征提取的帕金森病检测方法 [J]. *电子与信息学报*, 2023, 45 (10): 3502-3510.
- [10] YAN S, JIANG C, JANZEN A, et al. Neuronally derived extracellular vesicle  $\alpha$ -synuclein as a serum biomarker for individuals at risk of developing Parkinson disease [J]. *JAMA Neurol*, 2024, 81 (1): 59-68.
- [11] 邵暇荔, 张 泳, 金莉蓉, 等. 震颤为主型帕金森病与特发性震颤患者脑灌注对比分析 [J]. *中国现代神经疾病杂志*, 2023, 23 (6): 542-548.
- [12] 张小恒, 张馨月, 李勇明, 等. 面向帕金森病语音诊断的非监督两步式卷积稀疏迁移学习算法 [J]. *电子学报*, 2022, 50 (1): 177-184.
- [13] 季培琛, 李 晨. 基于多头注意力机制的多模态帕金森病安全检测系统 [J]. *计算机测量与控制*, 2024, 32 (3): 138-145.
- [14] 符宇辰. 基于语音的帕金森病分型识别方法研究 [D]. 南京: 南京邮电大学, 2022.
- [15] 郭东伟. 基于语音特征的帕金森病诊断算法研究 [D]. 重庆: 重庆大学, 2020.
- [16] OMARE, ABDELALIM M, et al. Enhancing parkinson's disease diagnosis accuracy through speech signal algorithm modeling [J]. *Computers, Materials & Continua*, 2022, 70 (2): 2953-2969.
- [17] WARULE P, MISHRA S, DEB S. Time-frequency analysis of speech signal using chirplet transform for automatic diagnosis of Parkinson's disease [J]. *Biomedical Engineering Letters*, 2023, 13: 613-623.
- [18] VIMALAJEEWA D, MCDONALD E, TUNG M, et al. Parkinson's disease diagnosis with gait characteristics extracted using wavelet transforms [J]. *IEEE Journal of Translational Engineering in Health and Medicine*, 2023, 11: 271-281.
- [19] DOTOV D, COCHEN V, DRISS V, et al. Coordination rigidity in the gait, posture, and speech of persons with Parkinson's disease [J]. *Journal of Motor Behavior*, 2023, 55 (4): 394-409.
- [20] RAHI B, LI M, QI M. A review of techniques on gait-based person re-identification [J]. *International Journal of Network Dynamics and Intelligence*, 2023, 2 (1): 66-92.
- [21] SITAPURE N, KWON J. Exploring the potential of time-series transformers for process modeling and control in chemical systems: an inevitable paradigm shift? [J]. *Chemical Engineering Research and Design*, 2023, 194: 461-477.
- [22] LIN Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition [J]. *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015: 1449-1457.
- [23] 宋振振, 陈兰岚, 娄晓光. 基于时序卷积网络的情感识别算法 [J]. *华东理工大学学报 (自然科学版)*, 2020, 46 (4): 564-572.